

**STATISTIQUE CANADA**  
**ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION**  
**VOLET MÉNAGES**  
**CYCLES 1 à 7 (1994-1995 à 2006-2007)**  
**DOCUMENTATION LONGITUDINALE**

**Juillet 2008**



Statistique  
Canada

Statistics  
Canada

**Canada**

## **NOTE AUX UTILISATEURS**

Cette documentation accompagne la diffusion de 7 cycles de données longitudinales du volet ménages de l'Enquête nationale sur la santé de la population (ENSP).

Ce document fournit un large éventail d'information sur l'ENSP : objectifs, contenu, plan d'échantillonnage, collecte, traitement, pondération, qualité, totalisation et accès aux données. Les chapitres 7, 8 et 11 donnent plus de détails concernant les différents sous-ensembles de répondants et les poids (échantillonnage et bootstrap) qui leur sont associés.

Ce document fait référence aux différents cycles de l'ENSP en utilisant parfois les années du cycle en question. À titre d'information, voici la correspondance entre les cycles et les années.

Cycle 1 = 1994-1995

Cycle 2 = 1996-1997

Cycle 3 = 1998-1999

Cycle 4 = 2000-2001

Cycle 5 = 2002-2003

Cycle 6 = 2004-2005

Cycle 7 = 2006-2007

Ce document s'adresse également aux utilisateurs du fichier partagé de microdonnées, c'est-à-dire aux ministères provinciaux de la Santé, à Santé Canada et à l'Agence de santé publique du Canada. Le fichier partagé comprend maintenant deux sous-ensembles (carré et complet – voir section 7) qui incluent les répondants ayant accepté de partager les données recueillies dans le cadre de l'ENSP (tous les cycles) de même que les poids d'échantillonnage et poids bootstrap correspondants. Ces sous-ensembles font également partie du fichier maître, mais sont destinés à l'usage des partenaires de partage. Les utilisateurs du fichier partagé ne doivent pas tenir compte des références spécifiques qui sont faites concernant les autres sous-ensembles de répondants.

**Liens rapides vers les pages WEB du site de Statistique Canada  
ENSP, Volet ménages – longitudinal**

**Documentation, ENSP, Volet ménages – longitudinal**

- [Cycle 7, Définitions et méthodes](#)
- [Cycle 7, Documents de références](#) - Sur cette page on retrouve entre autres, les liens vers :
  - Index des variables
  - Documentation longitudinale (ce document en format PDF)
  - Document sur les variables dérivées
  - Liste des variables dérivées, cycles 1 à 7
  - Résumé du contenu, cycle 1 à 8
  - Dictionnaire des données (fréquences arrondies)
- [Définitions et méthodes \(par cycle\)](#)

**Questionnaires**

- [Questionnaire ENSP, volet ménages - longitudinal, cycle 1 \(1994-1995\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 2 \(1996-1997\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 3 \(1998-1999\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 4 \(2000-2001\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 5 \(2002-2003\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 6 \(2004-2005\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 7 \(2006-2007\)](#)
- [Questionnaire ENSP, volet ménages - longitudinal, cycle 8 \(2008-2009\)](#)
- [Sommaire contenu ENSP, volet ménages - longitudinal cycles 1 à 7](#)

**Publication Internet de l'ENSP**

- [En santé aujourd'hui, en santé demain? Résultats de l'ENSP](#)

**TABLE DES MATIÈRES**

	Page
1. Introduction.....	1
2. Contexte .....	2
3. Objectifs .....	3
4. Contenu de l'enquête .....	4
4.1 Critères de sélection .....	4
4.2 Modifications apportées au contenu du cycle 7 (2006-2007) .....	5
4.3 Variables des interviews précédentes utilisées comme information complémentaire au cycle 7.....	5
5. Plan d'échantillonnage .....	6
5.1 Plan d'échantillonnage pour le cycle 1 (1994-1995) .....	6
5.1.1 Répartition de l'échantillon .....	6
5.1.2 Méthode d'exclusion .....	6
5.1.3 Prélèvement de l'échantillon .....	7
5.1.4 Plan d'échantillonnage au Québec .....	8
5.2 Échantillon longitudinal .....	9
6. Collecte des données.....	11
6.1 Conception du questionnaire et méthode de collecte des données .....	11
6.2 Essais .....	11
6.3 Interviews.....	11
6.4 Non-réponse et dépistage.....	13
7. Traitement des données.....	14
7.1 Vérification .....	14
7.2 Codage .....	14
7.3 Création de variables dérivées et groupées .....	15
7.4 Estimation et pondération .....	16
7.5 Définition du profil de réponse longitudinal (LONGPAT) .....	16
7.6 Définition d'une réponse complète et de la non-réponse.....	17
7.7 Sous-ensembles de répondants .....	17
8. Pondération.....	19
8.1 Pondération longitudinale .....	20
8.1.1 Point de départ : poids épuré du cycle 1 (1994-1995) .....	20
8.1.2 Ajustements apportés pour créer les différents poids longitudinaux.....	20
8.1.2.1 Poids longitudinal carré (WT64LS) .....	20
8.1.2.2 Poids longitudinal complet (WT6BLF).....	20
8.1.2.3 Poids longitudinal carré partagé (WT6BSLS) .....	29
8.1.2.4 Poids longitudinal complet partagé (WT6BSLF) .....	31
9. Qualité des données .....	32
9.1 Erreurs dues à l'échantillonnage.....	32
9.1.1 Méthode bootstrap pour l'estimation de la variance.....	32
9.2 Erreurs non dues à l'échantillonnage.....	33
9.2.1 Taux de réponse .....	34
9.2.1.1 Taux de réponse du cycle 1 (1994-1995) .....	34
9.2.1.2 Taux de réponse du cycle 2 (1996-1997) et du cycle 3 (1998-1999).....	36
9.2.1.3 Taux de réponse du cycle 4 (2000-2001), du cycle 5 (2002-2003) du cycle 6 (2004-2005) et du cycle 7 (2006-2007).....	37

9.2.2	Taux de refus .....	39
9.2.3	Taux de personnes non dépistées .....	40
9.2.4	Taux d'érosion.....	41
9.2.4.1	Taux d'érosion basé sur le sous-ensemble longitudinal complet....	41
9.2.5	Taux de refus et de « Ne sait pas » .....	44
9.2.5.1	Taux de refus et de « Ne sait pas » par item .....	44
9.2.5.2	Taux de refus et de « Ne sait pas » par personne .....	45
10.	Lignes directrices sur les totalisations, l'analyse et la diffusion .....	47
10.1	Lignes directrices sur l'arrondissement.....	47
10.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation.....	48
10.2.1	Définitions des genres d'estimations : de type nominal par opposition à quantitatif.....	48
10.2.2	Totalisation d'estimations de type nominal .....	49
10.2.3	Totalisation d'estimations quantitatives.....	49
10.3	Lignes directrices pour l'analyse statistique.....	50
10.4	Lignes directrices pour la diffusion.....	51
11.	Utilisation du fichier longitudinal .....	52
11.1	Utilisation des poids longitudinaux.....	52
11.2	Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap .....	52
11.3	Convention utilisée pour nommer les variables .....	53
11.3.1	Structure élémentaire des noms des variables .....	53
11.3.2	Positions 1 et 2 : Nom de la variable/ section du questionnaire.....	54
11.3.3	Position 3 : Type d'enquête.....	56
11.3.4	Position 4 : Variable de l'année de référence du cycle .....	57
11.3.5	Position 5 : Type de variable.....	57
11.3.6	Position 6 à 8 : Nom de la variable .....	57
12.	Accès aux données de l'ENSP.....	59
12.1	Centres de données de recherche.....	59
12.2	Télé-accès .....	59
12.3	Initiative de démocratisation des données.....	59
12.4	Rapports analytiques et totalisations personnalisées .....	59
13.	Technique d'analyse de données d'enquêtes longitudinales .....	61
13.1	Approche « jumelage de cycles » .....	61
13.2	Création du sous-ensemble modifié .....	63
13.3	Aspects méthodologiques de l'approche « jumelage de cycles ».....	65
13.4	Un exemple d'utilisation de l'approche « jumelage de cycles » : l'abandon du tabac .....	65

## **Liste des tableaux**

Tableau 5.A	Taille de l'échantillon longitudinal par province
Tableau 7.A	Distribution des décès selon l'année de décès
Tableau 7.B	Sous-ensembles de répondants
Tableau 8.A	Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes
Tableau 8.B	Variables pour l'ajustement de la non-réponse au cycle 2
Tableau 8.C	Variables pour l'ajustement de la non-réponse au cycle 3
Tableau 8.D	Variables pour l'ajustement de la non-réponse au cycle 4
Tableau 8.E	Variables pour l'ajustement de la non-réponse au cycle 5
Tableau 8.F	Variables pour l'ajustement de la non-réponse au cycle 6
Tableau 8.G	Variables pour l'ajustement de la non-réponse au cycle 7
Tableau 8.H	Variables pour l'ajustement de la non-réponse
Tableau 9.A	Information pertinente pour le calcul des taux de réponse du cycle 1
Tableau 9.B	Taux de réponse du Cycle 1
Tableau 9.C	Information pertinente pour le calcul des taux de réponse des cycles 2 et 3
Tableau 9.D	Taux de réponse du panel aux cycles 2 et 3
Tableau 9.E	Information pertinente pour le calcul des taux de réponse des cycles 4 à 7
Tableau 9.F	Taux de réponse du panel aux cycles 4, 5, 6 et 7
Tableau 9.G	Taux de refus par cycle
Tableau 9.H	Taux de personnes non dépistées par cycle
Tableau 9.I	Source de l'érosion par cycle — sous-ensemble longitudinal complet
Tableau 9.J	Taux de refus et de « Ne sait pas » par module
Tableau 9.K	Taux de refus et de « Ne sait pas » par personne
Tableau 10.A	Lignes directrices relatives à la variabilité d'échantillonnage
Tableau 11.A	Sous-ensembles de répondants et fichiers de poids bootstrap correspondants
Tableau 11.B	Variables longitudinales « constantes »
Tableau 13.A	Exemples de profil de réponse des répondants de l'ENSP
Tableau 13.B	Taux d'érosion
Tableau 13.C	Exemple de la structure du sous-ensemble modifié pour le jumelage

## **Liste des annexes**

Annexe A :	ENSP, volet ménages, changements au questionnaire pour le cycle 7 (2006-2007)
Annexe B :	ENSP, volet ménages, exemples de rétroaction des données du cycle 7 (2006-2007) et de questions de suivi

## **Autres documents de référence disponibles**

Questionnaire
Cliché d'article
Index alphabétique, Index par sujet
Dictionnaire de données
Documentation des variables dérivées, ENSP, cycles 1 à 7
Liste des variables dérivées, cycles 1 à 7
Résumé du contenu, ENSP, cycles 1 à 8

## 1. Introduction

L'Enquête nationale sur la santé de la population (ENSP) vise à recueillir des renseignements « longitudinaux » sur la santé de la population canadienne ainsi que des renseignements sociodémographiques connexes. Le premier cycle de l'enquête a eu lieu en 1994-1995. L'enquête sera menée tous les deux ans pendant dix cycles. L'ENSP a recueilli des renseignements de nature transversale et longitudinale durant les trois premiers cycles, mais depuis le cycle 4 (2000-2001) l'enquête est devenue uniquement longitudinale. La composante transversale du programme des enquêtes santé est menée par l'Enquête sur la santé dans les collectivités canadiennes (ESCC).

L'ENSP ne compte maintenant qu'un seul volet : le volet ménages. Depuis 2000-2001, le volet Nord est mené par l'ESCC plutôt que l'ENSP. Après 5 cycles de collecte (1994-1995 à 2002-2003), on a mis fin au volet établissements de soins de santé en raison du très grand nombre de décès dans l'échantillon.

La population cible du volet ménages de l'ENSP comprend les membres des ménages des dix provinces canadiennes en 1994-1995 à l'exclusion des personnes habitant dans une réserve indienne ou sur des terres de la Couronne, des résidents des établissements de soins de santé, des membres à temps plein des bases des Forces canadiennes et des personnes habitant certaines régions éloignées de l'Ontario et du Québec.

Sept cycles de collecte ont été complétés pour le volet ménages de l'ENSP : cycle 1 (1994-1995), cycle 2 (1996-1997), cycle 3 (1998-1999), cycle 4 (2000-2001), cycle 5 (2002-2003), cycle 6 (2004-2005) et cycle 7 (2006-2007).

Le volet ménages du cycle 7 de l'ENSP a recueilli des renseignements détaillés sur la santé des répondants longitudinaux sélectionnés au hasard lors du premier cycle de même que certains renseignements démographiques sur les membres du ménage des répondants longitudinaux. 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. On recueille également des renseignements sociodémographiques tels que l'âge, le sexe, le niveau de scolarité, le revenu du ménage et l'activité sur le marché du travail.

Le présent document a pour but de faciliter l'utilisation des fichiers maître et partagé longitudinaux du cycle 7, du volet ménages de l'ENSP. **Ces fichiers contiennent l'information recueillie auprès des répondants lors des cycles 1 à 7.** Les fichiers sont décrits de façon plus détaillée dans les chapitres suivants.

Pour les demandes d'information, les questions concernant l'accès aux données (Centres de données de recherche, accès à distance), les totalisations personnalisées, les renseignements généraux ou concernant l'utilisation des fichiers ou des données, veuillez vous adresser aux:

**Services d'information et d'accès aux données**, Division de la statistique de la santé :

Téléphone : 1-613-951-1746, Courriel : [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca), Télécopieur : 1-613-951-0792

## **2. Contexte**

À l'automne 1991, le Conseil national d'information sur la santé 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 des 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.

Depuis son lancement en 1994, l'ENSP fournit de l'information unique sur la santé des Canadiens, car elle répond aux besoins d'information sur la dynamique de la santé. L'ENSP est une enquête longitudinale basée sur un échantillon comptant 17 276 personnes réparties dans les dix provinces à travers le Canada. Tous les deux ans, les mêmes personnes donnent de l'information actuelle et détaillée sur leur état de santé physique et mentale, l'utilisation des services de santé, l'activité physique, la vie au travail et l'environnement social. Au cours de ces années de suivi, les données ont démontré comment une variété de facteurs peuvent contribuer à améliorer ou à détériorer la santé.

Alors que les données recueillies auprès de la population à un point fixe dans le temps procurent une image statique, les données longitudinales de l'ENSP révèlent les transitions vers une bonne ou mauvaise santé. La richesse des données de l'ENSP, c'est qu'elle permet aussi d'évaluer les liens entre les caractéristiques socioéconomiques et démographiques des individus avec leur état de santé et son évolution dans le temps.



### **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 facteurs économiques, sociaux, démographiques, professionnels et environnementaux ayant une influence sur la santé,
- 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 d'augmenter l'échantillon<sup>1</sup>,
- 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.

---

<sup>1</sup> Étant donné la nature longitudinale de l'ENSP, l'option d'augmenter l'échantillon n'est plus disponible, l'ESCC offre maintenant cette possibilité.

## **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 du questionnaire. La deuxième section décrit de façon sommaire les modifications apportées au contenu de l'enquête pour le cycle 7. La dernière section présente les variables des interviews précédentes qui ont été utilisées comme information complémentaire lors de la collecte du cycle 7.

Il est à noter que depuis le cycle 6 le contenu de l'ENSP est plutôt stable. Étant donné la nature longitudinale de l'Enquête, aucun nouveau contenu n'a été ajouté depuis. Toutefois, certains modules ou questions des cycles précédents sont répétés de façon cyclique dans le questionnaire afin que les utilisateurs des données puissent faire des comparaisons dans le temps.

**NOTA :** Vous trouverez à la page *iii* des liens vers les questionnaires de l'ENSP et vers le sommaire du contenu de l'ENSP pour 7 cycles disponibles dans le site Web de Statistique Canada.

### **4.1 Critères de sélection**

Le contenu de l'ENSP a été choisi selon les critères suivants :

- 1) les renseignements doivent être en rapport avec les objectifs des provinces 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<sup>2</sup>,
- 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.

Au cycle 1, une personne a été choisie au hasard dans chaque ménage (répondant longitudinal) pour répondre au questionnaire approfondi sur la santé (composante Santé - H06). Lors des trois premiers cycles, un certain nombre de renseignements (démographiques et santé) ont également été recueillis à propos de tous les membres de chaque ménage (composante Générale - H05) ainsi que sur le logement.

---

<sup>2</sup> Depuis le cycle 6, aucun nouveau contenu n'a été ajouté au questionnaire de l'ENSP. Ce critère n'est donc plus valide.

À partir du cycle 4, les questionnaires des composantes Générale et Santé ont été combinés en un seul questionnaire complété par les répondants longitudinaux. Les renseignements démographiques suivants : âge, sexe, état matrimonial, lien de parenté; sont recueillis à propos des autres membres du ménage. On recueille également de l'information concernant le logement.

Selon les critères susmentionnés, le questionnaire de l'ENSP porte 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, et la situation démographique et socio-économique. Par exemple, on évalue 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 est mesurée au moyen de questions sur les visites faites chez des professionnels de la santé (médecine traditionnelle et médecine alternative), les soins hospitaliers et la consommation de médicaments et autres remèdes. Les questions sur les déterminants de la santé ont trait à l'usage du tabac, la consommation d'alcool et l'activité physique. Il y a en outre des questions sur les examens préventifs qui demandent la fréquence et les raisons d'utilisation de ces services. Les questions sur les caractéristiques démographiques et socio-économiques couvrent, entre autres, l'âge, le sexe, le niveau de scolarité, le revenu du ménage et l'activité sur le marché du travail.

#### **4.2 Modifications apportées au contenu du cycle 7 (2006-2007)**

Tout comme au cycle précédent, les questions portant sur la santé sont posées en premier et sont suivies de celles sur les données socio-économiques (langue, niveau de scolarité, état de la population active et revenu).

Le contenu thématique du cycle 7 a été incorporé à la partie la plus appropriée du questionnaire. Le contenu thématique du cycle 7 comprend les modules suivants : Choix alimentaire (contenu thématique au cycle 5), Facteurs de stress chez les enfants et les adultes (« traumatismes ») (contenu thématique au cycle 4) et Insécurité alimentaire (contenu thématique au cycle 3).

L'annexe A présente en détail, les changements apportés au questionnaire du cycle 7 de l'ENSP.

#### **4.3 Variables des interviews précédentes utilisées comme information complémentaire au cycle 7**

Afin de réduire le fardeau de réponse, les questions posées lors des cycles précédents et dont la réponse ne peut changer dans le temps (le pays de naissance par exemple) ne sont plus posées.

Les variables pour lesquelles la réponse peut changer au cours du temps, mais uniquement dans certaines conditions (par exemple, le niveau de scolarité) ont été mises à jour au besoin.

Dans certains cas, les réponses recueillies lors des cycles précédents pouvaient être rappelées au répondant lors de l'interview du cycle 7, et ce, afin d'améliorer la qualité des données recueillies. Par exemple, les renseignements obtenus sur certaines conditions chroniques lors des interviews précédentes ont été utilisés pour aider à expliquer un changement au cycle 7. Voir l'annexe B pour plus de détails.

## 5. Plan d'échantillonnage

La population cible du volet ménages de l'ENSP comprend les membres des ménages des dix provinces canadiennes en 1994-1995 à l'exclusion des personnes habitant dans une réserve indienne ou sur des terres de la Couronne, des résidents des établissements de soins de santé, des membres à temps plein des bases des Forces canadiennes et des personnes habitant certaines régions éloignées de l'Ontario et du Québec. Ce chapitre décrit le plan d'échantillonnage utilisé au cycle 1 et explique comment l'échantillon longitudinal de 17 276 personnes a été obtenu.

### 5.1 Plan d'échantillonnage pour le cycle 1 (1994-1995)

L'Enquête sur la population active (EPA), remaniée en 1991, a servi de base pour le plan d'échantillonnage dans toutes les provinces, à l'exception du Québec où l'échantillon a été choisi à partir des ménages déjà interviewés par Santé Québec pour l'Enquête sociale et de santé de 1992-1993.

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

- la taille visée des échantillons nationaux et provinciaux,
- la décision de choisir un membre par ménage pour constituer le panel longitudinal, et
- le choix de l'EPA comme véhicule pour sélectionner l'échantillon.

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

#### 5.1.1 Répartition de l'échantillon

L'ENSP visait initialement un échantillon d'une taille de 19 600 ménages répondants. 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>3</sup>. Selon cette méthode, l'échantillon a été réparti proportionnellement à  $\sqrt{(0.804W_h^2 + 1/12^2)}$ , où  $W_h$  représente la proportion des ménages dans la province  $h$ ,  $h=1, \dots, 10$ , 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 du cycle 1 qui leur avait été alloué, mais pour des fins transversales seulement. Ces unités additionnelles ne font pas partie de l'échantillon longitudinal.

#### 5.1.2 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. 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

<sup>3</sup> Kish, L. (1988). Multipurpose Sample Design, Survey Methodology, 14, 19-32

personnes célibataires et les personnes âgées, seraient surreprésentées.

Par conséquent, une méthode d'exclusion pour accroître la représentativité des parents et des jeunes dans le panel a été adoptée. Pour ce faire, une partie de l'échantillon a été soumise à une sélection préliminaire. Une fois la liste des membres de ces ménages 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 le nombre prévu de ménages exclus au nombre de ménages visités dans chaque province.

La méthode d'exclusion des ménages qui n'ont 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 soumis à la sélection préliminaire 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.

### **5.1.3 Prélèvement de l'échantillon**

Un plan d'échantillonnage stratifié à plusieurs degrés a été utilisé pour le volet 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 la liste de logements présents dans chaque grappe choisie et à en sélectionner un certain nombre à partir de cette liste.

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'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 sondage 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, régions urbaines et régions rurales) à partir desquelles sont formées des strates géographiques 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'un certain nombre de non-réponses était attendu, et qu'une partie 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 d'un à six renouvellements par province, un renouvellement correspondant au sixième de l'échantillon total. On procède à la stabilisation de l'échantillon pour le maintenir au niveau voulu, notamment lorsqu'on réserve deux renouvellements, mais que la taille de l'échantillon requise représente seulement 1,5 renouvellement.

En raison des besoins particuliers de l'ENSP, deux modifications ont été apportées à cette stratégie d'échantillonnage. Le nombre de renouvellements « réservés » a été déterminé à l'échelle des strates plutôt qu'à l'échelle provinciale afin de répondre aux exigences relatives à la répartition infraprovinciale de l'échantillon pour des fins transversales au premier cycle. 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 regroupées de manière à former de plus grandes strates de l'ENSP comportant le nombre requis de grappes de l'échantillon. Une fois les strates regroupées, leurs grappes ont, elles aussi, été regroupées pour former des échantillons répétés.

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. Deux variables de plan, « stratum » et « replicat », se trouvent dans le fichier maître. Elles représentent la strate de l'EPA et l'échantillon répété de l'ENSP.

#### **5.1.4 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 classes de densité urbaine (la région métropolitaine de recensement de Montréal, les capitales régionales, les petites agglomérations urbaines et le secteur rural). Dans chaque secteur, les grappes ont été stratifiées selon des caractéristiques socio-économiques et sélectionnées à l'aide d'une méthode d'échantillonnage PPT. On a dénombré les grappes sélectionnées et on a prélevé des échantillons aléatoires de logements parmi celles-ci, à raison de 10 logements par grappe dans les grandes villes et de 20 ou 30 logements ailleurs.

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

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre classes de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à  $\sqrt{(2W_h^2 + 1/4^2)}$ , où  $W_h$  représente la part de la population pour la classe  $h$ ,  $h=1, 2, 3, 4$ . Dans chaque classe, on a tenté d'obtenir un sous-échantillon de l'ESS, qui, en ce qui concerne le membre sélectionné pour faire partie du panel, serait 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 été presque atteint.

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

## **5.2 Échantillon longitudinal**

L'échantillon longitudinal, aussi appelé panel longitudinal ou simplement panel, est composé des 17 276 personnes sélectionnées au cycle 1 qui ont au moins rempli la composante générale du questionnaire du cycle 1. Il comprend également les 2 022 enfants provenant du premier cycle de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ). Ces enfants ont été interviewés par l'ELNEJ pour l'ENSP au cycle 1 et sont interviewés par l'ENSP depuis le deuxième cycle. C'est cet échantillon longitudinal (17 276) qui a été et sera suivi dans tous les cycles de l'ENSP. Il est à noter que les échantillons supplémentaires ajoutés aux cycles 1, 2 et 3 à des fins transversales ne font pas partie de l'échantillon longitudinal.

L'échantillon longitudinal n'est pas renouvelé au fil du temps. Aucun membre du panel n'a été et ne sera classé hors du champ de l'enquête. La taille de l'échantillon longitudinal demeurera

donc égale à 17 276 pour tous les cycles. Par conséquent, au cycle 7, les membres du panel longitudinal étaient tous âgés d'au moins onze ans et l'échantillon longitudinal ne contenait aucune personne ayant immigré au Canada après 1994-1995.

Le nombre de personnes répondant à l'enquête diminue légèrement d'un cycle à l'autre en raison de l'érosion causée par la non-réponse (par exemple, les refus et les personnes non dépistées). Malgré l'érosion, l'échantillon longitudinal est toujours représentatif de la population de 1994-1995. L'érosion étant relativement faible (voir section 9.2.4), elle ne devrait pas entraîner de fortes augmentations de la variance des estimations. Il est à noter que les membres du panel de l'ENSP qui sont décédés ou qui ont déménagé en établissement de soins de santé font toujours partie de l'échantillon longitudinal et sont considérés comme étant répondants (voir section 7.6). Ces personnes ne contribuent donc pas à l'érosion du panel longitudinal de l'ENSP.

Le tableau 5.A présente la taille de l'échantillon longitudinal par province en 1994-1995 et montre également le nombre de personnes ayant répondu de façon complète aux sept cycles de l'ENSP.

**Tableau 5.A : Taille de l'échantillon longitudinal par province**

<b>Province</b>	<b>Échantillon longitudinal Cycle 1 (1994-1995)</b>	<b>Nombre de répondants réponse complète aux cycles 1 à 7</b>
Terre-Neuve	1 082	746
Île-du-Prince-Édouard	1 037	719
Nouvelle-Écosse	1 085	704
Nouveau-Brunswick	1 125	728
Québec	3 000	1 890
Ontario	4 307	2 546
Manitoba	1 205	805
Saskatchewan	1 168	824
Alberta	1 544	979
Colombie-Britannique	1 723	1 051
<b>Total</b>	<b>17 276</b>	<b>10 992</b>



## **6. Collecte des données**

### **6.1 Conception du questionnaire et méthode de collecte des données**

Les questions de l'enquête 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 spécifié 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 ou à l'intervieweur d'obtenir une rétroaction immédiate pour corriger les incohérences. L'ajout automatique de la période de référence d'après la date du jour est un autre avantage. Il est possible de « préinsérer » du texte ou des données selon les renseignements recueillis durant l'interview en cours et aussi durant les interviews des cycles précédents. Ceci permet à l'intervieweur de continuer son travail sans devoir revenir en arrière pour trouver des réponses antérieures. Cette possibilité permet notamment d'utiliser le nom ou le sexe du répondant pour 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 est adapté au répondant selon les données recueillies pendant l'interview en cours et lors des cycles précédents.

### **6.2 Essais**

L'application IAO 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. Du cycle 1 au cycle 6, deux essais sur le terrain ont été effectués. Quatre des bureaux régionaux de Statistique Canada ont participé à ces essais. Ces deux essais visaient principalement à observer les réactions des répondants face à de nouvelles questions ou nouveaux modules, 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 IAO (c.-à-d., questionnaire informatisé) ont aussi été soumises à des essais. Depuis le cycle 7, on n'effectue qu'un seul essai sur le terrain. Les échantillons des deux essais ont été jumelés en un seul. La majorité des non-répondants de longue date ont été retirés de l'échantillon. Un certain nombre a été conservé afin de pouvoir, entre autres, tester les procédures de dépistage. Le rôle de l'essai sur le terrain reste toutefois le même qu'auparavant.

### **6.3 Interviews**

Au cycle 7, la collecte des données du volet ménages a été répartie en quatre trimestres (commençant en juin, août et octobre 2006, et janvier 2007). Une autre période de collecte s'est tenue en juin 2007 pour effectuer des tentatives supplémentaires de suivi des non-répondants des trimestres précédents.

Au cycle 7, la collecte a été effectuée par des interviewers travaillant à partir des Centres d'appels de Statistique Canada situés à Edmonton, Sturgeon Falls, Sherbrooke et Halifax.

Une collecte spéciale est menée pour les membres du panel qui résident en établissements

de soins de santé. Pour ces répondants, l'interview a été menée en personne avec un questionnaire papier. On utilise le questionnaire du volet établissements de soins de santé et 272 répondants ont été interviewés de cette manière au cycle 7. Ils sont identifiés comme tels dans les fichiers maîtres et partagés (voir section 7.5). Il est à noter que le questionnaire du volet établissements de soins de santé recueille moins d'information que le volet ménages. Les variables non recueillies pour ces répondants sont codées avec un « 6 », c'est-à-dire « sans objet » dans les fichiers du volet ménages.

Tous les intervieweurs de l'ENSP sont des employés de Statistique Canada formés pour mener des enquêtes. La collecte des données de l'ENSP a été effectuée sous la structure de supervision et de contrôle mise en place par Statistique Canada. Les intervieweurs ont tous suivi une formation ciblée sur le contenu de l'ENSP et ont reçu un manuel de l'intervieweur à utiliser comme outil de référence.

Tous les membres vivants du panel longitudinal ont reçu par courrier une lettre d'introduction à l'enquête indiquant la tenue éminente de la collecte du cycle 7. Lors de cet envoi, les répondants du volet ménages ont également reçu une brochure présentant des informations générales sur l'enquête de même que quelques résultats tirés des données de l'ENSP, des extraits d'articles de journaux (En manchette) citant l'ENSP et, pour les remercier de leur participation, un outil permettant de calculer facilement l'indice de masse corporelle (IMC). De plus, de l'information sur l'ENSP était disponible pour les participants à l'enquête dans le site Web de Statistique Canada.

Aux cycles 6 et 7, on a effectué une enquête de suivi auprès d'un sous-échantillon des répondants à l'essai sur le terrain de novembre. L'objectif de ce suivi étant de vérifier les éléments suivants auprès de répondants :

- la clarté et utilité de la brochure, En manchette, Outil de calcul de l'IMC et du site Web de Statistique Canada pour les participants aux enquêtes,
- l'effet des documents sur leur motivation à participer,
- l'intérêt à recevoir de l'information supplémentaire sur l'ENSP (cycle 6).

Ces enquêtes de suivi indiquent que dans l'ensemble, les répondants qui reçoivent les documents sont satisfaits et les consultent. Toutefois, l'impact sur leur participation est minime. Des modifications à l'outil de calcul de l'IMC ont été apportées à la suite des suggestions reçues. Finalement, la majorité des répondants consultés souhaitaient recevoir de l'information additionnelle concernant l'enquête. À partir du Cycle 6, une question a donc été ajoutée à la fin du questionnaire pour identifier ces répondants. Ainsi, lors de l'envoi du matériel de remerciements, on ajoute du matériel d'information pour les répondants intéressés.

En général, on communique par téléphone avec les répondants du volet ménages. En fait, 99 % des interviews du cycle 7, volet ménages ont été effectuées par téléphone. On a procédé à des interviews en personne si le répondant n'avait pas le téléphone, à la demande du répondant ou lorsque ce dernier résidait en établissement de soins de santé. Le temps moyen de l'interview est d'un peu moins d'une heure.

Lors de la collecte, on a recueilli des renseignements généraux sur tous les membres du ménage (âge, sexe, liens de parenté entre les membres) du répondant longitudinal. Dans le cas des répondants longitudinaux âgés de 12 ans et plus, la déclaration par personne interposée n'a été permise que pour des raisons de maladie ou d'incapacité. Ces déclarations

par personne interposée représentent 5,7 % des données recueillies pour les répondants de 12 ans et plus. Toutes les interviews auprès des répondants âgés de moins de 12 ans ont été effectuées par personne interposée.

#### **6.4 Non-réponse et dépistage**

De nombreuses stratégies ont été adoptées pour réduire le nombre de cas de non-réponse. Par exemple, afin d'éviter de surcharger les intervieweurs, on a déterminé à partir des résultats des cycles précédents la taille maximale d'une tâche à attribuer à chaque intervieweur. La formation offerte aux intervieweurs incluait divers moyens pour réduire le nombre de cas de non-contact (par exemple, faire des appels ou des visites à divers moments de la journée) en se servant des renseignements de base recueillis au cours de l'interview précédente.

On a demandé aux intervieweurs de déployer tous les efforts raisonnables pour que les répondants longitudinaux acceptent de répondre au questionnaire de l'ENSP. Dans les cas où il n'y avait personne au domicile, plusieurs tentatives ont été faites pour contacter le répondant. Si le moment de l'appel (ou de la visite) était mal choisi, l'interviewer convenait d'un moment plus opportun pour rappeler (ou pour revisiter). S'il n'y avait pas de réponse (personne à la maison), de nombreux appels ont été effectués. Aux personnes qui ont refusé 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 répondant. Puis, l'intervieweur a fait un deuxième appel (ou une deuxième visite).

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 a fait l'objet d'un suivi lors des périodes de collecte subséquentes.

Le fait de ne pas dépister un répondant longitudinal est 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 autres contacts si recueillis au cycle précédent. En outre, on a formé les intervieweurs pour qu'ils se servent des outils publics 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 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 ordre de numéros. Le taux de non-réponse cumulatif lié à l'impossibilité de dépister le répondant longitudinal se chiffre à 5,4 % du panel total, ce qui est relativement faible pour le septième cycle de l'enquête. La section 9.2.3 traite plus en détail des taux de non-réponse due au non-dépistage.

On a tenté d'interviewer tous les membres du panel longitudinal qui ont déménagé au Canada ou aux États-Unis. On a également tenté de confirmer le lieu de résidence des membres du panel longitudinal qui ont déménagé à l'extérieur du Canada et des États-Unis. Aucune interview n'a été menée lorsque le membre demeure à l'extérieur du Canada et des États-Unis néanmoins l'information a été mise à jour pour le prochain cycle.

## 7. Traitement des données

### 7.1 Vérification

Une vérification des données a été d'abord effectuée lors de l'interview assistée par ordinateur (IAO). L'intervieweur ne pouvait pas saisir des valeurs situées hors de la fourchette permise. Les erreurs d'enchaînement étaient contrôlées à l'aide du système programmé de sauts de questions (commandes « passer à ») de l'IAO. Par exemple, le système d'IAO veillait à ce que toutes les questions ne s'appliquant pas au répondant ne soient pas posées. En cas de réponses incohérentes, des messages d'avertissement étaient donnés. Dans certains cas, l'incohérence a dû être résolue avant de poursuivre l'interview et dans d'autres cas, l'incohérence a été acceptée et aucune mesure correctrice n'a été prise. Après la collecte des données, un processus de vérification (plusieurs programmes) a été exécuté au bureau central. En général, les incohérences ont été corrigées en donnant la valeur « non déclarée » à une ou plusieurs des variables en cause. Aucune imputation n'a été effectuée.

### 7.2 Codage

Pour certaines questions de l'ENSP une réponse au long a été saisie (ex., médicaments, maladies, profession, lieu de travail). Lors du traitement des données, les réponses à ces questions sont codées conformément aux systèmes de classification type utilisés par Statistique Canada.

Depuis la diffusion du fichier de microdonnées du cycle 4, les données concernant l'industrie et la profession sont codées en utilisant le « Système de classification des industries de l'Amérique du Nord (SCIAN) » et la « Classification type des professions de 1991 (CTP-1991) », et ce, pour tous les cycles.

Depuis la diffusion du fichier de microdonnées du cycle 5, la classification des médicaments est fondée sur la Classification anatomique thérapeutique chimique (ATC) développée par l'Organisation mondiale de la santé (OMS) telle que contenue dans la Base de données sur les produits pharmaceutiques (BDPP) de Santé Canada (septembre 2005), et ce, pour tous les cycles. La liste complète des codes est disponible sur demande.

Depuis la diffusion du fichier de microdonnées du cycle 6, les problèmes de santé causant une limitation d'activités ont été codés conformément à la Classification statistique internationale des maladies et des problèmes de santé connexes, 10e révision (CIM-10), et ce, pour tous les cycles. De plus, le Système de codage supplémentaire pour les incapacités ostéo-articulaires et musculaires n'est plus utilisé et la codification basée sur la CIM-9 n'est plus disponible à Statistique Canada. Par conséquent, le fichier de microdonnées du cycle 7 et les suivants contiendront uniquement la codification en CIM-10, et ce, pour les cycles 1 à 7.

Plusieurs questions de l'ENSP proposent des catégories de réponses et permettent également de saisir une réponse au long (spécifiez). Lors du traitement des données, lorsqu'appropriées, ces réponses au long sont codées dans des catégories existantes. À l'occasion (surtout au cycle 1), de nouvelles catégories ont été créées.

Le décès d'un membre du panel est, lorsque possible, confirmé avec la Base canadienne de données de l'état civil – Décès. Lorsque le décès est confirmé, on saisit la cause principale du décès ainsi que la date du décès (si elle diffère de celle recueillie lors de l'enquête). La cause de décès est ensuite codée à l'aide de la CIM-10 seulement (la codification CIM-9 n'étant plus disponible à Statistique Canada). Les variables se rapportant à un répondant décédé sont représentées par un « 9 » (non déclaré) dans le fichier.

Il y a un total de 2 032 membres du panel qui sont décédés au cours des 7 premiers cycles de l'ENSP. Parmi ce nombre, 1 773 décès ont été confirmés avec la Base canadienne de données de l'état civil – Décès. L'année du décès est connue pour 254 des 259 cas restants. Il reste donc 5 cas pour lesquels ni l'année ni la cause de décès ne sont connues. Le tableau suivant présente le nombre de décès et le nombre de décès confirmés selon les années.

**Tableau 7.A : Nombre de décès selon l'année de décès**

<b>Année du décès</b>	<b>Nombre de décès</b>	<b>Nombre de décès confirmés avec la Base canadienne de données de l'état civil – Décès</b>
1994	25	25
1995	115	115
1996	152	149
1997	147	144
1998	188	184
1999	160	157
2000	180	176
2001	155	151
2002	182	181
2003	174	172
2004	191	188
2005	179	131
2006	146	0
2007	33	0
Inconnue	5	0
<b>TOTAL</b>	<b>2 032</b>	<b>1 773</b>

### **7.3 Création de variables dérivées et groupées**

Pour simplifier l'analyse des données, plusieurs variables ont été dérivées d'après certains éléments du questionnaire de l'ENSP. Par exemple, plusieurs variables peuvent être combinées pour former une nouvelle variable dérivée. Le cinquième caractère du nom des variables dérivées est habituellement un D (voir section 11.3 sur la convention utilisée pour nommer les variables). Pour plus de détails sur le calcul des variables dérivées, consulter le document « Enquête nationale sur la santé de la population, Documentation des variables dérivées, Cycles 1 à 7 ».

À partir de certaines variables, on a créé des variables dites groupées, c'est-à-dire que les valeurs ont été regroupées afin de créer une autre variable. Dans certains cas, les variables dites groupées sont des variables dérivées simplement obtenues par la réduction du nombre de catégories de réponse. Le cinquième caractère du nom des variables groupées est habituellement un G (voir section 11.3 sur la convention utilisée pour nommer les variables).

#### **7.4 Estimation et 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 échantillonnée en représente 50 dans la population. Selon cette terminologie, on peut dire que chaque personne a un facteur de pondération (ou poids) de 50.

La phase de pondération est l'étape où l'on calcule ce poids pour chaque personne. Il faut utiliser ce poids pour produire des estimations valables à partir des données de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes au Canada dont l'état de santé général s'est détérioré entre deux cycles, 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 de pondération utilisée par l'ENSP est décrite au chapitre 8.

#### **7.5 Définition du profil de réponse longitudinal (LONGPAT)**

À chaque cycle, selon le résultat de l'interview, chacun des membres du panel se voit assigner un des cinq statuts de réponse suivants :

Complété (1) : statut donné aux membres du panel ayant fourni une réponse complète à l'interview, c'est-à-dire ayant répondu à toutes les questions jusqu'à un point déterminé dans le questionnaire de l'ENSP.

Décédé (2) : statut donné aux membres du panel décédés.

En établissement (3) : statut donné aux membres du panel ayant déménagé en établissement de soins de santé. Ce statut indique qu'une réponse complète ou partielle a été obtenue via le véhicule de collecte du volet établissements de soins de santé.

Réponse partielle (4) : statut donné aux membres du panel ayant rempli partiellement le questionnaire de l'ENSP.

Non-réponse (5) : statut donné aux membres du panel si aucun des statuts précédents n'a été assigné.

Au fil des cycles, les statuts de réponse d'un même répondant sont concaténés en une seule variable appelée le « profil de réponse longitudinal » (LONGPAT). Cette variable est disponible dans le fichier de microdonnées de l'ENSP et permet de connaître rapidement le profil de réponse d'un membre du panel. Cette variable est aussi utilisée pour définir les différents sous-ensembles analytiques tels que décrits à la section 7.7.

Il arrive parfois qu'on découvre au cycle suivant que le statut de réponse d'un cycle précédent soit erroné. Les correctifs sont alors apportés au profil de réponse longitudinal. Par exemple, on peut découvrir qu'un membre du panel avec un statut « *non-réponse* » est décédé en appariant avec la Base canadienne de données sur l'état civil-Décès. Le statut « *décédé* » sera alors assigné à cette personne en date du décès (par exemple, la variable LONGPAT était 115555 au cycle 6 et elle est devenue 1155222 au cycle 7).

## **7.6 Définition d'une réponse complète et de la non-réponse**

Depuis le cycle 4, l'ENSP est strictement longitudinale. La définition d'une réponse n'est pas la même du point de vue longitudinal et transversal. Pour le panel longitudinal de l'ENSP, une « réponse complète » comprend les membres du panel qui ont l'un des statuts de réponse suivants : complété, décédé ou déménagé en établissement de soins de santé.

Par conséquent, la non-réponse inclut les membres du panel ayant l'un des statuts suivants : réponse partielle ou non-réponse.

## **7.7 Sous-ensembles de répondants**

Afin d'offrir davantage de flexibilité aux utilisateurs, un seul fichier maître de microdonnées a été créé pour le cycle 7 de l'ENSP. Ce fichier contient les 17 276 membres du panel longitudinal de l'ENSP, peu importe le type de réponse obtenue aux cycles 1 à 7. Dans ce fichier maître, quatre sous-ensembles de répondants ont été créés avec les poids de sondage correspondants et les variables indicatrices rendant leur identification plus facile. Pour plus d'information relative au calcul des poids longitudinaux de chacun des sous-ensembles de répondants, veuillez vous référer au chapitre 8. Pour l'utilisation des poids longitudinaux, veuillez consulter la section 11.1. Le tableau 7.B donne une description de ces quatre sous-ensembles de répondants en fonction du type de réponse.

**Tableau 7.B : Sous-ensembles de répondants**

<b>Sous-ensemble de répondants</b>	<b>Type de réponse</b>	<b>Variable indicatrice</b>	<b>Nombre de répondants</b>
Longitudinal carré	Panel entier : tous les membres du panel, peu importe le type de réponse aux cycles 1 à 7	Aucune, tous les enregistrements	17 276
Longitudinal complet	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1 à 7	WF6BLF	10 992
Longitudinal carré partagé	Tous les membres du panel, peu importe le type de réponse aux cycles 1 à 7, qui ont autorisé le partage des données.	WF6BSLS	16 007
Longitudinal complet partagé	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1 à 7 et qui ont autorisé le partage des données.	WF6BSLF	10 668

Les utilisateurs du fichier partagé, c'est-à-dire les ministères provinciaux, Santé Canada et l'Agence de santé publique du Canada<sup>4</sup>, doivent prendre note que le sous-ensemble « longitudinal carré partagé » qui comprend la variable indicatrice du sous-ensemble « complet partagé » est fourni séparément sur un CD-ROM avec les poids d'échantillonnage correspondants pour chacun des sous-ensembles. Les poids d'échantillonnage et les variables indicatrices des autres sous-ensembles ne sont pas disponibles sur le CD-ROM du fichier partagé.

---

<sup>4</sup> Le gouvernement fédéral a créé une nouvelle structure pour s'adapter aux besoins en santé. Santé Canada et la nouvelle Agence de santé publique du Canada (ASPC) relèvent du ministre de la santé. L'ASPC est l'un des partenaires de partage pour l'ENSP.



## 8. Pondération

Le présent chapitre décrit le processus de pondération longitudinale pour chacun des sous-ensembles de répondants décrits à la section 7.7. Le processus de pondération longitudinale est nécessairement différent de celui de la pondération transversale pour plusieurs raisons. Premièrement, les poids longitudinaux doivent représenter la probabilité de sélection de l'unité d'analyse au moment de la sélection de l'échantillon. Puisque l'échantillon longitudinal a été sélectionné en 1994-1995, les poids doivent représenter la probabilité de sélectionner l'individu au cycle 1, et non dans les cycles qui suivent. De plus, la définition d'une réponse longitudinale est différente d'une réponse transversale, ce qui nécessite des ajustements différents particuliers à chaque type de non-réponse. Les analystes doivent toujours utiliser les poids longitudinaux pour les études faites à partir des sous-ensembles de répondants. Les poids longitudinaux ont été conçus spécifiquement pour représenter la population cible de 1994-1995. Aux cycles 1, 2 et 3, des fichiers transversaux et longitudinaux ont été produits. Malgré le fait que les membres du panel faisaient partie à la fois des fichiers transversaux et longitudinaux, leurs poids n'étaient pas identiques dans ces deux types de fichier, mais plutôt ajustés afin de bien représenter la population d'intérêt.

Pour le cycle 7, quatre ensembles de poids ont été créés : WT64LS, WT6BLF, WT6BSLS et WT6BSLF. Le tableau 8.A présente les variables de poids et les variables indicatrices associées avec chaque sous-ensemble. Un membre du panel fait partie d'un sous-ensemble donné lorsque la variable indicatrice de celui-ci est égale à 1.

**Tableau 8.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes**

Sous-ensemble de répondants	Poids	Variable indicatrice
Longitudinal carré	WT64LS	Aucune, tous les enregistrements
Longitudinal complet	WT6BLF	WF6BLF
Longitudinal carré partagé	WT6BSLS	WF6BSLS
Longitudinal complet partagé	WT6BSLF	WF6BSLF

Seuls les poids WT6BLF, WT6BSLS et WT6BSLF ont été corrigés pour la non-réponse. Par contre, les quatre poids ont tous été poststratifiés selon des groupes d'âge-sexe<sup>5</sup> pour chaque province, conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996. La poststratification est utilisée afin de s'assurer que les quatre sous-ensembles de répondants représentent correctement la population cible de 1994-1995. Les sections suivantes décrivent en détail la méthode de pondération longitudinale de l'ENSP.

<sup>5</sup> La poststratification se fait en utilisant la date de naissance mise à jour plutôt que la variable d'âge au cycle 1 (DHC4\_AGE) qui, elle, n'est jamais mise à jour.

## 8.1 Pondération longitudinale

La procédure utilisée pour la pondération longitudinale est basée essentiellement sur la pondération effectuée pour l'échantillon transversal du cycle 1 de l'ENSP. Quelques ajustements avaient été apportés à ce poids transversal afin d'incorporer les échantillons supplémentaires utilisés uniquement à des fins transversales. Pour la pondération du panel longitudinal, ces ajustements ont été enlevés pour créer un poids « épuré ». Ce poids épuré est le point de départ pour obtenir les poids longitudinaux.

### 8.1.1 Point de départ : poids épuré du cycle 1 (1994-1995)

Le poids épuré du cycle 1 est obtenu à partir du poids de base de l'EPA pour toutes les provinces à l'exception du Québec, et à partir du poids de base de l'Enquête sociale et de santé pour le Québec. Plusieurs ajustements sont apportés à ce poids afin de tenir compte des particularités de l'ENSP et de tenir compte de la probabilité réelle de sélection de chaque membre du panel. Tous ces ajustements effectués au cycle 1 sont conservés pour les cycles suivants puisque l'échantillon longitudinal réfère toujours à la même population de référence, c'est-à-dire celle de 1994-1995.

Une description complète des procédures de pondération utilisées au cycle 1, et toujours valides pour les cycles subséquents, est incluse aux sections 11.3 et 11.4 de la documentation du fichier de microdonnée à grande diffusion du cycle 2. [http://www.statcan.ca/francais/sdds/document/3236\\_D7\\_T1\\_V2\\_F.pdf](http://www.statcan.ca/francais/sdds/document/3236_D7_T1_V2_F.pdf)

À partir de ce point, des ajustements ont été appliqués au poids épuré afin d'obtenir les différents ensembles de poids longitudinaux.

### 8.1.2 Ajustements apportés pour créer les différents poids longitudinaux

#### 8.1.2.1 Poids longitudinal carré (WT64LS)

Le poids longitudinal carré **WT64LS** doit être utilisé avec le sous-ensemble longitudinal carré. Il est calculé en poststratifiant le poids épuré du cycle 1 conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996, par groupe d'âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l'intérieur de chaque province. L'ajustement dû à la poststratification est donné par :

$$\frac{\textit{Estimation démographique dans une classe province-âge-sexe}}{\textit{Somme des poids épurés des membres longitudinaux dans une classe province-âge-sexe}}$$

#### 8.1.2.2 Poids longitudinal complet (WT6BLF)

Le sous-ensemble longitudinal complet inclut seulement les membres du panel qui ont fourni une réponse complète, c'est-à-dire les membres qui ont eu un statut « complété », « décédé » ou « déménagé en établissement », à chaque cycle. Les membres du panel exclus de ce sous-ensemble ont donc été non-répondants, c'est-à-dire qu'ils ont eu un statut « réponse partielle » ou « non-réponse », à un certain point durant les sept premiers cycles de l'enquête, et leur poids doit donc être redistribué pour compenser la non-

réponse.<sup>6</sup>

Le point de départ est le poids épuré du cycle 1 et des ajustements pour la non-réponse sont faits. Un ajustement différent est effectué pour la non-réponse à chacun des cycles et ces ajustements sont cumulatifs d'un cycle à l'autre. Par exemple, pour obtenir le poids du cycle 7, les ajustements pour la non-réponse aux cycles 2 à 7 sont appliqués successivement au poids épuré du cycle 1.

Les ajustements nécessaires en vue d'obtenir le poids longitudinal complet du cycle 7 sont décrits ci-dessous.

### **Ajustement 1 : Ajustement pour la non-réponse au cycle 2 (1996-1997)**

L'ajustement pour la non-réponse a été fait en utilisant l'approche des classes de pondération. Ces dernières représentent des groupes de répondants qui affichent la même propension à répondre à l'enquête. On se fonde sur les caractéristiques des répondants et des non-répondants au cycle 2 de l'enquête, obtenues d'après les données du cycle 1, pour déterminer à quelle classe de pondération ils appartiennent. On définit les classes 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 les 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 forts 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 classes de pondération.

L'algorithme CHAID (« Chi-Square Automatic Interaction Detection ») a été utilisé pour définir les classes de pondération. Un minimum de 30 unités par classe de pondération a été utilisé afin de produire des ajustements plus stables.

Des classes de pondération ont été créées séparément pour chaque province. Notez que dans ce cas, le terme province réfère à la province de résidence utilisée au moment de la sélection de l'échantillon en 1994-1995. Les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre longitudinal au cycle 1 ont été prises en considération. Quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les

---

<sup>6</sup> Voir la section 7.5 pour la définition du profil de réponse longitudinal et la section 7.6 pour la définition du statut de réponse complet pour l'ENSP.

caractéristiques personnelles provenant de la composante Santé n'ont pas été utilisées, car elles n'étaient pas disponibles pour plusieurs membres longitudinaux en 1994-1995.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 2 sont énumérées dans le tableau 8.B. Deux variables du plan d'échantillonnage du cycle 1 indiquant la présence dans le ménage de membres de moins de 25 ans et la présence de membres de moins de 12 ans ont été incluses. Une variable indiquant la non-réponse aux questions sur le revenu au cycle 1 ainsi qu'une variable indiquant si l'individu était âgé de 16 ans ou moins ont aussi été incluses. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.B.

**Tableau 8.B : Variables pour l'ajustement de la non-réponse au cycle 2**

DHC4_AGE	DHC4_MAR	GE34DURB	LFC4_1	SDC4DRAC
DHC4DECF	DHC4_OWN	HCC4DMDC	RAC4F1	SDC4GCB7
DHC4_DWE	GE34DCMA	INC4DIA5	SDC4DAIM	SEX

Afin de corriger pour les membres longitudinaux qui n'ont pas répondu au cycle 2, l'ajustement qui suit est apporté aux poids des répondants :

*Somme des poids de tous les membres longitudinaux*

*Somme des poids des membres longitudinaux répondants aux cycles 1 et 2*

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération.

**Ajustement 2 : Ajustement pour la non-réponse au cycle 3 (1998-1999)**

Les 15 672<sup>7</sup> enregistrements avec une réponse complète après les deux premiers cycles, autrement dit ceux qui ont un profil de réponse longitudinale 11, 12 ou 13<sup>8</sup>, servent de point de départ. Un ajustement pour la non-réponse est apporté aux cas avec un statut de « réponse longitudinale complète » après trois cycles (représentée par un des profils de réponse suivants : 111, 112, 113, 122, 131, 132 ou 133). Notez que tous les autres profils de réponse, c.-à-d. 114, 115, 134 et 135, forment la non-réponse. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 (profil 122) ou en établissement de soins de santé depuis le cycle 2 (profil 133) sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque

<sup>7</sup> Lors de la diffusion des données du cycle 2, le sous-ensemble longitudinal complet comptait 15 670 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y avait donc 15 672 cas avec une réponse complète après deux cycles plutôt que 15 670 cas.

<sup>8</sup> Voir la section 7.5 pour la définition du profil de réponse longitudinal et la section 7.6 pour la définition d'une réponse complète pour l'ENSP.

leur poids au cycle 2 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2 étaient probablement décédés ou en établissement de soins de santé.

L'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 3, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 2 seulement ont été considérées. Comme pour le cycle 2, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Cependant, contrairement à l'ajustement pour la non-réponse au cycle 2, les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 3.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 3 sont énumérées dans le tableau 8.C. Une variable du plan d'échantillonnage du cycle 1 indiquant le type de ménage (adulte ou enfant) a été incluse. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 2 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.C.

**Tableau 8.C : Variables pour l'ajustement de la non-réponse au cycle 3**

AD_6_1	DHC6_AGE	INS6_4	SDC6DAIM	SMS6_9A
AD_6_7	DHC6_MAR	INS6_6	SDC6_4P	SMS6_13A
ALC6WKY	DV_6_65J	LFC6_41	SDC6_5A	SMS6_13C
ALC6_3	EDC6_3	MHC6DWK	SDC6_5F	SMS6_13E
AM56_SHA	ES_6_80	MHC6_1A	SDC6_6B	SMS6_16D
AM66_PXY	GE36LMOV	MHC6_1B	SDC6_7A	SMS6_18A
AM66_SHA	HCC6F1	MHC6_1F	SDC6_7B	SMS6_18D
BPC6_10	HSC6DPAD	MHC6_1L	SDC6_7D	SP36_CPA
CCC6DNUM	HWS6_5	MHC6_13	SEX	SSC6D2
CCC6_1L	INC6DIA5	PC_6_40	SHS6_4	SSC6_3
CCC6_1N	INC6_1A	RPC6_3	SMC6_2	SSS6_2
DGC6_1D	INC6_3B	RSS6_1	SMC6_5	SSS6_4

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 3, l'ajustement suivant est apporté aux poids des répondants :

$$\frac{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 et 2}}{\text{Somme des poids des membres longitudinaux répondants aux cycles 1, 2 et 3}}$$

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Seuls les enregistrements avec les profils de réponse longitudinaux suivants ont été impliqués dans l'ajustement : 111 à 115, 131, 132, 134 et 135. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou en établissement de soins de santé depuis le cycle 2 ne font pas partie de cet ajustement.

**Ajustement 3 : Ajustement pour la non-réponse au cycle 4 (2000-2001)**

Les 14 631<sup>9</sup> enregistrements avec une réponse complète après trois cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou 3 ou en établissement de soins de santé depuis le cycle 2 ou 3 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2 ou 3 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2 ou 3 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 4, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 3 seulement ont été considérées. Comme pour le cycle 3, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 4.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 4 sont énumérées dans le tableau 8.D. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 3 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.D.

**Tableau 8.D : Variables pour l'ajustement de la non-réponse au cycle 4**

CCC8DANY	DGC8_1A	HCC8_1	PAC8_1A	SDC8_6A
CCC8_1C	DHC8_AGE	INC8DIA5	PAC8_1J	SDC8_7A
CCC8_1L	DHC8DECF	ISC8_1	PY_8DH1	SEX

<sup>9</sup> Lors de la diffusion des données du cycle 3, le sous-ensemble longitudinal complet comptait 14 619 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y avait donc 14 631 cas avec une réponse complète après trois cycles plutôt que 14 619 cas.

CCC8_1N	DHC8_OWN	NU_8_1B	RAC8F1	SSC8DEMO
CCC8_1V	FIC8F1	PAC8DFD	RPC8_2	SSC8DSOC
DGC8F1	GE38DURB	PAC8DLEI	SDC8_4A	TWC8_5

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 4, l'ajustement suivant est apporté aux poids des répondants :

**Somme des poids des membres longitudinaux répondants aux cycles 1, 2 et 3**  
**Somme des poids des membres longitudinaux répondants aux cycles 1 à 4**

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou 3 ou en établissement de soins de santé depuis le cycle 2 ou 3 ne font pas partie de cet ajustement.

**Ajustement 4 : Ajustement pour la non-réponse au cycle 5 (2002-2003)**

Les 13 597<sup>10</sup> enregistrements avec une réponse complète après quatre cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3 ou 4 ou en établissement de soins de santé depuis le cycle 2, 3 ou 4 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2, 3 ou 4 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2, 3 ou 4 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 5, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 4 seulement ont été considérées. Comme pour le cycle 4, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 5.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 5 sont énumérées dans le tableau 8.E. Trois variables du plan d'échantillonnage du cycle 1 ont été utilisées, indiquant la présence de membres du ménage qui ont moins de 12

<sup>10</sup> Lors de la diffusion des données du cycle 4, le sous-ensemble longitudinal complet comptait 13 582 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y avait donc 13 597 cas avec une réponse complète après quatre cycles plutôt que 13 582 cas.

ans, l'autre indiquant la présence de membres du ménage qui ont moins de 25 ans, et la dernière indiquant le type de ménage (adulte ou enfant). Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 4 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.E.

**Tableau 8.E : Variables pour l'ajustement de la non-réponse au cycle 5**

ALC0_3	DHC0_OWN	IMM	MHC0_1J	SMC0_2
ALC0DTYP	DHC0DL12	INC0DIA5	MHC0DCH	ST_0DC4
ALC0DWKY	DHC0DLE5	ISC0_1	MHC0DDS	ST_0DC5
AM60_SHA	GE30DURB	LSC0_1	PAC0DFD	ST_0DC6
BPC0_10	GHC0_21	LSC0DPFT	PAC0DLEI	ST_0DC8
CCC0DANY	HCC0DHPC	MHC0_16	SDC0_4A	ST_0DR2
DGC0F1	HSC0DHSI	MHC0_1A	SDC0_6A	ST_0DW3
DHC0_AGE	HWC0DSW	MHC0_1F	SEX	ST_0DW6

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 5, l'ajustement suivant est apporté aux poids des répondants :

*Somme des poids des membres longitudinaux répondants aux cycles 1 à 4*  
*Somme des poids des membres longitudinaux répondants aux cycles 1 à 5*

Cet ajustement est apporté selon les classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3 ou 4 ou en établissement de soins de santé depuis le cycle 2, 3 ou 4 ne font pas partie de cet ajustement.

***Ajustement 5 : Ajustement pour la non-réponse au cycle 6 (2004-2005)***

Les 12 559<sup>11</sup> enregistrements avec une réponse complète après cinq cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3, 4 ou 5 ou en établissement de soins de santé depuis le cycle 2, 3, 4 ou 5 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2, 3, 4 ou 5 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2, 3, 4 ou 5 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes

<sup>11</sup> Lors de la diffusion des données du cycle 5, le sous-ensemble longitudinal complet comptait 12 546 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y avait donc 12 559 cas avec une réponse complète après cinq cycles plutôt que 12 546 cas.



distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 6, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 5 seulement ont été considérées. Comme pour le cycle 5, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 6.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 6 sont énumérées dans le tableau 8.F. Deux variables du plan d'échantillonnage du cycle 1 indiquant la présence dans le ménage de membres de moins de 12 ans et la présence de membres de moins de 25 ans ont été utilisées. Une variable indiquant la non-réponse aux questions portant sur le revenu au cycle 5 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.F.

**Tableau 8.F : Variables pour l'ajustement de la non-réponse au cycle 6**

ALC2_2	DHC2DL12	LSC2_1	PAC2_3E	SSC2DTNG
ALC2_3	EDC2_4	LSC2_21	PAC2_3F	ST_2DC5
ALC2DWKY	EDC2D2	LSC2DPFT	PAC2DFD	ST_2DC6
CCC2_1C	GE32DURB	MHC2_1A	PAC2DFM	ST_2DC7
CCC2_1F	HSC2DEMO	MHC2_1C	RAC2_1C	ST_2DC8
CCC2_1L	HWC2DISW	MHC2_1F	RAC2_6D	ST_2DC9
CCC2DANY	HCC2DMDC	MHC2_1G	SEX	ST_2DW1
CCC2DNUM	IMM	PAC2_1F	SMC2_2	ST_2DW2
DHC2_AGE	INC2_1A	PAC2_3A	SMC2DTYP	ST_2DW5
DHC2_OWN	INC2DHH	PAC2_3B	SMC2DYRS	ST_2DW6
DHC2DHSZ	INC2DIA5			

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 6, l'ajustement suivant est apporté aux poids des répondants :

*Somme des poids des membres longitudinaux répondants aux cycles 1 à 5*  
*Somme des poids des membres longitudinaux répondants aux cycles 1 à 6*

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3, 4 ou 5 ou en établissement de soins de santé depuis le cycle 2, 3, 4 ou 5 ne font pas partie de cet ajustement.

**Ajustement 6 : Ajustement pour la non-réponse au cycle 7 (2006-2007)**

Les 11 619<sup>12</sup> enregistrements avec une réponse complète après six cycles servent de point de départ. Encore une fois, les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3, 4, 5, ou 6 ou en établissement de soins de santé depuis le cycle 2, 3, 4, 5, ou 6 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucun ajustement pour la non-réponse n'est fait puisque leur poids au cycle 2, 3, 4, 5, ou 6 a déjà été ajusté pour tenir compte du fait que certains non-répondants au cycle 2, 3, 4, 5, ou 6 étaient probablement décédés ou en établissement de soins de santé.

Comme pour les cycles précédents, l'approche des classes de pondération a été utilisée pour effectuer l'ajustement pour la non-réponse. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour l'ajustement de la non-réponse au cycle 7, les caractéristiques du ménage ainsi que les caractéristiques personnelles du membre du panel longitudinal au cycle 6 seulement ont été considérées. Comme pour le cycle 6, quelques caractéristiques reliées au plan de sondage de l'enquête ou au poids d'échantillonnage ont aussi été prises en considération dans le but d'incorporer le plan de sondage de l'enquête dans l'analyse. Les caractéristiques personnelles provenant de la composante Santé ont été utilisées puisqu'elles étaient disponibles pour tous les enregistrements impliqués dans l'ajustement pour la non-réponse au cycle 7.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse au cycle 7 sont énumérées dans le tableau 8.G. Deux variables du plan d'échantillonnage du cycle 1 ont été utilisées, l'une indiquant la présence dans le ménage de membres de moins de 25 ans et l'autre indiquant le type de ménage (adulte ou enfant). Une variable groupe d'âge du cycle 6 a aussi été incluse. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.G.

**Tableau 8.G : Variables pour l'ajustement de la non-réponse au cycle 7**

ALCA_2	DHCA_MAR	INCADHH	LSCA_21	SSCADAFF
ALCA_3	EDCA_4	INCADIAG5	PACADFD	STCADW1
DHCADECF	GHCA_2	INCA_3B	SDCA_6A	STCADW6
DHCADLVG	IMM	LSCADPFT	SMCA_2	STCADW7
DHCA_AGE				

<sup>12</sup> Lors de la diffusion des données du cycle 6, le sous-ensemble longitudinal complet comptait 11 593 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y avait donc 11 619 cas avec une réponse complète après six cycles plutôt que 11593 cas.

Afin de corriger pour les membres du panel qui n'ont pas répondu au cycle 7, l'ajustement suivant est apporté aux poids des répondants :

Somme des poids des membres longitudinaux répondants aux cycles 1 à 6  
Somme des poids des membres longitudinaux répondants aux cycles 1 à 7

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2, 3, 4, 5 ou 6 ou en établissement de soins de santé depuis le cycle 2, 3, 4, 5 ou 6 ne font pas partie de cet ajustement.

### **Ajustement 7 : Poststratification**

Le poids des unités faisant partie du sous-ensemble complet a finalement été poststratifié conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996, par groupe d'âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l'intérieur de chaque province. Ceci est fait pour s'assurer que la population de 1994-1995 est représentée correctement dans toute estimation provenant du fichier longitudinal. Cet ajustement de poststratification est donné par :

Estimation démographique dans une classe province-âge-sexe  
Somme des poids des membres longitudinaux répondants aux cycles 1 à 7  
dans une classe province-âge-sexe

Le poids final pour une réponse longitudinale complète, **WT6BLF**, est calculé en prenant d'abord le poids épuré du cycle 1, et en le multipliant ensuite par les ajustements 1 à 7.

### **8.1.2.3 Poids longitudinal carré partagé (WT6BSLS)**

Le sous-ensemble carré partagé est créé depuis le cycle 5. Ce dernier inclut les membres du panel qui ont autorisé le partage de l'information recueillie lors de toutes les interviews menées dans le cadre de l'ENSP avec les ministères provinciaux de la santé, Santé Canada et l'Agence de santé publique du Canada. Étant donné que ces partenaires reçoivent seulement les enregistrements de ceux qui ont accepté de partager leur information, un poids spécial, WT6BSLS, doit être créé de telle sorte que les estimations calculées à partir de ce sous-ensemble représentent correctement la population totale.

Afin de calculer ce poids, les deux ajustements suivants ont été apportés au poids épuré du cycle 1 : un ajustement pour la non-réponse (refus de partager ses données) et un ajustement pour la poststratification.

### **Ajustement 1 : Ajustement pour la non-réponse (refus de partager ses données)**

Comme pour le calcul du poids longitudinal complet (section 8.1.2.2), l'ajustement pour la non-réponse a été fait en utilisant l'approche des classes de pondération. Ces dernières représentent des groupes de membres du

panel qui affichent la même propension à accepter de partager leurs données. Les caractéristiques du ménage ainsi que les caractéristiques personnelles des membres du panel longitudinal pour les 7 cycles ont été considérées pour déterminer à quelle classe de pondération ils appartiennent. L'algorithme CHAID (« Chi-Square Automatic Interaction Detection ») a été utilisé pour définir ces classes de pondération et un minimum de 30 unités par classe a été utilisé afin de produire des ajustements plus stables. Des classes distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995.

Les variables choisies par l'algorithme CHAID afin de produire les classes de pondération pour corriger la non-réponse sont énumérées dans le tableau 8.H. Des variables indiquant si le membre du panel avait un statut « complété », « décédé », « déménagé en établissement », « réponse partielle » ou « non-réponse » aux cycles 3, 4, 6 et 7 ont été utilisées, de même qu'une variable indiquant s'il faisait partie du sous-ensemble longitudinal complet. Une variable du plan d'échantillonnage du cycle 1 indiquant la présence dans le ménage de membres de moins de 25 ans a été incluse, de même que des variables indiquant la non-réponse aux questions sur le revenu aux cycles 3 et 7. Veuillez consulter le dictionnaire des données pour une description complète des variables du tableau 8.H.

**Tableau 8.H : Variables pour l'ajustement de la non-réponse**

ALCA_2	DHC4DSZ	DHCB_OWN	EDCBD2	INCBDIAG5
DHC0DECF	DHC8_MAR	EDC2D2	IMM	SDC4_6A
DHC4_AGE	DHCB_MAR	EDC8D2	INCBDDHH	SEX

Afin de corriger pour les membres longitudinaux qui n'ont pas accepté de partager leurs données au cycle 7, l'ajustement qui suit est apporté aux poids de ceux qui ont accepté de le faire :

$$\frac{\text{Somme des poids de tous les membres longitudinaux}}{\text{Somme des poids des membres longitudinaux qui ont accepté de partager leurs données}}$$

Cet ajustement est apporté à l'intérieur de chacune des classes de pondération.

### **Ajustement 2 : Poststratification**

Le poids des unités faisant partie du sous-ensemble carré partagé a finalement été poststratifié conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996, par groupe d'âge (0-11, 12-24, 25-44, 45-64, 65 et plus) et par sexe à l'intérieur de chaque province. Ceci est fait pour s'assurer que la population de 1994-1995 est représentée correctement dans toute estimation provenant du fichier longitudinal. Cet ajustement de poststratification est donné par :

*Estimation démographique dans une classe province-âge-sexe  
Somme des poids des membres longitudinaux qui ont  
accepté de partager leurs données au cycle 7  
dans une classe province-âge-sexe*

Le poids final pour le sous-ensemble carré partagé, **WT6BSLS**, est obtenu en multipliant le poids épuré du cycle 1 par les ajustements 1 et 2.

#### **8.1.2.4 Poids longitudinal complet partagé (WT6BSLF)**

Comme pour le sous-ensemble carré partagé, le sous-ensemble complet partagé inclut les membres du panel qui ont accepté de partager l'information recueillie lors de toutes les interviews menées dans le cadre de l'ENSP avec les ministères provinciaux de la santé, Santé Canada et l'Agence de santé publique du Canada, mais seulement ceux qui ont fourni une réponse complète aux cycles 1 à 7. Étant donné que ces partenaires reçoivent seulement les enregistrements de ceux qui ont accepté de partager leur information, un poids spécial doit être créé de telle sorte que les estimations calculées à partir de ce sous-ensemble représentent correctement la population totale.

Un seul ajustement simple est apporté au poids longitudinal complet pour créer le poids du sous-ensemble complet partagé (fichier partagé). Cet ajustement est donné par :

*Somme des poids des membres longitudinaux répondants aux cycles 1 à 7,  
dans une classe province/profil de réponse longitudinal/âge-sexe  
Somme des poids des membres longitudinaux répondants aux cycles 1 à 7  
et ayant accepté de partager, dans une classe province/profil de réponse  
longitudinal/âge-sexe*

Il est à noter qu'aux cycles 3 à 7, quelques profils de réponse longitudinaux ont été regroupés afin de produire des ajustements plus stables. Le regroupement a été fait pour quelques classes province/âge-sexe qui n'avaient que quelques observations dans certains profils de réponse représentant les répondants décédés ou en établissement de soins de santé. Dans chaque cas, le profil de réponse problématique a été regroupé avec un autre profil de la même classe province/âge-sexe de telle sorte que la somme des poids corresponde toujours aux totaux de la population. Le poids final pour le sous-ensemble complet partagé, **WT6BSLF**, est obtenu en multipliant le poids longitudinal complet, **WT6BLF**, par cet ajustement. Puisque cet ajustement est fait en respectant les classes utilisées lors de l'ajustement pour la poststratification, aucune poststratification supplémentaire n'est nécessaire.

## 9. Qualité des données

L'évaluation de la qualité des données est un aspect important à toute enquête. L'étude de la qualité des données permet de vérifier la fiabilité et l'exactitude de l'information recueillie, ainsi que de s'en servir pour améliorer la qualité de la prochaine réalisation de l'enquête.

L'enquête permet de produire des estimations fondées sur l'information recueillie à partir d'un échantillon de personnes. On aurait pu obtenir des estimations quelque peu différentes si on avait effectué un recensement complet en utilisant le même questionnaire, les mêmes intervieweurs, les mêmes superviseurs, les mêmes méthodes de traitement, etc. que ceux utilisés pour l'enquête. La différence entre les estimations tirées de l'échantillon et celles qui découlent d'un dénombrement complet effectué dans des conditions semblables s'appelle l'erreur due à l'échantillonnage des estimations.

Les erreurs qui ne sont pas liées à l'échantillonnage peuvent être commises à presque toutes les étapes d'une enquête. Il est possible que les intervieweurs comprennent mal les instructions, que les répondants interprètent incorrectement les questions posées, que les réponses soient mal saisies ou que des erreurs se produisent au moment du traitement et de la totalisation des données. Tous ces exemples représentent des erreurs non dues à l'échantillonnage.

### 9.1 Erreurs dues à l'échantillonnage

Étant donné que les estimations d'une enquête par sondage comportent inévitablement une erreur due à l'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de l'ampleur de cette erreur due à l'échantillonnage. La mesure de l'importance éventuelle des erreurs dues à l'échantillonnage est fondée sur l'écart-type des estimations tirées des données de l'enquête. Cependant, en raison de la grande diversité des estimations que l'on peut tirer d'une enquête, l'écart-type d'une estimation est habituellement exprimé en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV), s'obtient en divisant l'écart-type de l'estimation par l'estimation elle-même et on l'exprime en pourcentage.

Par exemple, supposons qu'une personne estime que 25 % des Canadiens âgés de 12 ans et plus ont amélioré leur état de santé général entre les cycles 1 et 2 de l'enquête et que cette estimation comporte un écart-type de 0,003. On calcule alors le CV de cette estimation de la façon suivante :

$$(0,003/0,25) \times 100 \% = 1,20 \%$$

Statistique Canada utilise fréquemment les résultats du CV pour vérifier la qualité des estimations statistiques produites, et recommande vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ENSP d'en faire de même. Pour consulter les lignes directrices sur la façon d'interpréter les résultats du CV, se référer au tableau à la fin de la section 10.4.

#### 9.1.1 Méthode bootstrap pour l'estimation de la variance

Afin de déterminer la qualité de l'estimation et de calculer le CV, il est nécessaire de calculer l'écart-type. Les intervalles de confiance de même qu'un grand nombre de tests statistiques nécessitent également le calcul de l'écart-type de l'estimation.

Pour l'ENSP, on utilise un plan de sondage à plusieurs degrés, ce qui signifie qu'il n'y a pas de formule simple pour calculer les estimations de la variance. Par conséquent, on a recours à une méthode approximative. On applique la méthode bootstrap parce que l'information sur le plan d'échantillonnage doit être prise en compte lors du calcul des estimations de la variance. La méthode bootstrap permet d'effectuer ce calcul et, conjointement avec le programme Bootvar, elle demeure une méthode relativement facile à utiliser.

La méthode de ré-échantillonnage bootstrap utilisée pour l'ENSP sous-tend la sélection d'échantillons aléatoires simples, appelés les échantillons répétés, et le calcul de l'écart des estimations d'un échantillon répété à l'autre. Dans chaque strate, on sélectionne un échantillon aléatoire simple de  $(n-1)$  des  $n$  grappes avec remise pour former un échantillon répété. Il est à noter que puisque la sélection se fait avec remplacement, une grappe peut être choisie plus d'une fois. Pour chaque échantillon répété, on recalcule le poids de sondage de chaque enregistrement dans les grappes  $(n-1)$  sélectionnées. Ces poids sont par la suite poststratifiés en fonction de l'information démographique de la même façon que les pondérations du plan de sondage, ce qui permet d'obtenir les poids bootstrap finaux.

Le processus complet (sélectionner les échantillons aléatoires simples, recalculer et poststratifier les poids de chaque strate) est répété  $B$  fois,  $B$  prenant une grande valeur. En général, on utilise  $B=500$  pour l'ENSP afin de produire 500 poids bootstrap. Pour obtenir l'estimateur bootstrap de la variance, on doit calculer l'estimation ponctuelle de chacun des  $B$  échantillons. L'écart-type de ces estimations représente l'estimateur bootstrap de la variance. Statistique Canada a élaboré un programme qui peut effectuer tous ces calculs pour l'utilisateur : le programme Bootvar. Pour plus d'information concernant les poids bootstrap, veuillez vous référer au chapitre 11.2.

Le programme Bootvar est offert en formats SAS et SPSS. Il est constitué de macros qui calculent les variances de totaux, ratios, différences entre ratios, et pour des régressions linéaires et logistiques.

Le programme Bootvar est fourni avec les poids bootstrap et un document qui expliquent comment modifier et utiliser le programme selon les besoins de l'utilisateur.

## **9.2 Erreurs non dues à l'échantillonnage**

Beaucoup de temps et d'efforts ont été investis afin de 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 qui ont reçu une formation sur les procédures de l'enquête de même que sur les concepts utilisés dans le questionnaire. Au moment de la collecte, on a observé des interviews afin de déceler les problèmes et offrir des solutions lorsqu'approprié. La mise à l'essai de l'application IAO et les essais sur le terrain ont également été au nombre des procédures essentielles pour réduire au maximum 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 de 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 de non-réponse partielle, car

une fois le questionnaire débuté, les répondants ont tendance à le terminer. Dans la plupart des cas, la non-réponse partielle à l'enquête peut être attribuable au fait, que le répondant refuse de répondre, ne se souvient pas de l'information demandée ou n'est pas en mesure de fournir de l'information sur lui-même ou sur la personne pour laquelle il fournit une réponse par procuration. Il y a non-réponse totale lorsqu'il est impossible de dépister ou de joindre le répondant, qu'aucun membre du ménage ne peut fournir les renseignements voulus ou que le répondant refuse de participer à l'enquête. On a traité les cas de non-réponse en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. Voir la section 8.1.2 pour avoir de plus amples détails sur l'ajustement de la pondération pour la non-réponse.

Cette section présente de l'information concernant différentes composantes de la non-réponse. On aborde d'abord la non-réponse de façon globale en présentant les taux de réponse obtenus à chaque cycle. Suivent ensuite les sections relatives aux refus, aux membres du panel non dépistés et à l'érosion. Finalement, la non-réponse d'item est examinée brièvement.

### **9.2.1 Taux de réponse**

Ce chapitre présente les taux de réponse et décrit la façon dont ils sont calculés. Le calcul des taux de réponse du cycle 1 diffère de celui des cycles subséquents. Les taux de réponse du cycle 1 sont basés sur les 20 095 personnes faisant partie du champ d'enquête et choisies en vue de former le panel, alors qu'aux cycles subséquents ils sont basés sur les 17 276 personnes formant le panel. Autre différence importante, pour les trois premiers cycles, les taux de réponse des personnes choisies sont calculés pour la composante générale (H05) et la composante Santé (H06) (voir section 4.1). Comme l'enquête est devenue purement longitudinale au cycle 4 et qu'il n'y a plus de distinction entre ces deux composantes, il n'y a donc, à partir du cycle 4, qu'un seul taux de réponse pour le panel longitudinal.

#### **9.2.1.1 Taux de réponse du cycle 1 (1994-1995)**

Les taux de réponse du cycle 1 sont basés sur les 20 095 personnes faisant partie du champ de l'enquête et choisies en vue de former le panel. Par conséquent, les personnes qui faisaient partie des 3 165 ménages hors du champ de l'enquête (code de statut = 017, 018, 023, 024)<sup>13</sup> et des 2 983 ménages rejetés par la méthode d'exclusion (voir section 5.1.2) ont été exclues des calculs des taux de réponse du cycle 1.

##### ***Taux de réponse H05 des personnes choisies***

###### Nombre de personnes choisies ayant répondu à la composante H05

Toutes les personnes choisies à l'intérieur du champ de l'enquête

Pour la composante H05 du cycle 1, le taux de réponse des personnes choisies à l'échelle nationale était de **86,0 %**. À l'échelle provinciale, ce taux variait de 80,7 % en Ontario à 91,0 % en Alberta.

---

<sup>13</sup> 017 = Autre logement non admissible (p. ex. ambassade)  
018 = Ménage rejeté  
023 = En construction ou démolit  
024 = Logement vacant



### Taux de réponse H06 des personnes choisies

Nombre de personnes choisies ayant répondu à la composante H06  
Toutes les personnes choisies à l'intérieur du champ de l'enquête

Pour la composante H06, le taux de réponse des personnes choisies était de **83,6 %** à l'échelle du Canada, mais à l'échelle provinciale il variait de 77,8 % en Ontario à 89,1 % en Alberta.

L'information pertinente pour le calcul des taux de réponse est donnée dans le tableau 9.A et les taux de réponse à l'échelle nationale et provinciale sont présentés dans le tableau 9.B.

**Tableau 9.A : Information pertinente pour le calcul des taux de réponse du cycle 1**

Cycle 1 (1994-1995)				
Nombre de personnes choisies à l'intérieur du champ de l'enquête (1) = (2) + (4) ou (3) + (5)	Nombre de répondants		Nombre de non-répondants	
	H05 (2)	H06 (3)	H05 (4)	H06 (5)
20 095	17 276	16 794	2 819	3 301

**Tableau 9.B : Taux de réponse du Cycle 1**

Province	Taux de réponse du Cycle 1 (1994-1995)	
	H05 (2) / (1)	H06 (3) / (1)
Terre-Neuve/Labrador*	89,3 %	86,9 %
Île-du-Prince-Édouard	87,6 %	84,9 %
Nouvelle-Écosse	85,4 %	82,1 %
Nouveau-Brunswick	88,1 %	86,0 %
Québec	87,5 %	85,6 %
Ontario	80,7 %	77,8 %
Manitoba	89,5 %	87,0 %
Saskatchewan	88,5 %	86,9 %
Alberta	91,0 %	89,1 %
Colombie-Britannique	85,2 %	82,8 %
<b>Canada</b>	<b>86,0 %</b>	<b>83,6 %</b>

\* Le Labrador ne faisait pas partie de Terre-Neuve lors de la sélection de l'échantillon.

### 9.2.1.2 Taux de réponse du cycle 2 (1996-1997) et du cycle 3 (1998-1999)

Les taux de réponse des cycles 2 et 3 sont fondés sur les 17 276 personnes qui forment le panel longitudinal. Les personnes avec un statut de réponse « *décédé* » ou « *déménagé en établissement* » sont considérées comme des unités répondantes aux fins longitudinales. Par contre, les personnes avec un statut « *réponse partielle* » sont considérées comme des unités non répondantes (voir section 7.6). Aucun membre du panel n'est classé hors du champ d'enquête.

#### ***Taux de réponse du panel à la composante H05***

Nombre de membres du panel ayant répondu à la composante H05  
ou qui sont décédés ou qui ont déménagé en établissement de soins de santé  
Nombre de membres du panel

À l'échelle canadienne, les taux de réponse du panel étaient de **93,6 %** au cycle 2 et de **88,9 %** au cycle 3. À l'échelle provinciale, ces taux de réponse variaient de 90,4 % en Colombie-Britannique à 96,1 % à Terre-Neuve au cycle 2 et de 84,2 % en Colombie-Britannique à 92,5 % à Terre-Neuve au cycle 3.

#### ***Taux de réponse du panel à la composante H06***

Nombre de membres du panel ayant répondu à la composante H06 ou  
qui sont décédés ou qui ont déménagé en établissement de soins de santé  
Nombre de membres du panel

À l'échelle canadienne, les taux de réponse du panel étaient de **92,8 %** au cycle 2 et de **88,3 %** au cycle 3. À l'échelle provinciale, ces taux de réponse variaient de 89,8 % en Colombie-Britannique à 94,8 % à Terre-Neuve au cycle 2 et de 83,9 % en Colombie-Britannique à 92,0 % à Terre-Neuve au cycle 3.

L'information pertinente pour le calcul des taux de réponse des cycles 2 et 3 est donnée dans le tableau 9.C et les taux de réponse à l'échelle nationale et provinciale sont présentés dans le tableau 9.D.

**Tableau 9.C : Information pertinente pour le calcul des taux de réponse des cycles 2 et 3**

Cycle	Nombre de membres du panel <sup>14</sup>					
	Décédés (1)	En établissement (2)	Répondants complets		Non-répondants	
			H05 (3)	H06 (4)	H05 (5)	H06 (6)
2	289	62	15 819	15 687	1 112	1 238
3	615	114	14 647	14 532	1 912	2 015

**Tableau 9.D : Taux de réponse du panel aux cycles 2 et 3**

Province	Taux de réponse			
	Cycle 2 (1996-1997)		Cycle 3 (1998-1999)	
	H05 <u>(1) + (2) + (3)</u> 17 276	H06 <u>(1) + (2) + (4)</u> 17 276	H05 <u>(1) + (2) + (3)</u> 17 276	H06 <u>(1) + (2) + (4)</u> 17 276
Terre-Neuve/Labrador*	96,1 %	94,8 %	92,5 %	92,0 %
Île-du-Prince-Édouard	95,0 %	94,3 %	91,3 %	90,7 %
Nouvelle-Écosse	94,6 %	94,0 %	90,1 %	89,4 %
Nouveau-Brunswick	94,8 %	94,3 %	89,5 %	89,2 %
Québec	95,1 %	94,1 %	89,3 %	88,3 %
Ontario	92,1 %	91,4 %	87,6 %	86,9 %
Manitoba	95,4 %	94,4 %	91,1 %	90,6 %
Saskatchewan	94,7 %	94,0 %	91,0 %	91,0 %
Alberta	91,8 %	91,5 %	88,7 %	88,4 %
Colombie-Britannique	90,4 %	89,8 %	84,2 %	83,9 %
<b>Canada</b>	<b>93,6 %</b>	<b>92,8 %</b>	<b>88,9 %</b>	<b>88,3 %</b>

\* Le Labrador ne faisait pas partie de Terre-Neuve lors de la sélection de l'échantillon.

### 9.2.1.3 Taux de réponse du cycle 4 (2000-2001), du cycle 5 (2002-2003) du cycle 6 (2004-2005) et du cycle 7 (2006-2007)

Comme dans les cycles 2 et 3, les taux de réponse longitudinaux des cycles 4, 5, 6 et 7 sont obtenus à l'aide des 17 276 membres du panel longitudinal. Les personnes avec un statut de réponse « décédé » ou « déménagé en

<sup>14</sup> Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Les nombres présentés dans ce tableau reflètent la situation suite à ces modifications.

établissement » sont considérées comme des unités répondantes aux fins longitudinales. Par contre, les personnes avec un statut « réponse partielle » sont considérées comme des unités non répondantes (voir section 7.6). Aucun membre du panel n'est classé hors du champ d'enquête. Comme l'enquête est maintenant purement longitudinale et on ne fait plus de distinction entre les questionnaires H05 et H06, un seul taux de réponse est calculé et il est équivalent au H06.

**Taux de réponse**

$$\frac{\text{Nombre de membres du panel répondants ou qui sont décédés ou qui ont déménagé en établissement de soins de santé}}{\text{Nombre de membres du panel}}$$

À l'échelle canadienne, les taux de réponse du panel sont de **84,9 %** au cycle 4, **80,8 %** au cycle 5 et **77,6 %** au cycle 6. Au cycle 7, le taux de réponse est de **77,0 %**.

L'information pertinente pour le calcul des taux de réponse des cycles 4, 5, 6 et 7 est donnée dans le tableau 9.E et les taux de réponse à l'échelle nationale et provinciale sont présentés dans le tableau 9.F.

**Tableau 9.E : Information pertinente pour le calcul des taux de réponse des cycles 4 à 7**

Cycle	Nombre de membres du panel <sup>15</sup>			
	Décédés (1)	En établissement (2)	Répondants complets (3)	Non- répondants (4)
4	977	133	13 560	2 606
5	1 311	161	12 483	3 321
6	1 680	144	11 590	3 862
7	2 032	148	11 119	3 977

<sup>15</sup> Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Les nombres présentés dans ce tableau reflètent la situation suite à ces modifications.

Tableau 9.F : Taux de réponse du panel aux cycles 4, 5, 6 et 7

Province	Taux de réponse = $\frac{(1) + (2) + (3)}{17276}$			
	Cycle 4	Cycle 5	Cycle 6	Cycle 7
Terre-Neuve/Labrador*	89,2 %	82,6 %	80,9 %	82,6 %
Île-du-Prince-Édouard	88,2 %	84,1 %	80,6 %	82,5 %
Nouvelle-Écosse	87,0 %	81,6 %	81,0 %	82,4 %
Nouveau-Brunswick	83,8 %	79,4 %	75,8 %	78,3 %
Québec	85,5 %	79,8 %	75,2 %	77,3 %
Ontario	82,3 %	80,1 %	75,0 %	72,6 %
Manitoba	89,0 %	82,7 %	81,6 %	78,4 %
Saskatchewan	90,9 %	84,8 %	82,9 %	81,0 %
Alberta	83,3 %	80,3 %	79,3 %	77,6 %
Colombie-Britannique	80,8 %	77,9 %	76,5 %	73,5 %
<b>Canada</b>	<b>84,9 %</b>	<b>80,8 %</b>	<b>77,6 %</b>	<b>77,0 %</b>

\* Le Labrador ne faisait pas partie de Terre-Neuve lors de la sélection de l'échantillon.

### 9.2.2 Taux de refus

Dans une enquête longitudinale, la non-réponse à un ou plusieurs cycles, est très coûteuse. Elle rompt d'une certaine façon la séquence temporelle d'information disponible auprès d'un répondant, ce qui peut complexifier les analyses faites sur ces données. Pire encore, une non-réponse chronique à l'enquête réduit la taille de l'échantillon disponible pour l'analyse, réduisant ainsi le potentiel d'observer des résultats statistiquement significatifs. Par conséquent, de nombreux efforts doivent être déployés à l'ENSP afin de minimiser la non-réponse.

Malgré tous les efforts faits pour convertir les refus (voir section 6.4), ils demeurent la plus grande source de non-réponse pour l'ENSP. Même si l'intention est de suivre les 17 276 membres du panel au fil du temps, il y a des unités qui ne sont pas envoyées à la collecte à chaque cycle, tel que les refus fermes. Notez aussi que lorsque le décès d'un membre du panel a été confirmé par appariement avec les fichiers de mortalités ou que l'année de décès d'un membre est connue, ce membre est considéré répondant pour le reste de l'enquête et il n'est plus envoyé à la collecte.

Deux taux de refus peuvent être calculés pour chaque cycle, un basé sur les membres du panel qui sont envoyés à la collecte et l'autre basé sur tous les 17 276 membres du panel. Le tableau 9.G, qui présente les deux taux pour les cycles

2 à 7, démontre que les taux de refus augmentent à chaque cycle, mais tendent à se stabiliser. Toutefois, il faut noter que les efforts pour convertir les refus en valent vraiment la peine, car des membres du panel qui avaient refusé de répondre pendant quelques cycles consécutifs ont accepté de participer de nouveau à l'enquête.

**Tableau 9.G : Taux de refus par cycle**<sup>16</sup>

Cycle	Nombre de membres du panel envoyés à la collecte	Nombre de refus lors de la collecte	Taux de refus basé sur les membres envoyés à la collecte	Nombre de refus qui n'ont pas été envoyés à la collecte	Nombre total de refus	Taux de refus basé sur le panel complet (17 276)
2	17 266	538	3,1 %	1	539	3,1 %
3	16 582	601	3,6 %	460	1 061	6,2 %
4	16 186	1 014	6,3 %	514	1 528	8,9 %
5	15 616	1 301	8,3 %	664	1 965	11,5 %
6	14 743	1 071	7,3 %	1 213	2 284	13,4 %
7	13 857	844	6,1 %	1 444	2 288	13,2 %

### 9.2.3 Taux de personnes non dépistées

Après les refus, l'incapacité de dépister un membre du panel est la deuxième plus grande source de non-réponse à l'ENSP. Malgré les nombreux efforts des intervieweurs (décrits à la section 6.4), le taux cumulatif attribuable aux personnes non dépistées augmente cycle après cycle. Cependant, de nombreuses tentatives ont été faites pour maintenir ce taux le plus faible possible. Deux taux de personnes non dépistées peuvent être calculés pour chaque cycle, un basé sur les membres du panel qui sont envoyés à la collecte et l'autre basé sur tous les 17 276 membres du panel. Le tableau 9.H, qui présente les deux taux pour les cycles 2 à 7 démontre que les taux de personnes non dépistées augmentent à chaque cycle.

<sup>16</sup> Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Les nombres présentés dans ce tableau reflètent la situation suite à ces modifications.

Tableau 9.H : Taux de personnes non dépistées par cycle<sup>17</sup>

Cycle	Nombre de membres du panel envoyés à la collecte	Nombre de non-dépistés lors de la collecte	Taux de non-dépistés basé sur les membres envoyés à la collecte	Nombre de non-dépistés qui n'ont pas été envoyés à la collecte	Nombre total de non-dépistés	Taux de non-dépistés basé sur le panel complet (17 276)
2	17 266	295	1,7 %	0	295	1,7 %
3	16 582	359	2,2 %	0	359	2,1 %
4	16 186	500	3,1 %	0	500	2,9 %
5	15 616	698	4,5 %	0	698	4,0 %
6	14 743	869	5,9 %	2	871	5,0 %
7	13 857	686	5,0 %	247	933	5,4 %

### 9.2.4 Taux d'érosion

L'érosion dans le cadre d'une enquête longitudinale est une réduction de la taille d'échantillon attribuable aux non-répondants. Pour les cinq premiers cycles, l'érosion de l'échantillon de l'ENSP était définie uniquement selon qu'un membre du panel fasse ou non partie du sous-ensemble complet. Ainsi, aussitôt qu'une non-réponse était observée pour un membre du panel, celui-ci faisait partie de l'érosion. Ce taux permettait de bien caractériser la situation envers les utilisateurs de ce sous-ensemble de données, mais peignait toutefois un portrait plutôt pessimiste de la situation réelle avec l'échantillon. L'érosion peut être définie de différente façon dépendamment de la méthode d'analyse utilisée Cette section présente les résultats du calcul de l'érosion basé sur le sous-ensemble complet. Une autre façon de présenter l'érosion est discutée à la section 13 où on y explique l'approche du « jumelage de cycles », une technique d'analyse permettant de réduire l'impact de l'érosion. Pour de plus amples renseignements sur les différents sous-ensembles, voir la section 7.7

#### 9.2.4.1 Taux d'érosion basé sur le sous-ensemble longitudinal complet

L'érosion ici est définie relativement au fait qu'un membre du panel fasse ou non partie du sous-ensemble complet. Ainsi, aussitôt qu'une non-réponse est observée pour un membre du panel, celui-ci fait partie de l'érosion. Deux taux d'érosion différents sont calculés : un taux qui présente l'érosion entre deux cycles consécutifs et un taux qui présente l'érosion cumulative en fonction de l'échantillon initial. Ces deux taux sont calculés en utilisant le nombre de répondants dans le sous-ensemble complet.

<sup>17</sup> Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Les nombres présentés dans ce tableau reflètent la situation suite à ces modifications.

**Information pertinente pour le calcul des taux d'érosion :**

Nombre de membres du panel de l'échantillon longitudinal :	17 276
Nombre de personnes dans le sous-ensemble complet du cycle 2 :	15 672 <sup>18</sup>
Nombre de personnes dans le sous-ensemble complet du cycle 3 :	14 631 <sup>19</sup>
Nombre de personnes dans le sous-ensemble complet du cycle 4 :	13 597 <sup>20</sup>
Nombre de personnes dans le sous-ensemble complet du cycle 5 :	12 559 <sup>21</sup>
Nombre de personnes dans le sous-ensemble complet du cycle 6 :	11 619 <sup>22</sup>
Nombre de personnes dans le sous-ensemble complet du cycle 7 :	10 992

**Taux d'érosion d'un cycle à l'autre du sous-ensemble complet :**

Cycle 2 (1996-1997) :	$\frac{17\,276 - 15\,672}{17\,276} = \frac{1\,604}{17\,276} = 9,3 \%$
Cycle 3 (1998-1999) :	$\frac{15\,672 - 14\,631}{15\,672} = \frac{1\,041}{15\,672} = 6,6 \%$
Cycle 4 (2000-2001) :	$\frac{14\,631 - 13\,597}{14\,631} = \frac{1\,034}{14\,631} = 7,1 \%$
Cycle 5 (2002-2003) :	$\frac{13\,597 - 12\,559}{13\,597} = \frac{1\,038}{13\,597} = 7,6 \%$
Cycle 6 (2004-2005) :	$\frac{12\,559 - 11\,619}{12\,559} = \frac{940}{12\,559} = 7,5 \%$
Cycle 7 (2006-2007) :	$\frac{11\,619 - 10\,992}{11\,619} = \frac{627}{11\,619} = 5,4 \%$

<sup>18</sup> Lors de la diffusion des données du cycle 2, le sous-ensemble longitudinal complet comptait 15 670 enregistrements. Avec l'information disponible au cycle 7, nous sommes retournés dans les cycles précédents et avons pu, entre autres, confirmer que certaines non-réponses étaient en fait un décès et que certains décès auraient dû être une non-réponse. Certains cas, qui n'étaient pas complets jusqu'à présent, le sont devenus alors que d'autres qui l'étaient ne l'étaient plus. Suite à ces modifications, il y a donc 15 672 cas avec une réponse complète après deux cycles plutôt que 15 670 cas.

<sup>19</sup> Pour les mêmes raisons que dans la note précédente, suite aux modifications, le sous-ensemble longitudinal complet du cycle 3 compte 14 631 enregistrements au lieu de 14 619.

<sup>20</sup> Pour les mêmes raisons que dans la note précédente, suite aux modifications, le sous-ensemble longitudinal complet du cycle 4 compte 13 597 enregistrements au lieu de 13 582.

<sup>21</sup> Pour les mêmes raisons que la note précédente, suite aux modifications, le sous-ensemble longitudinal complet du cycle 5 compte 12 559 enregistrements au lieu de 12 546.

<sup>22</sup> Pour les mêmes raisons que la note précédente, suite aux modifications, le sous-ensemble longitudinal complet du cycle 6 compte 11 619 enregistrements au lieu de 11 593.



**Taux d'érosion cumulatifs du sous-ensemble complet :**

$$\begin{aligned} \text{Cycle 2 (1996-1997)} &: \frac{17\,276 - 15\,672}{17\,276} = \frac{1\,604}{17\,276} = 9,3\% \\ \text{Cycle 3 (1998-1999)} &: \frac{17\,276 - 14\,631}{17\,276} = \frac{2\,645}{17\,276} = 15,3\% \\ \text{Cycle 4 (2000-2001)} &: \frac{17\,276 - 13\,597}{17\,276} = \frac{3\,679}{17\,276} = 21,3\% \\ \text{Cycle 5 (2002-2003)} &: \frac{17\,276 - 12\,559}{17\,276} = \frac{4\,717}{17\,276} = 27,3\% \\ \text{Cycle 6 (2004-2005)} &: \frac{17\,276 - 11\,619}{17\,276} = \frac{5\,657}{17\,276} = 32,7\% \\ \text{Cycle 7 (2006-2007)} &: \frac{17\,276 - 10\,992}{17\,276} = \frac{6\,284}{17\,276} = 36,4\% \end{aligned}$$

Tel qu'observé typiquement dans les enquêtes longitudinales, le taux d'érosion entre les cycles 1 et 2 est considérablement plus élevé (9,3 %) que ceux observés par la suite. Ces derniers sont demeurés assez constants au cours des cycles. Cumulativement, on observe qu'après sept cycles, environ 36 % du panel fait partie de l'érosion reliée au sous-ensemble complet. Le tableau 9.1 présente les plus importantes sources de l'érosion au fil des cycles. Comme on l'a déjà mentionné, les refus et l'incapacité de dépister demeurent les plus importantes sources.

**Tableau 9.1 : Source de l'érosion par cycle — sous-ensemble longitudinal complet**

Source de l'érosion	Érosion cycle par cycle						Érosion cumulative
	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 7
Refus	30,6 %	45,4 %	52,2 %	51,6 %	48,0 %	41,3 %	43,8 %
Incapacité de dépister	17,1 %	12,5 %	16,8 %	15,6 %	20,3 %	21,7 %	17,0 %
Réponse partielle au cycle 1	30,0 %	s/o	s/o	s/o	s/o	s/o	7,7 %
Réponse partielle	6,9 %	8,9 %	5,9 %	8,3 %	6,4 %	5,7 %	7,1 %
Personne à la maison/ Sonnerie - Pas de réponse	1,7 %	4,7 %	8,2 %	9,3 %	14,0 %	22,7 %	8,5 %
Déménagés hors du Canada	4,7 %	7,5 %	5,1 %	1,9 %	2,5 %	2,7 %	4,2 %
Pas d'interview, problème de santé mentale/physique	s/o	s/o	2,4 %	4,9 %	2,2 %	1,6 %	1,7 %
Autres non-réponses	9,0 %	21,0 %	9,3 %	8,3 %	6,6 %	4,3 %	10,1 %

## 9.2.5 Taux de refus et de « Ne sait pas »

### 9.2.5.1 Taux de refus et de « Ne sait pas » par item

Les taux par item ont été calculés à partir du nombre de refus ou de « Ne sait pas » ainsi que du nombre de valeurs valides pour chaque variable, sous-module et module du questionnaire. Les variables dérivées ainsi que celles ayant peu de réponses valides ne sont pas présentées. Les valeurs valides excluent les réponses codées « sans objet » ou « non déclarées ». Le tableau 9.J montre la proportion des refus et des « Ne sait pas » pour les modules de l'ENSP au cycle 7.

**Tableau 9.J : Taux de refus et de « Ne sait pas » par module**

Module	Taux de Refus	Taux de « Ne sait pas »
Tous	0,54 %	1,05 %
Administration	0,06 %	0,07 %
Variables du dossier du ménage (DHC)	0,09 %	0,06 %
État de santé général (GHC)	0,19 %	0,25 %
Sommeil (SLC)	0,18 %	0,86 %
Taille et Poids (HWC)	0,18 %	0,17 %
Nutrition (NU_, FV_, SK_, MK_)	0,39 %	1,14 %
Santé préventive (PHC)	0,49 %	1,03 %
Utilisation des soins de santé (HCC)	0,22 %	0,38 %
Limitations des activités (RAC)	0,36 %	0,88 %
Problèmes de santé chroniques (CCC)	0,04 %	0,81 %
État de santé (HSC)	0,19 %	0,46 %
Activité physique (PAC)	1,49 %	1,40 %
Exposition aux rayons UV (TUC)	1,26 %	0,84 %
Mouvements répétitifs (RPC)	0,19 %	1,38 %
Blessures (IJC)	0,15 %	0,39 %
Stress (STC, ST_)	0,25 %	2,33 %
Consommation de médicaments (DGC)	0,10 %	0,55 %
Usage du tabac (SMC)	0,57 %	1,38 %
Consommation d'alcool (ALC)	0,29 %	0,83 %
Santé mentale (MHC)	0,58 %	0,74 %
Soutien social (SSC)	0,19 %	0,72 %
Sociodémographie (SDC) – langue	0,77 %	0,41 %

Module	Taux de Refus	Taux de « Ne sait pas »
Niveau de scolarité (EDC)	0,44 %	0,38 %
Population active (LSC)	0,55 %	0,73 %
Revenu (INC)	1,88 %	3,10 %
Insécurité alimentaire (FI_)	0,64 %	0,51 %

Dans ce tableau on peut voir que les taux de refus par module sont très faibles et varient de 0,04 % à 1,88 %. Le taux global de refus est de 0,54 %. D'un cycle à l'autre, on observe que ce sont souvent les mêmes variables et modules qui présentent un taux de refus supérieur à celui des autres variables ou modules. Comme au cycle précédent, le module du revenu a encore le taux le plus élevé (1,88 %). Le module de l'activité physique a le deuxième taux le plus élevé, soit de 1,49 %. Le module de l'insécurité alimentaire, qui n'était pas inclus aux 3 derniers cycles, a un taux de refus de 0,64 % ce qui est légèrement supérieur au taux global. Si on va plus loin dans l'analyse, bien que ce ne soit pas présenté dans le tableau, les taux de certains sous-modules de la population active sont parmi les plus hauts taux de refus des cycles 2 à 7 atteignant 1,03 % au cycle 7. En général, les taux de refus des variables varient entre 0,0 % et 2,11 %, à l'exception de certaines variables reliées au revenu total et à l'activité physique dont les taux de refus peuvent atteindre 3,48 % et 2,29% respectivement.

Les taux de « Ne sait pas » par module sont faibles et varient de 0,06 % à 3,10 %. Le taux global de « Ne sait pas » est de 1,05 %. D'un cycle à l'autre, on observe que ce sont souvent les mêmes variables et modules qui présentent un taux de « Ne sait pas » supérieur aux autres. Par exemple, le module du revenu a un taux de 3,10 %, et certains sous-modules reliés à l'emploi et aux problèmes de santé chronique ont des taux de « Ne sait pas » de 2,64 % et 3,86 % respectivement, soit parmi les plus élevés, et ce, à tous les cycles. Le module de stress a le deuxième taux le plus élevé, soit 2,33 % et trois de ses quatre sous-modules ont des taux supérieurs à 2,5 %. Entre autres, le sous-module sur les facteurs de stress chez les enfants et les adultes, inclus pour la dernière fois au cycle 4, atteint 2,71 % au cycle 7. De façon générale, les taux de « Ne sait pas » par variables varient entre 0,0 % et 21,8 % et ce sont des variables de revenu qui ont les taux les plus élevés.

#### 9.2.5.2 Taux de refus et de « Ne sait pas » par personne

Des taux de refus et de « Ne sait pas » ont aussi été calculés pour déterminer le pourcentage de questions auxquelles une personne refuse de répondre ou répond « Ne sait pas ». Ces taux ne sont déterminés que pour les personnes ayant un statut de réponse « complété » lors de l'interview du cycle 7. Le tableau 9.K montre la proportion des personnes qui n'ont refusé à aucune question, qui ont refusé de répondre à moins de 1 % ou à moins de 3 % des questions posées. Parallèlement, les proportions des personnes qui n'ont jamais répondu « Ne sait pas », qui

ont répondu « Ne sait pas » à moins de 0,5 %, à moins de 1 % et à moins de 5 % des questions posées sont présentées dans ce même tableau.

**Tableau 9.K : Taux de refus et de « Ne sait pas » par personne**

	Refus à 0 % des questions	Refus à moins de 1 % des questions	Refus à moins de 3 % des questions	« Ne sait pas » à 0 % des questions	« Ne sait pas » à moins de 0,5 % des questions	« Ne sait pas » à moins de 1 % des questions	« Ne sait pas » à moins de 5 % des questions
Total	90,2 %	95,7 %	96,3 %	48,6 %	70,2 %	86,0 %	95,8 %
Hommes	90,2 %	95,6 %	96,1 %	50,4 %	70,9 %	86,4 %	95,7 %
Femmes	90,2 %	95,9 %	96,5 %	47,1 %	69,6 %	85,7 %	95,9 %
< 12 ans	100,0 %	0,0 %	0,0 %	92,9 %	92,9 %	100,0 %	0,0 %
12-24 ans	93,2 %	96,9 %	97,1 %	43,3 %	68,1 %	84,1 %	94,1 %
25-44 ans	92,9 %	96,1 %	96,3 %	60,9 %	79,8 %	91,6 %	96,6 %
45-64 ans	89,9 %	96,1 %	96,7 %	51,2 %	74,1 %	89,6 %	97,2 %
> = 65 ans	84,5 %	93,8 %	95,1 %	29,9 %	51,1 %	73,3 %	93,4 %
Par procuration	88,3 %	94,9 %	95,9 %	39,8 %	51,1 %	69,7 %	94,9 %
Pas par procuration	90,3 %	95,8 %	96,4 %	49,1 %	71,2 %	86,9 %	95,8 %

On voit que les taux de « Ne sait pas » ont beaucoup plus de variation que les taux de refus quand ils sont séparés par sexe, groupe d'âge et type d'interview. Au cycle 7, 90,2 % ont accepté de répondre à toutes les questions et un peu plus de 96% ont des taux de refus inférieurs à 3 % des questions. Pour les réponses « Ne sait pas », il y a 48,6 % des personnes avec un taux de 0 %, c'est-à-dire près de la moitié des personnes n'ont jamais répondu « ne sait pas ». Il y a environ 70 % des personnes qui ont répondu « ne sait pas » à moins de 0,5 % des questions. Lorsque l'on regarde 1 % et 5 % des questions posées, ce taux grimpe à 86 % et 96 % respectivement. Ceci démontre que les personnes qui refusent de répondre ou répondent « Ne sait pas », ne le font que pour très peu de questions.

## **10. Lignes directrices sur les totalisations, l'analyse et la diffusion**

Cette section de la documentation décrit les lignes directrices que les utilisateurs devraient suivre pour créer des tableaux, analyser, publier ou diffuser de quelque façon que ce soit les données de l'ENSP. Ces lignes directrices devraient permettre aux utilisateurs de produire des chiffres qui se rapprochent de ceux produits par Statistique Canada, tout en étant capables d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

### **10.1 Lignes directrices sur l'arrondissement**

Pour que les estimations de l'ENSP qui seront publiées correspondent à celles de Statistique Canada, nous conseillons aux utilisateurs de se conformer aux lignes directrices suivantes pour l'arrondissement de ces estimations. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles le sont en réalité.

- a) Les estimations dans le corps d'un tableau statistique devraient être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on veut arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques devraient être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.
- c) Les moyennes, les proportions, les taux et les pourcentages devraient être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs ou les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on veut arrondir à un seul chiffre par cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou de rapports) devraient être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.
- e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

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

Le plan d'échantillonnage utilisé pour l'ENSP n'était pas autopondéré, c'est-à-dire que les personnes faisant partie de l'échantillon n'ont pas toutes le même poids d'échantillonnage. Pour produire des estimations simples, y compris des tableaux statistiques, les utilisateurs doivent employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir des différents sous-ensembles ne pourront être considérées comme représentatives de la population cible de 1994-1995 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 variable du poids, certains progiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

### 10.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatif

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ENSP, il est utile de décrire les deux principaux genres d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites.

#### ***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 classe définie. Le nombre de personnes ayant cessé de fumer entre les cycles est un exemple de ce type d'estimation. Une estimation du nombre de personnes possédant une caractéristique donnée peut également être appelée une estimation d'agrégat.

Exemple de question de type nominal :

*Actuellement, est-ce que ... fume(z) des cigarettes tous les jours, à l'occasion ou jamais? (SMCB\_2)*

- Tous les jours
- À l'occasion
- Jamais

#### ***Estimations quantitatives :***

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

Un exemple d'estimation quantitative est l'augmentation moyenne du nombre de cigarettes fumées par jour par des fumeurs quotidiens qui ont augmenté leur consommation entre deux cycles.

Exemple de question quantitative :

*Actuellement, combien de cigarettes est-ce que... fume(z) chaque jour?*  
(**SMCB\_4**)

|\_|\_| Nombre de cigarettes

### 10.2.2 Totalisation d'estimations de type nominal

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

- en additionnant le poids final des enregistrements contenant la caractéristique voulue pour le numérateur ( $\hat{X}$ ),
- en additionnant le poids final des enregistrements contenant la caractéristique voulue pour le dénominateur ( $\hat{Y}$ ), puis
- en divisant l'estimation du numérateur par celle du dénominateur.

### 10.2.3 Totalisation d'estimations quantitatives

Pour obtenir l'estimation d'une somme ou d'une moyenne pour une variable quantitative, on procède aux étapes suivantes (seule l'étape a) est nécessaire pour obtenir l'estimation pour une somme) :

- en multipliant la valeur de la variable étudiée par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements étudiés afin d'obtenir le numérateur ( $\hat{X}$ ),
- en additionnant le poids final des enregistrements présentant la caractéristique étudiée pour obtenir le dénominateur ( $\hat{Y}$ ), puis
- en divisant l'estimation du numérateur par l'estimation du dénominateur.

Par exemple, pour estimer le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours, on calcule d'abord le numérateur ( $\hat{X}$ ) en faisant la somme des produits de la variable **SMCB\_4** par le poids final pour tous les enregistrements pour lesquels la valeur de la variable **SMCB\_2** est « tous les jours ». On obtient le dénominateur ( $\hat{Y}$ ) en additionnant le poids final de tous les enregistrements pour lesquels la valeur de la variable **SMCB\_2** est « tous les jours ». Le nombre moyen de cigarettes fumées chaque jour par les personnes qui fument tous les jours est finalement obtenu en divisant ( $\hat{X}$ ) par ( $\hat{Y}$ ).

### **10.3 Lignes directrices pour l'analyse statistique**

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

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

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

Auparavant, on produisait des tableaux de CV pour les données transversales. Des tableaux de CV n'ont pas été créés pour les fichiers longitudinaux, car il existe un très grand nombre de combinaisons de variables possibles pour l'analyse. Pour estimer correctement la variance, les responsables de l'ENSP recommandent d'utiliser la méthode bootstrap. Grâce à cette méthode, la complexité de la pondération et le plan d'échantillonnage de l'enquête sont intégrés dans le calcul de la variance. Un programme d'estimation de la variance bootstrap de format SAS a été élaboré, offrant une documentation d'accompagnement et des exemples d'utilisation, pour faciliter le calcul de la variance à l'aide de la méthode bootstrap. Le programme permet également de calculer le coefficient de variation correspondant. Une version semblable du programme est aussi offerte en SPSS. Il importe que les utilisateurs apprennent à l'utiliser, car le programme générera les estimations de variances exactes pour évaluer la qualité des estimations mises en tableaux. On recommande fortement d'utiliser ce programme plutôt que l'approche des poids transformés. Certains progiciels statistiques, comme STATA, peuvent lire l'information des strates et des grappes servant à l'estimation de la variance, ce qui améliore la qualité de l'estimation, mais ne tient pas compte des différents ajustements appliqués aux poids.



#### 10.4 Lignes directrices pour la diffusion

Avant de diffuser ou de publier des estimations de totaux et de proportions tirées du fichier maître, les utilisateurs doivent d'abord déterminer le nombre de répondants dans l'échantillon ayant la caractéristique à l'étude (par exemple, le nombre de répondants qui fument lorsqu'on s'intéresse à la proportion de fumeurs pour une population donnée). Si ce nombre est inférieur à 10, l'estimation pondérée ne doit pas être diffusée, quelque que soit la valeur de son coefficient de variation. Cela tient au fait que la possibilité d'obtenir une variance artificiellement faible est plus grande si l'échantillon compte moins de 10 personnes. Pour les estimations pondérées basées sur des échantillons composés de 10 répondants ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes décrites dans le tableau 10.A.

**Tableau 10.A : Lignes directrices relatives à la variabilité d'échantillonnage**

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	0,0 – 16,5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	16,6 – 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 E (ou d'une autre manière similaire).
Inacceptable	Plus de 33,3	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 F (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 ».

## 11. Utilisation du fichier longitudinal

### 11.1 Utilisation des poids longitudinaux

Le fichier maître du cycle 7 contient les 17 276 membres du panel (incluant les répondants qui sont occasionnellement ou définitivement non-répondants). On a créé quatre sous-ensembles de répondants (section 7.7) auxquels correspond un ensemble de poids distinct (section 8.1). Des variables indicatrices ont été créées afin d'identifier les enregistrements appartenant à chacun des sous-ensembles de répondants (tableau 8.A). Les enregistrements qui ne font pas partie d'un sous-ensemble de répondants particulier ont une variable indicatrice et un poids égal à 0 pour ce sous-ensemble de répondants particulier. **Pour constituer le sous-ensemble de répondants d'intérêt, sélectionnez les enregistrements où la variable indicatrice appropriée est égale à 1.**

Le poids WT64LS est appelé « poids longitudinal carré » et s'applique à l'ensemble des 17 276 membres qui constituent l'échantillon longitudinal. Il est à noter que tous les cas de non-réponse devront être pris en considération lors des calculs.

Le poids WT6BLF est appelé « poids longitudinal complet » et s'applique aux 10 992 enregistrements qui forment le sous-ensemble « complet ».

Le poids WT6BSLS est appelé « poids longitudinal carré partagé » et s'applique aux 16 007 enregistrements qui forment le sous-ensemble « carré partagé ».

Le poids WT6BSLF est appelé « poids longitudinal complet partagé » et s'applique aux 10 668 enregistrements du sous-ensemble « complet partagé ».

### 11.2 Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap

Les poids bootstrap sont nécessaires au calcul de l'estimation de la variance. La section 9.1.1 contient de l'information sur la méthode bootstrap à appliquer pour le calcul de l'estimation de la variance. À chacun des sous-ensembles de répondants correspond un ensemble de poids bootstrap. Pour les données du cycle 7, quatre ensembles différents de poids bootstrap ont été créés, soit le carré, le carré partagé, le complet et le complet partagé. Le tableau 11.A présente les sous-ensembles de répondants et le nom des fichiers de poids bootstrap correspondants.

**Tableau 11.A : Sous-ensembles de répondants et fichiers de poids Bootstrap correspondants**

Sous-ensemble de répondants	Nom du fichier contenant les poids Bootstrap	Nombre de répondants
Longitudinal carré	B5long	17 276
Longitudinal complet	B5lngf	10 992
Longitudinal carré partagé	B5long_share	16 007
Longitudinal complet partagé	B5lngf_share	10 668

En raison de la complexité du plan d'échantillonnage, les utilisateurs devraient utiliser le programme Bootvar fourni pour calculer la variance. Les données de sortie standards pour l'écart-type provenant d'autres logiciels statistiques, comme SAS et SPSS, peuvent nettement sous-estimer la variance d'une estimation dans le cadre de cette enquête. **Il incombe à l'utilisateur d'assurer la qualité/fiabilité des estimations qu'il produit en respectant les lignes directrices énoncées au chapitre 10 et de calculer correctement la variance pour toutes les estimations.** Le non-respect de cette consigne pourrait conduire à une mauvaise interprétation des résultats et compromettre la qualité des travaux de recherche.

Certains logiciels statistiques peuvent intégrer l'information sur les strates et les grappes comme données d'entrée dans le traitement analytique, ce qui génère une estimation de variance qui se rapproche beaucoup plus de l'estimation de la variance réelle. Cependant, ces logiciels ne tiennent pas compte des divers ajustements de pondération qui peuvent, dans certains cas, influencer considérablement sur les estimations de la variance.

### **11.3 Convention utilisée pour nommer les variables**

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 de différentes périodes de collecte et de divers volets de l'ENSP. Les exigences qui suivent devaient être satisfaites : d'une part, limiter les noms des variables à au plus huit caractères 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, 2000-2001, 2002-2003, 2004-2005 et 2006-2007) dans le nom pour 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, 1996-1997, 1998-1999, 2000-2001, 2002-2003, 2004-2005 et 2006-2007. Les noms des variables pour ces questions ne devraient différer qu'à la position réservée à l'année qui identifie le cycle durant lequel les données ont été recueillies. Cette convention est appliquée pendant toute la durée de l'enquête longitudinale et est adoptée pour tous les volets de l'ENSP, soit le volet ménages, le volet établissements de soins de santé et, auparavant, le volet Nord et les suppléments.

#### **11.3.1 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 : Cycle

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, DHC6\_AGE, DHC8\_AGE, DHC0\_AGE, DHC2\_AGE, DHCA\_AGE et DHCB\_AGE est la suivante :
- DH :** dans la section du questionnaire sur la démographie et la composition du ménage;
- C :** questions qui font partie du contenu principal de l'enquête auprès des ménages;
- 4/6/8/0/2/A/B :** variable du cycle 1 (1994-1995) / variable du cycle 2 (1996-1997) / variable du cycle 3 (1998-1999) / variable du cycle 4 (2000-2001) / variable du cycle 5 (2002-2003) / variable du cycle 6 (2004-2005) / variable du cycle 7 (2006-2007);
- \_ :** Contenu thématique du cycle;
- AGE :** nom de la variable.

### **11.3.2 Positions 1 et 2 : Nom de la variable/ section du questionnaire**

AD	Dépendance à l'égard de l'alcool
AL	Alcool
AM	Administration de l'enquête
AP	Attitudes face aux parents
BF	Allaitement
BP	Tension artérielle
CC	Problèmes de santé chroniques
CO	Capacité de faire face au stress - contenu acheté par l'Alberta, cycles 1 et 2
CE	Contact — Sortie
DG	Consommation de médicaments
DH	Démographie et ménage
DV	Dentiste — visites
ED	Niveau de scolarité
ES	Services d'urgence
EX	Examen de la vue
FH	Antécédents personnels et familiaux
FI	Insécurité alimentaire — contenu acheté par DRHC, cycle 3
FS	Vaccination antigrippale
FV	Consommation de fruits et légumes
GE	Identificateurs géographiques (méthodologie)
GH	État de santé général
HC	Utilisation de soins de santé
HI	Informations sur la santé
HH	Ménage
HN	Numéro d'assurance-maladie
HS	État de santé
HV	VIH
HW	Taille et poids, Image corporelle
IJ	Blessures
IN	Revenu

**Positions 1 et 2 : Nom de la variable/ section du questionnaire (suite)**

IS	Assurance
LF/LS	Population active
MH	Santé mentale
MK	Consommation de lait
NU	Nutrition
PA	Activités physiques
PC	Examen général
PH	Santé préventive
PR	Province
PY	Ressources psychologiques (estime de soi, contrôle, sentiment de cohésion)
RA	Limitation des activités
RH/MB	Antécédents résidentiels
RP	Mouvements répétitifs
RS	Sécurité routière — conduite sous influence
RT	Rationalité — Contenu acheté par l'Alberta, cycle 1
SC	Soins personnels
SD	Renseignements sociodémographiques
SH	Santé sexuelle
SK	Consommation de boissons gazeuses
SL	Sommeil
SM	Usage du tabac
SP	Identificateurs d'échantillon (méthodologie)
SS	Soutien social
ST	Stress
SV	Utilisation de soins de santé
TA	Variantes du tabagisme (Enquête promotion santé 1998)
TU	Bronzage et exposition aux rayons UV
TW	Incapacité au cours des 2 dernières semaines
VS	Sécurité personnelle et violence
WF	Variable indicatrice des sous-ensembles
WT	Poids d'échantillonnage des sous-ensembles

Quelques variables d'identification importantes ne suivent pas la convention : p. ex. : REALUKEY, PERSONID, CYCLE, SUBCYCLE, DESIGPRV, STRATUM et REPLICAT.

Il existe aussi quelques variables qui sont « constantes ». Le tableau 11.B présente les variables qui n'apparaissent qu'une seule fois dans le fichier de données. Le nom de ces variables ne suit pas la convention.

**Tableau 11.B : Variables longitudinales « constantes »**

Variables longitudinales	Concept
AOI	Âge au moment de l'immigration
COB	Pays de naissance
COBC	Code pour le pays de naissance
COBGC	Pays de naissance (7 groupes) – groupé
COD10	Code pour la cause de décès (CIM-10)

Variables longitudinales	Concept
DESIGPRV	Province de résidence en 1994
DOB	Jour de naissance
DOD	Jour de décès
HWB	Poids à la naissance
HWBG1	Poids à la naissance - groupé
IMM	Indicateur d'immigration pour le répondant
MOB	Mois de naissance
MOD	Mois de décès
REPLICATE	Replicat
SEX	Sexe
STRATUM	Strate
YOB	Année de naissance
YOD	Année de décès
YOI	Année de l'immigration au Canada

### 11.3.3 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 du contenu principal répétées à chaque cycle
- F Supplément sur l'insécurité alimentaire
- 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)
- Questions thématiques spécifiques à un cycle qui ne sont pas répétées chaque fois (par ex. choix alimentaires aux cycles 3, 5 et 7)
- 3 Variables d'administration de l'enquête pour le volet 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) (par exemple les variables des poids, de l'accord de partage ou des dates d'interview).

### 11.3.4 Position 4 : Variable de l'année de référence du cycle

4	Cycle 1 (1994-1995)
6	Cycle 2 (1996-1997)
8	Cycle 3 (1998-1999)
0	Cycle 4 (2000-2001)
2	Cycle 5 (2002-2003)
A	Cycle 6 (2004-2005)
B	Cycle 7 (2006-2007)
C	Cycle 8 (2008-2009)
D	Cycle 9 (2010-2011)
E	Cycle 10 (2012-2013)

### 11.3.5 Position 5 : Type de variable

Position 5	Type de variable	Description
–	Variable recueillie	Variable qui figure directement sur le questionnaire
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables recueillies (p. ex. Système de classification des industries de l'Amérique du Nord (SCIAN))
D	Variable dérivée	Variable calculée d'après une ou plusieurs variables recueillies ou codées, ordinairement pendant le traitement au Bureau central (p. ex. Système de classification type de l'état de la santé (CHSMS-HUI3))
F	Variable indicatrice	Variable calculée par l'application informatique à partir d'une ou de plusieurs variables recueillies (comme une variable dérivée) aux fins de son utilisation ultérieure durant l'interview (p. ex. indicateur de travail). Cela peut indiquer aussi qu'une réponse élaborée a été recueillie pour cette variable (p. ex. indicateur de limitation d'activité)
G	Variable groupée	Variables recueillies, codées, supprimées ou dérivées agrégées en groupes (p. ex., groupes d'âge)
L	Variable longitudinale dérivée	Variable calculée d'après des variables provenant d'au moins deux cycles d'enquête

### 11.3.6 Position 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 « Choisissez toutes les réponses appropriées », on se sert des lettres correspondant à chaque catégorie possible de réponse : ainsi, Q1 (Choisissez toutes les réponses appropriées) 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 DHC4\_AGE au cycle 1 (1994-1995), DHC6\_AGE au cycle 2 (1996-1997), etc. 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 STC2\_C1, la première question sur le stress au travail a été nommée STC2\_W1. 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 indique dans quelle section la question figure.



## 12. Accès aux données de l'ENSP

### 12.1 Centres de données de recherche

Le souci de protéger la confidentialité des données longitudinales écarte la possibilité d'émettre un fichier de microdonnées à grande diffusion (FMGD). Cependant, l'accès aux fichiers maîtres de microdonnées longitudinales, incluant les données des cycles 1 à 7 est possible via le programme de Centres de données de recherche (CDR) de Statistique Canada. Le programme des CDR s'inscrit dans une initiative de Statistique Canada, du Conseil de recherches en sciences humaines du Canada (CRSH) et de consortiums d'universités visant à renforcer la capacité de recherche sociale du Canada et à soutenir le milieu de la recherche sur les politiques.

Les centres de données de recherche (CDR) permettent aux chercheurs d'accéder aux microdonnées d'enquêtes sur les ménages et sur la population. Les centres comptent des employés de Statistique Canada. Ils sont exploités en vertu des dispositions de la Loi sur la statistique et sont administrés conformément à toutes les règles de confidentialité. Ils ne sont accessibles qu'aux chercheurs dont les projets ont été approuvés et qui ont prêté serment en qualité de personnes réputées être employées de Statistique Canada. On trouve des CDR à l'étendue du pays. Ainsi, les chercheurs n'ont pas à se déplacer vers Ottawa pour avoir accès aux microdonnées de Statistique Canada. Vous pouvez trouver de l'information supplémentaire sur le site du Programme des centres de données de recherche à [http://www.statcan.ca/francais/rdc/index\\_f.htm](http://www.statcan.ca/francais/rdc/index_f.htm).

### 12.2 Télé-access

Les chercheurs qui n'ont pas accès à un CDR ont une autre option qui est le service de télé-access de la Division de la statistique de la santé. Ce service permet aux chercheurs de développer et de tester des programmes informatiques sur leur ordinateur en utilisant les fichiers fictifs qui sont des répliques des fichiers maîtres. Les chercheurs envoient ensuite leurs programmes à une adresse électronique. Les programmes sont tournés en utilisant les fichiers maîtres de microdonnées dans un environnement de serveur interne. Par la suite, les fichiers de sortie sont vérifiés pour la confidentialité et retournés aux chercheurs par courriel. Pour plus d'information relative au service, veuillez contacter l'équipe des Services d'information et d'accès aux données à [nphs-ensp@statcan.ca](mailto:nphs-ensp@statcan.ca).

### 12.3 Initiative de démocratisation des données

Des FMGD sont disponibles pour les trois premiers cycles de l'ENSP et procurent un accès élargi aux fichiers transversaux de l'Enquête. Les FMGD de l'ENSP sont aussi accessibles via l'Initiative de démocratisation des données (IDD) des universités et collèges participants. Pour de l'information à ce sujet, veuillez consulter le site de Statistique Canada sous [http://www.statcan.ca/francais/Dli/dli\\_f.htm](http://www.statcan.ca/francais/Dli/dli_f.htm). Les FMGD de l'ENSP des cycles 1, 2 et 3 peuvent également être achetés. Veuillez communiquer par courriel avec l'équipe des Services d'information et d'accès aux données à [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca) ou avec un des bureaux régionaux de Statistique Canada.

### 12.4 Rapports analytiques et totalisations personnalisés

Lors de la diffusion des données du cycle 5 de l'ENSP, des résultats tirés de l'enquête ont été présentés dans une publication Internet gratuite intitulée « *En santé aujourd'hui, en santé demain? Résultats de l'Enquête nationale sur la santé de la population* ». La publication

(catalogue 82-618M) présente une série d'articles basés sur les données longitudinales de l'ENSP. Ces articles traitent d'enjeux importants liés à la santé (le tabac, la santé des immigrants, l'obésité, le vieillissement). Pour consulter cette publication, utilisez le lien suivant :

<http://www.statcan.ca/bsolc/francais/bsolc?catno=82-618-M&CHROPG=1>.

Des tableaux longitudinaux Cansim (gratuits) sont aussi disponibles sur le site Web de Statistique Canada. Ils présentent les changements, d'un cycle à l'autre de l'ENSP, de l'usage du tabac, de l'auto-évaluation de la santé, de l'indice de masse corporelle et de l'activité physique. On peut accéder à ces tableaux Cansim, en consultant le lien pour la publication Internet (ci-dessus) et puis en choisissant « Tableaux de données » dans les choix offerts à gauche de la page principale.

Il est fréquent que des articles basés sur les données de l'ENSP apparaissent dans les Rapports sur la santé, une revue trimestrielle produite par la Division de la statistique de la santé. Ce produit est disponible en format imprimé sous le numéro de catalogue 82-003-XPX ou en format électronique (gratuit) dans le site de Statistique Canada sous le numéro de catalogue 82-003-XIF. Pour plus d'information, consulter les liens suivants :

<http://www.statcan.ca/bsolc/francais/bsolc?catno=82-003-X&CHROPG=1>

<http://www.statcan.ca/bsolc/francais/bsolc?catno=82-003-S&CHROPG=1>

Des totalisations personnalisées fondées sur les données transversales de l'ENSP peuvent aussi être obtenues à frais recouvrables. Aucune totalisation n'est produite avec les données longitudinales de l'ENSP. Pour plus de renseignements ou pour des estimations de faisabilité et des coûts, communiquez avec les Services d'information et d'accès aux données de la Division de la statistique de la santé à [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca).

### 13. Technique d'analyse de données d'enquêtes longitudinales

Les enquêtes longitudinales comme l'ENSP ont l'avantage d'avoir un immense potentiel analytique qui s'accroît avec le temps. Ces enquêtes permettent aux analystes de la santé d'étudier, les événements qui affectent la santé d'un individu, d'étudier leurs effets et leurs causes et de produire des taux d'incidence. L'analyse de ces données met également en lumière la complexité du réseau des relations qui existent entre la santé d'un individu et l'intensité de son exposition à divers facteurs de risque. Cet avantage permet l'étude de questions de recherche plus complexes sur la santé. En revanche, certains enjeux réduisent ce potentiel analytique et peuvent même poser un risque à long terme aux enquêtes longitudinales.

L'analyse de données d'enquêtes longitudinales fait face à des enjeux différents de ceux provenant d'enquêtes transversales. D'une part, répondre à des questions de recherche plus approfondies amène fréquemment les utilisateurs de données à utiliser des techniques d'analyse plus complexes, c'est-à-dire souvent moins connues et moins documentées. D'autre part, l'érosion des données (voir la section 9.2.4) à travers le temps est l'un des plus importants enjeux de l'analyse de données longitudinales. Ajoutée au risque d'introduire un biais dans les estimations, une diminution de la taille de l'échantillon due à l'érosion peut s'avérer problématique non seulement pour la diversité des analyses statistiques mais pour leur qualité.

De plus, l'approche analytique la plus commune et la plus simple, est d'utiliser le sous-ensemble longitudinal complet (voir sections 7.6 et 7.7). Ce sous-ensemble est plus sensible au phénomène de l'érosion dont l'effet direct est de réduire à chaque cycle le nombre d'individus dans ce sous-ensemble. Après sept cycles, l'ENSP observe un taux d'érosion cumulatif du sous-ensemble complet de 36,4 %. En d'autres termes, 63,6 % des répondants du panel original ont une réponse complète (voir section 9.2.4) aux sept premiers cycles de l'ENSP.

Le but de ce chapitre est de fournir aux utilisateurs un outil pratique pouvant réduire l'effet de l'érosion dans leurs analyses statistiques et ainsi prolonger le potentiel analytique de l'enquête. Conséquemment, atténuer l'effet de l'érosion augmentera l'exactitude des estimations et la pertinence de l'enquête.

#### 13.1 Approche « jumelage de cycles »

Le jumelage de cycles (également connue sous le nom de « regroupement des observations répétées » - "Pooling of Repeated Observations") est une approche analytique permettant de réduire l'impact de l'érosion. Elle consiste à utiliser un sous-ensemble de cycles d'un répondant comme unité d'analyse au lieu de considérer tous les cycles d'un répondant comme unité d'analyse. Cette approche est un cas particulier d'un type de modèles statistiques appelés « modèles marginaux »<sup>23</sup>.

Bien que l'exemple fictif et l'application décrits dans cette section, se limitent à deux cycles consécutifs complétés d'un même répondant comme unité d'analyse, cette approche peut s'adapter à différents scénarios. La durée d'exposition aux facteurs de risque aurait pu être plus courte ou plus longue, le nombre de cycles (ou de mesures répétées) utilisés aurait pu être 1 ou supérieur à 2, là où les variables dépendantes pourraient être continues ou discrètes avec des états temporaires ou permanents tels la mort ou certaines maladies chroniques.

<sup>23</sup> Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. New York : Wiley.

### Exemple fictif

Avec cette approche « jumelage de cycles », après sept cycles de collecte, un répondant peut donc fournir jusqu'à six unités d'analyse s'il appartient au sous-ensemble complet : cycles 1-2, cycles 2-3, cycles 3-4, cycles 4-5, cycles 5-6 et cycles 6-7 (voir répondant A dans tableau 13.A). Si un répondant n'a pas répondu aux cycles 3, 4 et 7 par exemple, mais a répondu à tous les autres, il peut produire deux unités d'analyse : cycles 1-2 et cycles 5-6 (voir répondant B du tableau 13.A). Ce dernier ne fait pas partie du sous-ensemble complet du cycle 7 (étant donné la non-réponse aux cycles 3, 4 et 7) et serait automatiquement exclu des analyses utilisant ce sous-ensemble. En revanche, en utilisant l'approche « jumelage de cycles », deux unités d'analyse proviennent de ce répondant. Il suffit qu'un répondant ait répondu à deux cycles consécutifs pour qu'il fournisse une unité d'analyse. D'après le tableau 13.A, les répondants A, B, C et D vont produire respectivement 6, 2, 4 et 1 unités d'analyse.

**Tableau 13.A : Exemples de profil de réponse des répondants de l'ENSP**

Répondant	Cycles avec une réponse complète						
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7
A	X	X	X	X	X	X	X
B	X	X			X	X	
C	X	X	X	X	X		X
D	X		X	X			X

D'après ce tableau fictif, le sous-ensemble complet du cycle 7 ne comprendrait qu'un répondant sur les quatre puisque dans ce cas l'unité d'analyse est le répondant qui a répondu à tous les cycles de l'enquête. Le taux d'érosion après sept cycles serait alors de 75 %.

Lorsque l'approche « jumelage de cycles » est utilisée, l'unité d'analyse est un enregistrement contenant deux cycles consécutifs d'un répondant. Ainsi, d'après l'exemple du tableau 13.A, le nombre potentiel d'unités d'analyse est de 24 (6 combinaisons possibles de deux cycles consécutifs multipliées par 4 répondants). Comme calculé plus haut, le nombre total d'unités d'analyse est de 12 (6+2+4+1), c'est-à-dire 50 % du nombre potentiel d'unités d'analyse et par conséquent le taux d'érosion est de 50 % comparativement à 75 % pour le sous-ensemble complet.

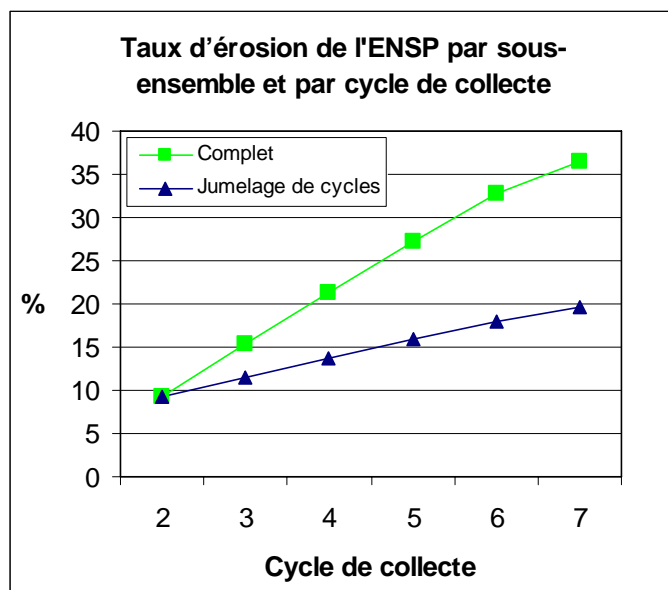
En appliquant cette approche à l'ENSP, l'échantillon potentiel (maximal) après sept cycles de collecte est de 103 656 (6 × 17 276) unités d'analyse (l'ENSP contient 17 276 répondants multipliés par 6 unités d'analyse possible après sept cycles). Après les sept cycles de collecte, l'érosion a réduit le nombre d'unités d'analyse à 83 321, c'est-à-dire près de 80 % de la taille de l'échantillon potentielle, soit un taux d'érosion de 20 %.

En revanche, le taux d'érosion du sous-ensemble complet est de 36 % après sept cycles. En prenant l'unité d'analyse de l'approche « jumelage de cycles », le taux d'érosion diminue d'environ 16 points de pourcentage par rapport à celui observé avec le sous-ensemble complet (voir tableau 13.B et figure 13.A).

**Tableau 13.B : Taux d'érosion**

Cycle	Sous-ensemble	
	Complet	Jumelage de cycles
2	9,3 %	9,3 %
3	15,3 %	11,4 %
4	21,3 %	13,7 %
5	27,3 %	15,9 %
6	32,7 %	18,0 %
7	36,4 %	19,6 %

**Figure 13.A : Taux d'érosion de l'ENSP tirés du tableau 13.B par sous-ensemble et cycle de collecte**



### 13.2 Création du sous-ensemble modifié

Avant d'être en mesure d'analyser les données avec l'approche « jumelage de cycles », il est nécessaire de modifier la structure du sous-ensemble carré de l'ENSP. Le sous-ensemble carré de l'ENSP contient 17 276 enregistrements ou unités d'analyse, où chaque enregistrement représente l'information de tous les cycles d'un seul répondant. La transformation de ce sous-ensemble consiste à modifier l'unité d'analyse afin qu'elle contienne l'information de deux cycles consécutifs d'un seul répondant. Pour le cycle 7, le nombre d'enregistrements de ce sous-ensemble modifié sera alors de 83 321 (voir section

13.1).

En prenant l'exemple du tableau 13.A, l'enregistrement du répondant A est réécrit en 6 unités d'analyse, les répondants B, C et D en 2, 4 et 1 unités d'analyse respectivement (voir tableau 13.C). Voici les étapes à suivre pour modifier le sous-ensemble carré de l'ENSP et faire l'analyse de données avec l'approche « jumelage de cycles ».

*Étape à suivre pour la création du sous-ensemble modifié de l'ENSP :*

Étape 1 : Calculer le nombre de jumelages possibles de deux cycles consécutifs pour chacun des 17 276 répondants de l'ENSP.

Étape 2 : Cloner chaque répondant de l'ENSP le même nombre de fois que le nombre de jumelages calculé à l'étape 1.

Étape 3 : Identifier les variables longitudinales d'intérêt pour les cycles jumelés.

Étape 4 : Identifier le poids d'échantillon du cycle 1, soit le poids longitudinal carré (WT64LS), ainsi que l'ensemble correspondant de poids bootstrap (B5long) (voir section 11.1 et 11.2).

Étape 5 : Conserver seulement l'information résultante des étapes 3, 4 et 5. Cette étape augmentera l'efficacité des programmes analytiques.

**Tableau 13.C : Exemple de la structure du sous-ensemble modifié pour le jumelage**

Répondant	Cycles Jumelés	Poids Cycle 1	Poids Bootstrap	Variables d'intérêt
A	1-2	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
A	2-3	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
A	3-4	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
A	4-5	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
A	5-6	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
A	6-7	$W_A$	BS1 <sub>A</sub> à BS500 <sub>A</sub>	
B	1-2	$W_B$	BS1 <sub>B</sub> à BS500 <sub>B</sub>	
B	5-6	$W_B$	BS1 <sub>B</sub> à BS500 <sub>B</sub>	
C	1-2	$W_C$	BS1 <sub>C</sub> à BS500 <sub>C</sub>	
C	2-3	$W_C$	BS1 <sub>C</sub> à BS500 <sub>C</sub>	
C	3-4	$W_C$	BS1 <sub>C</sub> à BS500 <sub>C</sub>	
C	4-5	$W_C$	BS1 <sub>C</sub> à BS500 <sub>C</sub>	
D	3-4	$W_D$	BS1 <sub>D</sub> à BS500 <sub>D</sub>	

### 13.3 Aspects méthodologiques de l'approche « jumelage de cycles »

Tel qu'expliqué à la section 10.3, le plan de sondage et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance. L'approche « jumelage de cycles » amène une complexité supplémentaire : les unités d'analyse provenant d'un même répondant sont fortement corrélées entre elles.

À l'aide des modèles marginaux (Fitzmaurice et autres (2004)), il suffit de considérer les unités d'analyse de l'approche « jumelage de cycles » comme un degré supplémentaire au plan d'échantillonnage de l'ENSP afin de tenir compte de la dépendance entre les unités d'analyse dans le calcul de la variance. La corrélation entre les unités d'analyse d'un même répondant se retrouve alors intégrée à l'effet de plan de l'échantillon. En créant un sous-ensemble modifié tel que décrit précédemment et en utilisant la méthode bootstrap pour son calcul de la variance, l'analyste s'assure de calculer la variance correctement.

Étant donné cette forte corrélation entre les unités d'analyse d'un même répondant, le choix de la méthode de calcul de la variance est plus important. L'utilisation d'une des méthodes décrites à la section 10.3, comme la méthode bootstrap (en SAS ou SPSS) ou certains logiciels comme STATA ou SUDAAN, est fortement recommandée.

### 13.4 Un exemple d'utilisation de l'approche « jumelage de cycles » : l'abandon du tabac

L'abandon du tabac est une des étapes les plus importantes que les fumeurs puissent franchir pour améliorer leur santé. Ainsi, comprendre les facteurs associés à l'abandon du tabac est important pour les programmes de santé publique qui visent à réduire le taux de tabagisme. L'analyste, Margot Shields (2005)<sup>24</sup>, a utilisé l'ENSP pour étudier les facteurs associés à l'abandon du tabac des personnes âgées de 18 ans et plus qui étaient des fumeurs quotidiens.

Les cycles 1 à 5 de l'ENSP ont été utilisés dans cette analyse. L'ENSP suit les mêmes personnes au fil du temps en les interviewant tous les deux ans. Un fichier d'analyse a été créé pour examiner successivement les cycles par pair. La population cible d'intérêt était tous les répondants de 18 ans et plus qui ont déclaré être des fumeurs quotidiens. Les fumeurs quotidiens étaient considérés comme ayant abandonné le tabac si, au cycle suivant, ils déclaraient ne pas avoir fumé du tout durant les deux dernières années.

Le fichier d'analyse a été créé pour identifier les fumeurs quotidiens de 18 ans et plus aux quatre cycles (1, 2, 3 et 4) et déterminer s'ils avaient abandonné le tabac au cycle suivant deux ans plus tard (c'est-à-dire, aux cycles 2, 3, 4 et 5 respectivement). Pour qu'un enregistrement fasse partie du fichier d'analyse, deux conditions devaient être satisfaites :

- la personne devait être un fumeur quotidien de 18 ans et plus au cycle de base<sup>25</sup>;
- la consommation de tabac devait être connue au cycle suivant deux ans plus tard (c'est-à-dire, on savait si la personne fumait encore ou avait abandonné le tabac).

Ainsi, le fichier d'analyse contenait des personnes ayant contribué à plusieurs enregistrements puisqu'elles pouvaient avoir déclaré être des fumeurs quotidiens à plus

---

<sup>24</sup> M. Shields, (2005), Cheminement vers l'abandon du tabac, Rapport sur la santé, Vol. 16, No. 3. Statistique Canada, Catalogue 82-003.

<sup>25</sup> Dans une paire de cycles jumelés, le cycle de base correspond au premier cycle. Par exemple pour les cycles jumelés 3 et 4, le cycle de base est le cycle 3.

d'un cycle d'enquête.

L'exemple fictif ci-dessous va illustrer ce qui précède. Supposons que seulement deux facteurs de risque soient étudiés pour modéliser l'abandon du tabac : le nombre de cigarettes fumées quotidiennement ( $x$ ) et l'âge d'initiation au tabac ( $z$ ). Soit,  $y$  une variable binaire qui prend la valeur « 1 » si un fumeur quotidien devient non-fumeur au cycle subséquent deux ans plus tard, autrement  $y$  prend la valeur de « 0 ». L'information de  $x$ ,  $y$  et  $z$  est recueillie à chaque cycle.

Soit  $p$  la probabilité que  $y$  prenne la valeur de « 1 », c'est-à-dire que la personne ait abandonné le tabac. Supposons que le modèle d'intérêt soit un modèle logistique :

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 x + \alpha_2 z,$$

et que l'objectif de l'étude est d'estimer  $\alpha_0$ ,  $\alpha_1$  et  $\alpha_2$  ainsi que les écarts-types de leurs estimés. Supposons que seulement les quatre premiers cycles de l'ENSP soient utilisés pour cette analyse et qu'il n'y ait que quatre personnes dans le fichier l'ENSP (tous des hommes). Par conséquent, on obtient le fichier d'analyse suivant :

Cycle	Identification de la personne	x	Z	Y <sup>26</sup>
1	1	$x_{11}$	$z_{11}$	0
1	2	$x_{21}$	$z_{21}$	1
1	3	$x_{31}$	$z_{31}$	1
1	4	$x_{41}$	$z_{41}$	0
2	1	$x_{12}$	$z_{12}$	0
2	4	$x_{42}$	$z_{42}$	0
3	1	$x_{13}$	$z_{13}$	0
3	3	$x_{33}$	$z_{33}$	1
3	4	$x_{43}$	$z_{43}$	1

- La personne 1, qui était un fumeur quotidien aux quatre cycles contribue à trois enregistrements dans le fichier d'analyse puisqu'il avait 18 ans ou plus et était un fumeur quotidien aux trois cycles de base (1, 2 et 3).

<sup>26</sup>

Veuillez noter que cette variable est dérivée du cycle subséquent deux ans plus tard.



- La personne 2 ne contribue que pour un enregistrement : il était un fumeur quotidien au cycle 1, avait abandonné le tabac au cycle 2 et n'a pas répondu à l'enquête aux cycles 3 et 4.
- La personne 3 contribue à deux enregistrements : il était un fumeur quotidien au cycle 1, avait abandonné le tabac au cycle 2, avait repris l'usage du tabac au cycle 3 et avait abandonné le tabac à nouveau au cycle 4.
- La personne 4 contribue à trois enregistrements : il était un fumeur quotidien aux trois premiers cycles et avait abandonné le tabac au cycle 4.

Le profil de réponse peut se résumer ainsi :

Identification de la personne	Cycle 1	Cycle 2	Cycle 3
1	X	X	X
2	X		
3	X		X
4	X	X	X

Supposons que pour l'analyse, un logiciel de régression logistique soit utilisé et requière un fichier de données qui inclut les poids d'enquête ainsi que les poids bootstrap. Les poids d'enquête et les poids bootstrap du cycle 1 sont utilisés. Le fichier d'analyse doit être fusionné avec les poids bootstrap du cycle 1 et devrait avoir la structure suivante :

Enregistrement	Personne	Wt	X	Z	Y	bs1	bs2	---
1	1	$w_1$	$x_{11}$	$z_{11}$	0	$bs1_1$	$bs2_1$	---
2	1	$w_1$	$x_{12}$	$z_{12}$	0	$bs1_1$	$bs2_1$	---
3	1	$w_1$	$x_{13}$	$z_{13}$	0	$bs1_1$	$bs2_1$	---
4	2	$w_2$	$x_{21}$	$z_{21}$	1	$bs1_2$	$bs2_2$	---
5	3	$w_3$	$x_{31}$	$z_{31}$	1	$bs1_3$	$bs2_3$	---
6	3	$w_3$	$x_{33}$	$z_{33}$	1	$bs1_3$	$bs2_3$	---
7	4	$w_4$	$x_{41}$	$z_{41}$	0	$bs1_4$	$bs2_4$	---
8	4	$w_4$	$x_{42}$	$z_{42}$	0	$bs1_4$	$bs2_4$	---
9	4	$w_4$	$x_{43}$	$z_{43}$	1	$bs1_4$	$bs2_4$	---

Où  $w_i$  est la variable du poids d'enquête pour la personne  $i$  et  $bsj_i$  est le  $j^{\text{ième}}$  poids bootstrap correspondant de la même personne. Veuillez prendre note que les poids d'enquête et les poids bootstrap sont les mêmes pour les observations multiples d'une même personne.

Cette approche utilise les mêmes informations (poids d'enquête et poids bootstrap) et le même logiciel qui serait utilisé si le modèle d'intérêt était adapté à une seule observation par personne. Cette approche nécessite simplement que le poids pour une personne et les poids bootstrap soient assignés à de multiples observations pour cette personne.

La question est de savoir quel poids d'enquête et quels poids bootstrap sont les plus appropriés, mais les variables de poids du premier cycle d'enquête sont souvent le meilleur choix. Ceci permet à l'analyste de maximiser la taille de l'échantillon en utilisant de l'information partielle lorsqu'une personne ne répond pas à tous les cycles d'enquête.

Bien que des complexités supplémentaires soient associées aux réponses partielles des personnes, ce sujet n'est pas traité dans ce document.

**Annexe A : ENSP, volet ménages, changements au questionnaire pour le cycle 7 (2006-2007)**

**1. Introduction**

Cette annexe décrit les changements apportés au questionnaire entre le cycle 6 (2004-2005) et le cycle 7 (2006-2007). Certaines questions posées au cycle 6 ne l'étaient plus au cycle 7 et d'autres questions non posées au cycle 6 ont été ajoutées au cycle 7. Les changements entre les deux cycles sont décrits en détail au point 3. Globalement, mentionnons que les principaux ajouts au cycle 7 ont été faits au module sur la nutrition (on répète les 12 questions thématiques sur les choix alimentaires du cycle 5) ainsi qu'au module sur le stress (on répète les 7 questions thématiques sur les facteurs de stress chez les enfants et les adultes (traumatismes) du cycle 4) et on répète aussi les 3 questions thématiques du module sur insécurité alimentaire du cycle 3. Seules les 16 questions thématiques sur la capacité de faire face au stress du module sur le stress ont été enlevées au cycle 7.

**2. Changements à la structure du questionnaire**

En faisant abstraction des questions ajoutées ou retirées au cycle 7, l'ordre du questionnaire est le même qu'au cycle 6 sauf pour une seule question sur l'usage du tabac qui a été déplacée.

**3. Changements au contenu principal de l'enquête**

Pour ce qui suit, les noms de variables externes du questionnaire sont utilisés. Quelques-uns des noms internes peuvent avoir été modifiés pour assurer la cohérence dans le questionnaire, ceci est causé par la suppression ou l'addition de questions. Les noms des variables du fichier maître et du fichier partagé suivent la convention utilisée pour nommer les variables.

Parmi le contenu thématique, le sommeil et l'image corporelle (Cycle 5) font dorénavant partie du contenu principal.

**Sections pour lesquelles aucune modification n'a été apportée**

- Variables du dossier ménage
- État de santé général
- Taille et poids (HW)
- Santé préventive (PH)
- Utilisation des soins de santé (HC)
  - Soins de santé à domicile
- Limitation des activités (RA)
- Problèmes de santé chroniques (CC)
  - Allergies alimentaires
  - Autres allergies
  - Asthmes
  - Fibromyalgie
  - Arthrite ou rhumatisme, excepté la fibromyalgie
  - Maux de dos

- Hypertension
- Migraines
- Bronchite chronique ou emphysème
- Diabète
- Épilepsie
- Maladie cardiaque
- Cancer
- Ulcères à l'intestin ou à l'estomac
- Troubles dus à un accident vasculaire cérébral
- Incontinence urinaire
- Troubles intestinaux
- La maladie d'Alzheimer ou autre forme de démence cérébrale
- Cataractes
- Glaucome
- Problèmes de la thyroïde
- Autre problème de santé chronique
- État de santé (HS)
  - Vision
  - Ouïe
  - Élocution
  - Mobilité
  - Mains et doigts
  - Sentiments
  - Mémoires
  - Pensée
  - Douleurs et malaises
- Activités physiques (PA)
- Exposition aux rayons ultraviolets (UV) (TU)
- Mouvements répétitifs (RP)
- Blessures (IJ)
- Stress (ST)
  - Problèmes actuels
  - Stress au travail
  - Contrôle
- Consommation de médicaments (DG)
- Usage du tabac (SM)
- Santé mentale (MH)
- Soutien social (SS)
- Langue (SD)
- Niveau de scolarité (ED)
- Population active (LF)
  - Lien au travail
  - Recherche de travail – dernières 4 semaines
  - Lien à l'emploi précédent
  - Description de l'emploi
  - Absences/Heures
  - Autre emploi
  - Semaines employées

- Recherche de travail
- Revenu (IN)
- Numéro provincial d'assurance-maladie et Administration (AM)
  - Numéro provincial d'assurance-maladie
  - Administration

### **Sections pour lesquelles des modifications ont été apportées**

#### **Nutrition (NU)**

- Choix alimentaire
  - Toutes les questions (NU\_B\_1A, NU\_B\_1C à NU\_B\_1E, NU\_B\_2A à NU\_B\_2C, NU\_B\_3A à NU\_B\_3D et NU\_B\_3G) de cette sous-section thématique du cycle 5 sont répétées au cycle 7 afin de pouvoir comparer les données entre les cycles.

#### **Stress (ST)**

- Facteurs de stress chez les enfants et les adultes (« traumatismes »)
  - Toutes les questions (ST\_B\_T1 à ST\_B\_T7) de cette sous-section thématique du cycle 4 sont répétées afin d'avoir les données pour les répondants qui étaient âgés de 18 ans au cycle 7.
- Stratégies pour faire face au stress
  - Toutes les questions (CO\_A\_1 à CO\_A\_16) de cette sous-section thématique du cycle 6 ont été enlevées au cycle 7.

#### **Insécurité alimentaire (FI)**

- Toutes les questions (FI\_B\_1 à FI\_B\_3) de cette sous-section thématique du cycle 3 sont répétées au cycle 7 afin de pouvoir comparer les données entre les cycles.

#### **4. Achat de contenu**

Aucun achat de contenu au cycle 7.

**Annexe B : ENSP volet ménages, exemples de variables provenant de l'interview précédente utilisées comme information complémentaire au cycle 7 (2006-2007)**

Hypertension, mammographie, test PAP	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.). Lors du cycle 3, les questions ont été répétées. 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. Lors des cycles 4, 5, 6 et 7, dans le cas où le répondant avait déclaré avoir eu un test lors d'une interview antérieure, seule la question « À quand remontait la dernière fois? » a été posée.
Limitation des activités	On s'est servi des données indiquant si le répondant souffrait ou non d'une incapacité lors de l'interview précédente. 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 de l'interview précédente concernant certains problèmes de santé chroniques (asthme, fibromyalgie, arthrite, hypertension, migraine, diabète, épilepsie, ulcères à l'estomac ou à l'intestin et les séquelles d'un accident cérébral vasculaire) pour essayer d'obtenir des explications concernant le changement. Si le problème était nouveau, on a saisi la date de sa manifestation.
Usage du tabac	Si, lors de la dernière interview, un fumeur quotidien a fourni l'âge auquel il/elle a commencé à fumer tous les jours, cette réponse a été utilisée au cycle 7. Un fumeur occasionnel ou non-fumeur au cycle 7 qui a déclaré fumer (ou avoir déjà fumé) tous les jours lors de l'interview précédente, on a utilisé un indicateur d'usage quotidien. En cas de changement de situation, on a posé des questions pour obtenir une explication.
Renseignements sociodémographiques	Pour chaque répondant, on a utilisé un indicateur montrant que les renseignements concernant le pays de naissance ont été recueillis. 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 précédente. 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.
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 ou que le répondant avait donné un numéro différent de celui enregistré lors du dernier cycle, on a demandé au répondant de donner son numéro d'assurance-maladie à nouveau.