

**Programme d'enquêtes sur la santé de la population**

**Enquête nationale sur la santé de la population**

**Cycle 4 (2000-2001), volet ménages**

**Documentation longitudinale**

**Mai 2002**

### **NOTE AUX UTILISATEURS**

Depuis la collecte du cycle 4 en 2000-2001, l'Enquête nationale sur la santé de la population (ENSP) est de nature strictement longitudinale.

Afin de donner une plus grande flexibilité aux utilisateurs, un seul fichier maître de microdonnées a été créé pour le cycle 4 de l'ENSP. Ce fichier contient les 17 276 membres du panel longitudinal de l'ENSP peu importe le type de réponse obtenu aux cycles 1 à 4. Dans ce fichier maître quatre sous-ensembles de répondants ont été créés avec les poids de sondage correspondants ainsi que des variables indicatrices rendant leur identification plus facile.

La documentation longitudinale du cycle 4 de l'ENSP fournit un large éventail d'information sur l'enquête : objectifs, développement de contenu, plan d'échantillonnage, collecte, traitement, qualité, pondération, totalisation et accès aux données. Les chapitres 7 et 9 donnent plus de détails concernant les différents sous-ensembles de répondants et le poids qui leur est associé

Ce guide s'adresse également aux utilisateurs du fichier partagé de microdonnées. Le fichier partagé inclut les répondants qui ont accepté de partager leurs données au cycle 4 de l'ENSP de même que le poids d'échantillonnage correspondant. Ce groupe de répondants est un sous-ensemble du fichier maître de microdonnées. Les utilisateurs du fichier partagé ne doivent pas tenir compte des références qui sont faites concernant les autres sous-ensembles de répondants.

Finalement, ce document fait référence aux différents cycles de l'ENSP en utilisant parfois les années du cycle en question. À titre de référence 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

## TABLE DES MATIÈRES

	Page
1. Introduction.....	6
2. Contexte.....	7
3. Objectifs.....	8
4. Contenu de l'enquête.....	9
4.1 Critères de sélection.....	9
4.2 Modifications apportées au contenu du cycle 4 (2000-2001).....	10
4.3 Rétroaction des données du cycle 3 (1998-1999) et questions de suivi.....	11
5. Plan d'échantillonnage.....	12
5.1 Plan d'échantillonnage pour le cycle 1 (1994-1995).....	12
5.1.1 Répartition de l'échantillon.....	12
5.1.2 Méthode d'exclusion.....	13
5.1.3 Prélèvement de l'échantillon.....	13
5.1.4 Plan d'échantillonnage au Québec.....	15
5.2 Échantillon longitudinal.....	16
6. Collecte des données.....	18
6.1 Conception du questionnaire et méthode de collecte des données.....	18
6.2 Essais.....	18
6.3 Interviews.....	18
6.4 Non-réponse et dépistage.....	19
7. Traitement des données.....	21
7.1 Vérification.....	21
7.2 Codage.....	21
7.3 Création de variables dérivées et groupées.....	21
7.4 Estimation et pondération.....	22
7.5 Sous-ensembles de répondants.....	22
7.6 Définition d'une réponse complète.....	23
8. Taux de réponse et taux d'érosion.....	24
8.1 Taux de réponse du cycle 1 (1994-1995).....	24
8.2 Taux de réponse du cycle 2 (1996-1997).....	25
8.3 Taux de réponse du cycle 3 (1998-1999).....	26
8.4 Taux de réponse du cycle 4 (2000-2001).....	27
8.5 Taux d'érosion.....	28
8.6 Taux de non-réponse attribuables aux personnes non-dépistées.....	29

9.	Pondération .....	30
9.1	Pondération longitudinale.....	30
9.1.1	Poids épurés du cycle 1 (1994-1995).....	31
9.1.2	Poids longitudinaux.....	31
9.1.2.1	Poids longitudinal carré (WT64LS).....	31
9.1.2.2	Poids longitudinal complet (WT60LF).....	31
9.1.2.3	Poids longitudinal complet cycles 1 et 4 (WT60LFE) .....	36
9.1.2.4	Poids longitudinal complet partagé (WT60SLF).....	38
10.	Qualité des données .....	39
10.1	Erreurs dans l'enquête .....	39
10.1.1	Erreurs non dues à l'échantillonnage .....	39
10.1.2	Erreurs dues à l'échantillonnage .....	40
10.1.2.1	Méthode bootstrap pour l'estimation de la variance .....	40
10.1.2.2	Programme Bootvar pour l'estimation de la variance .....	41
11.	Lignes directrices sur les totalisations, l'analyse et la diffusion .....	42
11.1	Lignes directrices sur l'arrondissement.....	42
11.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation.....	43
11.2.1	Définitions des genres d'estimations : de type nominal par opposition à quantitatif .....	43
11.2.2	Totalisation d'estimations de type nominal .....	44
11.2.3	Totalisation d'estimations quantitatives .....	44
11.3	Lignes directrices pour l'analyse statistique .....	44
11.4	Lignes directrices pour la diffusion.....	45
12.	Utilisation du fichier maître longitudinal.....	47
12.1	Utilisation des poids longitudinaux.....	47
12.2	Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap.....	47
12.3	Convention utilisée pour nommer les variables .....	48
12.3.1	Structure élémentaire des noms des variables.....	48
12.3.2	Positions 1 et 2 : Nom de la variable / section du questionnaire.....	49
12.3.3	Position 3 : Type d'enquête.....	50
12.3.4	Position 4 : Variable de l'année de référence du cycle .....	51
12.3.5	Position 5 : Type de variable.....	51
12.3.6	Position 6 à 8 : Nom de la variable .....	51
13.	Accès aux données de l'ENSP.....	53
13.1	Fichiers de microdonnées.....	53
13.2	Rapports analytiques et totalisations .....	53

### **Liste des tableaux**

Tableau 5.A	Taille de l'échantillon longitudinal par province
Tableau 7.A	Sous-ensembles de répondants
Tableau 8.A	Dépistage, cycles 2, 3 et 4
Tableau 9.A	Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes.
Tableau 9.B	Variables pour la correction de la non-réponse au cycle 2 - WT60LF
Tableau 9.C	Variables pour la correction de la non-réponse au cycle 3 - WT60LF
Tableau 9.D	Variables pour la correction de la non-réponse au cycle 4 - WT60LF
Tableau 11.A	Lignes directrices relatives à la variabilité d'échantillonnage
Tableau 12.A	Sous-ensembles des répondants et fichiers de poids bootstrap correspondants
Tableau 12.B	Variables longitudinales

### **Liste des annexes**

Annexe A :	ENSP, volet ménages, Contenu de la composante Santé au cours des 4 cycles
Annexe B :	ENSP, volet ménages, Changements au questionnaire pour le cycle 4 (2000-2001)
Annexe C :	ENSP volet ménages, Exemples de rétroaction des données du cycle 3 (1998-1999) et de questions de suivi

### **Autres documents de référence**

Questionnaire
Cliché d'enregistrements
Index alphabétique
Index par sujet
Dictionnaire des données
Documentation des variables dérivées, ENSP, cycles 1 à 4

## **1. Introduction**

L'Enquête nationale sur la santé de la population (ENSP) vise à recueillir des renseignements sur la santé de la population canadienne ainsi que des renseignements socio-démographiques connexes. Le premier cycle de l'enquête a d'abord eu lieu en 1994-1995 et elle aura lieu tous les deux ans pendant une période de 20 ans. 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. Des renseignements de nature transversale sont désormais recueillis grâce à l'Enquête sur la santé dans les collectivités canadiennes (ESCC). L'ENSP compte maintenant deux volets : l'enquête auprès des ménages et l'enquête auprès des établissements de santé. À partir du cycle 4, l'enquête dans le Nord sera uniquement transversale et sera menée par l'ESCC plutôt que l'ENSP.

Le volet ménages de l'ENSP vise les membres des ménages de toutes les provinces, à l'exclusion de la population des réserves indiennes et terres de la Couronne, des résidents des établissements de soins de santé, des membres à plein temps des Forces armées canadiennes et de certaines régions éloignées du Québec et de l'Ontario. Le volet Établissements de soins de santé inclut les résidents séjournant à long terme dans des établissements de soins de santé (séjours de plus de six mois) comportant au moins quatre lits dans toutes les provinces à l'exclusion des territoires et des membres à plein temps des Forces armées canadiennes. Quatre cycles de collecte sont complétés pour le volet Ménages de l'ENSP : l'ENSP cycle 1 (1994-1995), l'ENSP cycle 2 (1996-1997), l'ENSP cycle 3 (1998-1999) et l'ENSP cycle 4 (2000-2001).

Le volet ménages du cycle 4 de l'ENSP a recueilli des renseignements détaillés sur la santé des répondants longitudinaux sélectionnés au hasard lors du premier cycle de même que des renseignements démographiques sur tous les membres du ménage des répondants longitudinaux.

Le questionnaire détaillé sur la santé 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 tels que l'âge, le sexe, le niveau de scolarité, l'appartenance ethnique, le revenu du ménage et l'activité sur le marché du travail.

Le présent document a pour but de faciliter la manipulation du fichier maître longitudinal du cycle 4, volet ménages contenant les résultats de l'ENSP. Ce fichier est décrit de façon plus détaillée dans les chapitres suivants. Pour toute question au sujet des données ou de leur utilisation, veuillez vous adresser à :

Renseignements sur le contenu de l'enquête ou le télé-accès :

Section ENSP, Division de la statistique de la santé      Téléphone : 1-613-951-1653  
Télécopieur : 1-613-951-4198  
Courriel : [nphs-ensp@statcan.ca](mailto:nphs-ensp@statcan.ca)

Totalisations personnalisées ou renseignements généraux sur les données :

Services personnalisés à la clientèle  
Division de la statistique de la santé      Téléphone : 1-613-951-1746  
Courriel : [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca)

## **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 brosser 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.

### 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,
- permettre de coupler les données de l'enquête à des données administratives courantes, telles que les statistiques de l'état civil, les mesures environnementales, les variables liées aux collectivités ainsi que l'utilisation des services de santé.



## **4. Contenu de l'enquête**

Les objectifs décrits au chapitre 3 représentent l'orientation générale de l'ENSP, notamment en ce qui concerne le type de renseignements à recueillir. La première section de ce chapitre présente les critères généraux utilisés pour choisir le contenu de l'enquête et résume le contenu des diverses sections et changements. La deuxième section décrit en détails les modifications apportées au contenu de l'enquête pour réaliser le cycle 4 de l'ENSP ainsi que son nouveau contenu. La dernière section présente les variables du cycle 3 qui ont été utilisées lors de la collecte du cycle 4. De l'information détaillée additionnelle sur le contenu de l'enquête de 2000-2001 et sur les changements de contenu durant les quatre cycles est disponible dans les annexes A, B et C.

### **4.1 Critères de sélection**

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

- 1) Les renseignements doivent être en rapport avec les objectifs des provinces et des territoires en matière de santé et faciliter le contrôle de ces objectifs. Dans les cas où ces objectifs n'existent pas, par exemple au niveau national, on peut déterminer le contenu de l'enquête à partir des politiques et des programmes,
- 2) Les renseignements ne doivent pas faire double emploi avec ceux provenant d'autres sources,
- 3) Pour mieux faire comprendre la santé et ses déterminants, l'enquête doit fournir de nouveaux renseignements dans des domaines qui n'ont pas encore été étudiés de manière approfondie,
- 4) L'enquête doit porter principalement sur les comportements ou sur les problèmes de santé qui se prêtent à la prévention, au traitement ou à l'intervention,
- 5) L'enquête doit permettre de recueillir des renseignements sur les problèmes de santé qui causent le plus de souffrance ou qui coûtent le plus cher aux personnes touchées, à l'ensemble de la population ou aux régimes de soins santé,
- 6) L'enquête doit fournir des renseignements sur les facteurs liés non seulement à la maladie mais aussi à la bonne santé.

Lors des trois premiers cycles un certain nombre de renseignements ont été recueillis auprès de tous les membres de chaque ménage (composante Générale, H05, par exemple les caractéristiques socio-démographiques et les renseignements sur l'activité relative au marché du travail). De plus, une personne avait été choisie au hasard dans chaque ménage pour une interview plus approfondie (composante Santé, H06, par exemple des renseignements sur l'état de santé) : le répondant longitudinal. À partir du cycle 4, seul le répondant longitudinal fournit tous les renseignements provenant des questionnaires des composantes Générale et Santé.

Selon les critères susmentionnés, le questionnaire 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 a évalué l'état de santé à l'aide de questions sur la perception personnelle de la santé, la capacité fonctionnelle, les problèmes de santé chroniques et la limitation des activités. L'utilisation des services de santé a été mesurée au moyen de questions sur les visites faites 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 avaient trait à l'usage du tabac, la consommation d'alcool et l'activité physique. Il y avait en outre des questions sur les examens préventifs qui demandaient la fréquence et les raisons d'utilisation de ces services. Les questions sur les caractéristiques démographiques et socio-économiques couvraient l'âge, le sexe, le niveau de scolarité, l'appartenance ethnique, la race, le revenu du ménage et l'activité sur le marché du travail.

Une copie du questionnaire du cycle 4 (2000-2001) de l'ENSP volet ménages est fournie avec la documentation de même que sur le site Web de Statistique Canada « [www.statcan.ca](http://www.statcan.ca) », veuillez sélectionner «Produits et services», « publications gratuites », « Santé », et ensuite sélectionnez la catégorie «Enquête nationale sur la santé de la population». Veuillez consulter le chapitre 13 pour d'autres détails sur l'information disponible concernant le site Web de Statistique Canada.

#### **4.2 Modifications apportées au contenu du cycle 4 (2000-2001)**

Avec l'apparition de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), il y a eu plusieurs changements au questionnaire de l'ENSP- volet Ménages. Depuis que l'ENSP est devenue strictement longitudinale, certains éléments du contenu ont été déplacés dans l'ESCC (notamment la section sur l'incapacité au cours des deux dernières semaines et certaines questions sur l'endroit où ont été fournis les soins) d'autres ont été éliminés (par ex. certaines conditions chroniques) et l'ordre du questionnaire a changé. Les questions de suivi ont été modifiées et certaines questions de nature constante (par ex. le pays de naissance) ont été éliminées. Il y a eu des améliorations au contenu et au libellé.

Comme les questions ne sont posées dorénavant qu'au répondant longitudinal, il n'est plus nécessaire de distinguer la composante « Générale » (caractéristiques socio-démographiques et certaines questions sur la santé pour tous les membres des ménages) de la composante « Santé » (questionnaire plus spécifique sur la santé du répondant longitudinal). Ces deux ensembles de questions ont donc été combinés en un seul questionnaire. Des renseignements démographiques sur tous les membres du ménage (âge, sexe, liens de parenté entre les membres) sont toujours recueillis.

L'ordre du questionnaire a changé de manière à refléter l'ordre de la composante « Générale » des cycles précédents (et l'ordre d'autres enquêtes sociales comme l'Enquête sociale générale) : les données sur la santé sont en premier, suivies des données socio-économiques (langue, appartenance ethnique, race, niveau de scolarité, état de la population active et revenu) qui sont maintenant à la fin de l'enquête. Il n'y a pas eu de contenu acheté pour ce cycle.

Comme pour chaque cycle il y avait un contenu thématique. Le thème principal du cycle 4 portait sur le stress. Les questions thématiques du cycle 1 ont été reprises au cycle 4. Elle couvrait les problèmes actuels, les événements récents, les facteurs de stress chez les enfants et les adultes, le stress au travail, l'estime et le contrôle de soi. Une nouvelle section sur le stress perçu a été ajoutée à la section sur l'état de santé général. Des questions détaillées sur l'arthrite, le diabète et les maladies cardiaques ont aussi été ajoutées au contenu thématique.

La section sur les blessures a été modifiée de façon à inclure du nouveau contenu et à normaliser les concepts selon ceux de la CIM-10 (Classification statistique internationale des maladies et des problèmes de santé connexes. 10<sup>ième</sup> révision). La section sur la Population active comporte aussi des changements importants visant à ajouter du nouveau contenu et à normaliser les concepts selon ceux de l'Enquête sur la population active.

Pour plus d'information, voir les annexes A et B.

### **4.3 Rétroaction des données du cycle 3 (1998-1999) et questions de suivi**

Afin de réduire le fardeau de réponse, les questions posées lors des cycles précédents et dont la réponse n'a pas changé dans le temps (le pays de naissance par exemple) n'ont pas été posées à nouveau. 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. Certaines réponses recueillies lors des cycles précédents ont été rappelées au répondant lors de l'interview du cycle 4. Cette technique de rappel a permis d'améliorer la qualité de la collecte. Par exemple, les renseignements sur certaines conditions chroniques obtenus lors de l'interview du cycle 3 ont été utilisés pour aider à expliquer un changement au cycle 4. Voir l'annexe C pour plus de détails.

## **5. Plan d'échantillonnage**

La population cible du volet ménages de l'ENSP vise les membres des ménages de toutes les 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 Forces armées canadiennes et 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 (17 276 personnes) de l'ENSP a été formé.

### **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'on a choisi un échantillon 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/territoriaux,
- la décision de choisir un membre par ménage pour constituer le panel longitudinal et
- le choix de l'EPA remaniée 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. 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>1</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/le territoire  $h$ ,  $h=1, \dots, 12$ , d'après le Recensement de 1991. Cette répartition a permis de déterminer la taille de l'échantillon principal de chaque province. Quatre provinces ont choisi d'acheter des unités d'échantillonnage additionnelles pour augmenter l'échantillon 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.

---

<sup>1</sup> Kish, L. (1988). Multipurpose Sample Design, Survey Methodology, 14, 19-32

### **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 personnes célibataires et les personnes âgées, seraient sur-représentées.

Par conséquent, on a adopté une méthode d'exclusion pour accroître la représentativité des parents et des jeunes dans le panel. Une partie de l'échantillon a été soumise à une sélection préliminaire. Une fois la liste des membres 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 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 des listes de ménages pour chaque grappe et à sélectionner des logements ou ménages à partir de ces listes.

Dans chaque province, sauf au Québec, l'ENSP a fait appel à la méthode d'échantillonnage polyvalente mise au point pour le remaniement de l'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 quelque 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-réponse 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 de un à six renouvellements par province, un renouvellement correspondant au sixième de l'échantillon total. On procède à la stabilisation de l'échantillon pour le maintenir au niveau voulu, notamment lorsqu'on réserve deux renouvellements mais que la taille de l'échantillon requise représente seulement 1,5 renouvellements.

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 infra-provinciale 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.

Par suite de ces modifications, l'échantillon de grappes de l'ENSP peut être considéré comme un échantillon stratifié répété dont les strates sont constituées de groupes de strates de l'EPA et dont les échantillons répétés sont habituellement des échantillons indépendants, répartis de manière identique et composés de quatre grappes chacun. Les quelques exceptions à cette règle ne devraient pas avoir d'incidence appréciable sur les résultats de l'enquête. Deux variables de plan, « stratum » et « replicat » se trouvent sur 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 catégories de densité urbaine (la région métropolitaine de recensement de Montréal, les capitales régionales, les petites agglomérations urbaines et le secteur rural). Dans chaque secteur, les grappes ont été stratifiées selon des caractéristiques socio-économiques et sélectionnées à l'aide d'une méthode d'échantillonnage PPT. On a dénombré les grappes sélectionnées et on a prélevé des échantillons aléatoires de logements parmi celles-ci, à raison de 10 logements par grappe dans les grandes villes, et de 20 ou 30 logements ailleurs.

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

La taille de l'échantillon de l'ENSP a d'abord été répartie parmi les quatre catégories de densité urbaine. Afin d'éviter la concentration de l'échantillon à Montréal, la répartition a été proportionnelle à  $\sqrt{(2W_h^2 + 1/4^2)}$ , où  $W_h$  représente la part de la population pour la catégorie  $h$ ,  $h=1,2,3,4$ . Dans chaque catégorie, on a tenté d'obtenir un sous-échantillon de l'ESS qui serait, en ce qui concerne le membre sélectionné pour faire partie du panel, proportionnel aux populations des quatre types de ménages. Pour ce faire, on a tiré de l'ESS un nombre suffisant de ménages pour donner le rendement voulu de ménages ayant des enfants (le groupe le plus sous-représenté), après quoi on a supprimé la part excédentaire des trois autres groupes de ménages. On a donc prélevé un échantillon initial supérieur de près de 50 % à l'échantillon requis. Après 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 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. C'est cet échantillon longitudinal qui a été suivi aux cycles 2, 3 et 4 et qui sera suivi au cours des cycles à venir 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 ne se renouvelle pas au fil du temps. Aucun membres 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 4, les membres du panel longitudinal étaient tous âgés d'au moins six 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 les non-répondants, les refus et les personnes qui n'ont pu être retracées. Malgré l'érosion, l'échantillon longitudinal est toujours représentatif de la population de 1994-1995. L'érosion étant relativement faible (voir section 8.5), 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 santé font toujours partie de l'échantillon longitudinal et sont considérés comme étant répondants. 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 aux quatre cycles de l'ENSP (ce compte comprend les personnes décédées).



**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, 2, 3 et 4</b>
Terre-Neuve	1 082	905
Île-du-Prince-Édouard	1 037	856
Nouvelle-Écosse	1 085	843
Nouveau-Brunswick	1 125	893
Québec	3 000	2 393
Ontario	4 307	3 236
Manitoba	1 205	996
Saskatchewan	1 168	985
Alberta	1 544	1 189
Colombie-Britannique	1 723	1 286
<b>Total</b>	<b>17 276</b>	<b>13 582</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 pré-insérer 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'insérer le nom du répondant ou de faire les accords en genre à l'intérieur même des questions. Il est également possible de programmer des intervalles ou des réponses acceptables d'après les données recueillies durant l'interview. En d'autres termes, le questionnaire 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. De plus, à chaque cycle, 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 faces 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 (i.e., questionnaire informatisé) ont aussi été soumis à des essais. La mise à l'essai de l'application s'est poursuivie de façon continue jusqu'au début de la collecte des données de l'enquête auprès des ménages.

### **6.3 Interviews**

La collecte des données a été répartie en quatre trimestres (débutant en juin, août et novembre 2000, et février 2001). Une autre période de collecte s'est tenue en juin 2001 avec des tentatives supplémentaires de dépistage des non-répondants des trimestres précédents.

Les intervieweurs étaient des employés à temps partiel formés spécifiquement pour mener des enquêtes selon la méthode d'interview assistée par ordinateur et la majorité étaient des intervieweurs expérimentés de l'Enquête sur la population active (EPA). On a appliqué à

la collecte des données de l'ENSP la structure de supervision et de vérification de l'EPA. Les intervieweurs ont suivi une formation ciblée sur le contenu de l'ENSP et ils 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 4 de l'ENSP. Ils ont également reçu une brochure présentant entre autre quelques résultats tirés des données de l'ENSP.

En général on communique d'abord par téléphone avec les répondants de l'échantillon des ménages et plus de 95 % des interviews au cycle 4 ont été effectuées par téléphone. On a procédé à des interviews sur place si le répondant n'avait pas le téléphone, dans le cadre d'un dépistage en personne d'un répondant ou à la demande d'un répondant. L'interview complète auprès de chaque ménage a duré en moyenne une heure.

Dans chaque logement, on a recueilli des renseignements généraux sur tous les membres du ménage (âge, sexe, liens de parenté entre les membres) auprès de la personne qui se trouvait à la maison au moment de l'appel de l'intervieweur. Dans le cas des répondants longitudinaux, la déclaration par personne interposée n'a été permise que pour des raisons de maladie ou d'incapacité. Ces déclarations par personne interposée représentent 4,7 % des données recueillies pour les répondants de 12 ans et plus. En revanche, presque toutes les interviews auprès des répondants âgés de moins de 12 ans ont été effectuées par personne interposée.

#### **6.4 Non-réponse et dépistage**

On a demandé aux intervieweurs de déployer tous les efforts raisonnables pour que les 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 ménage. 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 ménage. Puis, l'intervieweur a fait un deuxième appel (ou une deuxième visite). Si le moment de cet appel (ou de cette visite) était mal choisi, l'intervieweur a convenu avec la personne intéressée d'un moment plus opportun pour rappeler ou pour revenir.

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

Les cas de refus ont fait l'objet d'un suivi par les intervieweurs principaux, les surveillants de projet ou d'autres intervieweurs, afin de tenter de convaincre les répondants de participer à l'enquête. Afin de maximiser le taux de réponse, un grand nombre de cas de non-réponse ont fait l'objet d'un suivi lors des périodes de collecte subséquentes.

Le fait de ne pas dépister un répondant longitudinal 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 contacts si ces renseignements avaient été recueillis dans un cycle précédent. En outre, on a formé les intervieweurs pour qu'ils se servent des outils disponibles, comme les annuaires téléphoniques locaux et l'assistance annuaire. Les cas pour lesquels ces démarches n'ont donné aucun résultat ont été transmis à un intervieweur chevronné spécialement formé pour dépister les répondants. Les intervieweurs de dépistage avaient accès aux répertoires téléphoniques de l'ensemble du Canada et à des annuaires par numéros. Le taux de non-réponse lié à l'impossibilité de dépister le répondant longitudinal se chiffre à 2,9 % pour le cycle 4, ce qui est relativement faible.

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.

## **7. Traitement des données**

### **7.1 Vérification**

La vérification des données a été effectuée lors de l'interview assistée par ordinateur. 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 toute question ne s'appliquant pas au répondant ne soit pas posée. En cas de réponse contradictoire des messages d'avertissement étaient donnés. Dans certains cas, la contradiction a dû être résolue avant de poursuivre l'interview et dans d'autres cas, aucune mesure corrective n'a été prise. Après la collecte des données, un programme de vérification a été exécuté au bureau central. En général, les incohérences ont été corrigées en donnant la valeur « non déclaré » pour l'une des variables en cause ou les deux. Aucune imputation n'a été effectuée.

### **7.2 Codage**

Pour plusieurs questions auxquelles il est possible de saisir une réponse écrite, on a codé les renseignements inscrits de façon à les classer dans de nouvelles catégories distinctes ou dans une catégorie énumérée existante si les renseignements inscrits étaient le calque d'une catégorie déjà énumérée. Dans la mesure du possible (par exemple pour la profession, l'industrie, les maladies), on a codé les renseignements conformément aux systèmes de classification type utilisés pour le Recensement de la population ou pour d'autres enquêtes de Statistique Canada, comme l'Enquête sur la santé et les limitations d'activités et le cycle 6 de l'Enquête sociale générale.

Dans les fichiers du cycle 3, les données du cycle 3 concernant l'industrie et l'occupation ont été codées en utilisant le « Système de classification des industries de l'Amérique du Nord (SCIAN) » et la « Classification type des professions de 1991 (SOC-91) ». Cependant, les données des cycles 1 et 2 concernant l'industrie et l'occupation sont restées codées à partir de la « Classification type des industries de 1980 (SIC-80) » et de la « Classification type des professions de 1980 (SOC-80) ». Dans le fichier maître du cycle 4 (contenant tous les membres du panel longitudinal), les données des cycles 1 et 2 ont été « recodées » et par conséquent les quatre cycles sont uniformément codées avec SCIAN et SOC-91.

### **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. Le cinquième caractère du nom des variables dérivées est habituellement un D (voir section 12.3 sur la convention utilisée pour nommer les variables). Dans certains cas, les variables dérivées sont simplement obtenues en groupant des catégories de réponse. Dans d'autres, plusieurs variables sont combinées pour

en former une nouvelle. 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 4 ».

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. Le cinquième caractère du nom des variables groupées est habituellement un G (voir section 12.3 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 représente 50 membres de la population. Dans la terminologie de ce texte, on peut dire que chaque personne a un coefficient de pondération (ou poids) de 50.

La phase de pondération est l'étape où l'on calcule ce poids pour chaque personne. Il faut utiliser ce poids pour produire des estimations valables à partir des résultats de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes au Canada dont l'état général de santé 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 9.

#### **7.5 Sous-ensembles de répondants**

Afin d'offrir davantage de flexibilité aux utilisateurs, un seul fichier maître de microdonnées a été créé pour le cycle 4 de l'ENSP. Ce fichier contient les 17 276 membres du panel longitudinal de l'ENSP peu importe le type de réponse obtenu aux cycles 1 à 4.

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 9. Pour l'utilisation des poids longitudinaux veuillez consulter la section 12.1. Le tableau 7.A donne une description de ces quatre sous-ensembles de répondants en fonction du type de réponse.

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

<b>Sous-ensembles de répondants</b>	<b>Type de réponse</b>	<b>Nombre de répondants</b>
Longitudinal carré	Panel entier : tous les membres du panel peu importe le type réponse aux cycles 1, 2, 3 et 4.	17 276
Longitudinal complet	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1, 2, 3 et 4.	13 582
Longitudinal complet C1 et C4	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1 et 4, peu importe le type de réponse aux cycles 2 et 3.	14 321
Longitudinal complet partagé	Tous les membres du panel qui ont qui ont fourni une réponse complète aux cycles 1, 2, 3 et 4 et qui ont autorisé le partage des données au cycle 4.	13 110

Les utilisateurs du fichier partagé doivent prendre note que le sous-ensemble «complet partagé» est fourni séparément sur un CD-ROM avec les poids d'échantillonnage correspondant. Les poids d'échantillonnage et les variables indicatrices des autres sous-ensembles ne sont pas disponibles sur le CD-ROM du fichier partagé.

## **7.6 Définition d'une réponse complète**

Les trois derniers sous-ensembles de données ont été créés en utilisant la définition de « réponse complète ». 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 fourni une réponse complète à l'interview c'est à dire qui ont répondu à toutes les questions jusqu'à un point déterminé dans le questionnaire de l'ENSP).
- **Les membres du panel décédés.** Le décès d'un membre du panel a été, lorsque possible, confirmé avec la Base canadienne de données de l'état civil - Décès. Lorsque le décès est confirmé, on saisit la cause reliée au décès. Les variables des personnes décédées sont représentées par un « 9 » c.-à-d. non déclaré dans le fichier.
- **Les membres du panel qui ont déménagé en établissement de soins de santé.** La collecte de données pour personnes se fait via le volet établissement de santé. Les données recueillies sont ramenées sur le fichier du volet ménages de l'ENSP. Toutefois, le volet établissement recueille moins d'information que le volet ménages. Les variables manquantes de ces répondants sont représentées par un « 6 » c.-à-d. sans-objet.

## **8. Taux de réponse et taux d'érosion**

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 que les cycles subséquents 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 général (H05) et la composante Santé (H06). Comme l'enquête est devenue purement longitudinale au cycle 4 et qu'il n'y a plus de distinction entre ces deux composantes, le taux de réponse du panel longitudinal n'est calculé qu'une seule fois au cycle 4.

### **8.1 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 d'un ménage hors du champ de l'enquête (code de statut = 017, 018, 023, 024)<sup>2</sup> ont été exclues des calculs des taux de réponse du cycle 1.

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

nbre 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 de l'ENSP, 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.

#### ***Taux de réponse H06 des personnes choisies***

nbre 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.

---

<sup>2</sup> 017 = Autre logement non admissible (p. ex. ambassade)  
018 = Ménage rejeté  
023 = En construction ou démolit  
024 = Logement vacant



***Information pertinente pour le calcul des taux de réponse :***

Nombre de personnes choisies à l'intérieur du champ de l'enquête : 20 095  
Nombre de répondants à la composante H05 parmi les personnes choisies : 17 276  
Nombre de répondants à la composante H06 parmi les personnes choisies : 16 794

Nombre de non-répondants à la composante H05 parmi les personnes choisies : 2 819  
Nombre de non-répondants à la composante H06 parmi les personnes choisies : 3 301

***Calcul des taux de réponse des personnes choisies :***

$$\text{Taux H05 parmi les personnes choisies} = \frac{17\,276}{17\,276 + 2\,819} = \frac{17\,276}{20\,095} = 86,0 \%$$

$$\text{Taux H06 parmi les personnes choisies} = \frac{16\,794}{16\,794 + 3\,301} = \frac{16\,794}{20\,095} = 83,6 \%$$

**8.2 Taux de réponse du cycle 2 (1996-1997)**

Les taux de réponse du cycle 2 sont fondés sur les 17 276 personnes qui forment le panel longitudinal. Les personnes qui sont décédées ou qui ont déménagé en établissement de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête.

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

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H05}}{\text{nbre de membres du panel}} \\ \text{ou qui sont décédés ou qui ont déménagé en établissement de santé}$$

À l'échelle canadienne, le taux de réponse du panel était de **93,6 %**. À l'échelle provinciale, ce taux variait de 90,4 % en Colombie-Britannique à 96,2 % à Terre-Neuve.

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

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H06 ou}}{\text{nbre de membres du panel}} \\ \text{qui sont décédés ou qui ont déménagé en établissement de santé}$$

À l'échelle canadienne, le taux de réponse du panel était de **92,8 %**. À l'échelle provinciale, ce taux variait de 89,6 % en Colombie-Britannique à 95,1 % à Terre-Neuve.

***Information pertinente pour le calcul des taux de réponse :***

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés <sup>3</sup> :	287
Nombre de membres du panel qui ont déménagé en établissement de santé :	62
Nombre de membres du panel répondants au H05 :	15 819
Nombre de membres du panel répondants au H06 :	15 687
Nombre de membres du panel non-répondants au H05 :	1 108
Nombre de membres du panel non-répondants au H06 :	1 240

***Calcul des taux de réponse du panel :***

$$\text{Taux de réponse du panel au H05} = \frac{15\,819 + 287 + 62}{15\,819 + 287 + 62 + 1\,108} = \frac{16\,168}{17\,276} = 93,6 \%$$

$$\text{Taux de réponse du panel au H06} = \frac{15\,687 + 287 + 62}{15\,687 + 287 + 62 + 1\,240} = \frac{16\,036}{17\,276} = 92,8 \%$$

### **8.3 Taux de réponse du cycle 3 (1998-1999)**

Tout comme au cycle 2, les taux de réponse longitudinaux du cycle 3 sont fondés sur le panel longitudinal complet. Les personnes qui sont décédées ou qui ont déménagé en établissement de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête.

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

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H05}}{\text{ou qui sont décédés ou qui ont déménagé en établissement de santé}} \\ \text{nbre de membres du panel}$$

À l'échelle canadienne, le taux de réponse du panel était de **88,9 %**. À l'échelle provinciale, ce taux variait de 84,2 % en Colombie-Britannique à 92,5 % à Terre-Neuve.

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

$$\frac{\text{nbre de membres du panel ayant répondu à la composante H06}}{\text{ou qui sont décédés ou qui ont déménagé en établissement de santé}} \\ \text{nbre de membres du panel}$$

À l'échelle canadienne, le taux de réponse du panel était de **88,2 %**. À l'échelle provinciale, ce taux variait de 83,9 % en Colombie-Britannique à 92,0 % à Terre-Neuve.

---

<sup>3</sup> À ce moment, les décès n'avaient pas été confirmés à l'aide de la Base canadienne de données sur l'état civil – décès. Il a été découvert que quatre de ces personnes étaient vivantes au cycle 3.

***Information pertinente pour le calcul des taux de réponse :***

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés <sup>4</sup> :	599
Nombre de membres du panel qui ont déménagé en établissement de santé	114
Nombre de membres du panel répondants au H05 :	14 647
Nombre de membres du panel répondants au H06 :	14 532
Nombre de membres du panel non-répondants au H05 :	1 916
Nombre de membres du panel non-répondants au H06 :	2 031

***Calcul des taux de réponse du panel :***

$$\text{Taux de réponse du panel au H05} = \frac{14\,647 + 599 + 114}{14\,647 + 599 + 114 + 1\,916} = \frac{15\,360}{17\,276} = 88,9 \%$$

$$\text{Taux de réponse du panel au H06} = \frac{14\,532 + 599 + 114}{14\,532 + 599 + 114 + 2\,031} = \frac{15\,245}{17\,276} = 88,2 \%$$

**8.4 Taux de réponse du cycle 4 (2000-2001)**

Comme dans les cycles 2 et 3, les taux de réponse longitudinaux du cycle 4 sont obtenus à l'aide des 17 276 membres du panel longitudinal. Les personnes qui sont décédées ou qui ont déménagé en établissement de santé sont considérées comme une réponse aux fins longitudinales. Aucun membre du panel n'est classé comme étant hors du champ d'enquête. Comme l'enquête est maintenant purement longitudinale et ne fait plus de distinction entre les questionnaires H05 et H06, un seul taux de réponse est donc calculé.

***Taux de réponse***

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

À l'échelle du Canada, le taux de réponse du panel était de **84,8 %**. À l'échelle provinciale, ce taux variait de 80,5 % en Colombie-Britannique à 90,9 % en Saskatchewan.

---

<sup>4</sup> À ce moment, les décès n'avaient pas été confirmés à l'aide de la Base canadienne de données sur l'état civil - décès. Il a été découvert qu'une de ces personnes était vivante au cycle 4.

***Information pertinente pour le calcul des taux de réponse :***

Nombre de membres du panel longitudinal :	17 276
Nombre de membres du panel décédés :	957
Nombre de membres du panel qui ont déménagé en établissement de santé :	135
Nombre de membres du panel répondants :	13 559
Nombre de membres du panel non-répondants :	2 625

***Calcul du taux de réponse du panel :***

$$\text{Taux de réponse du panel} = \frac{13\,559 + 957 + 135}{13\,559 + 957 + 135 + 2\,625} = \frac{14\,651}{17\,276} = 84,8 \%$$

**8.5 Taux d'érosion**

L'érosion est une réduction de la taille d'un échantillon attribuable aux non-répondants, aux passages hors du champ d'enquête et aux personnes qu'on ne peut dépister. On peut calculer deux taux d'érosion différents : un taux qui présente l'érosion à la fin de chaque cycle et un taux qui présente l'érosion cumulative en fonction de l'échantillon initial. Pour calculer ces deux taux, on utilise le nombre de personnes du sous-ensemble complet. Pour de plus amples renseignements sur le sous-ensemble complet, voir la section 7.5.

***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 :	15 670
Nombre de personnes dans le sous-ensemble complet du cycle 3 :	14 619
Nombre de personnes dans le sous-ensemble complet du cycle 4 :	13 582

***Taux d'érosion à la fin de chaque cycle***

$$\text{Cycle 2 (1996-1997) : } \frac{17\,276 - 15\,670}{17\,276} = \frac{1\,606}{17\,276} = 9,3 \%$$

$$\text{Cycle 3 (1998-1999) : } \frac{15\,670 - 14\,619}{15\,670} = \frac{1\,051}{15\,670} = 6,7 \%$$

$$\text{Cycle 4 (2000-2001) : } \frac{14\,619 - 13\,582}{14\,619} = \frac{1\,037}{14\,619} = 7,1 \%$$

**Taux d'érosion cumulatifs**

Cycle 2 (1996-1997) :	$\frac{17\ 276-15\ 670}{17\ 276}$	=	$\frac{1\ 606}{17\ 276}$	=	9,3 %
Cycle 3 (1998-1999) :	$\frac{17\ 276-14\ 619}{17\ 276}$	=	$\frac{2\ 657}{17\ 276}$	=	15,4 %
Cycle 4 (2000-2001) :	$\frac{17\ 276-13\ 582}{17\ 276}$	=	$\frac{3\ 694}{17\ 276}$	=	21,4 %

**8.6 Taux de non-réponse attribuables aux personnes non-dépistées**

L'incapacité de dépister un membre du panel de l'échantillon longitudinal représente un type de non-réponse. 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. Le tableau 8.A résume le dépistage pour les cycles 2, 3 et 4.

**Tableau 8.A : Dépistage, cycles 2, 3 et 4**

	Cycle 4			Total
	Dépisté et répondant	Dépisté et non-rép.	Non-dépisté	
Cycle 2 Dépisté et répondant	14 426	1 413	320	16 159
Cycle 3 Dépisté et répondant	14 071	900	189	15 160
Cycle 3 Dépisté et non-rép.	303	504	52	859
Cycle 3 Non-dépisté	52	9	79	140
Cycle 2 Dépisté et non-rép.	202	587	32	821
Cycle 3 Dépisté et répondant	151	29	6	186
Cycle 3 Dépisté et non-rép.	44	553	9	606
Cycle 3 Non-dépisté	7	5	17	29
Cycle 2 Incapacité de dépister	103	41	152	296
Cycle 3 Dépisté et répondant	60	10	11	81
Cycle 3 Dépisté et non-rép.	4	10	8	22
Cycle 3 Non-dépisté	39	21	133	193
<b>Total</b>	<b>14731</b>	<b>2041</b>	<b>504</b>	<b>17 276</b>

Par exemple, parmi les 15 160 membres du panel qui sont des répondants dépistés aux cycles 2 et 3, 14 071 étaient des répondants dépistés au cycle 4, 900 étaient des non-répondants dépistés au cycle 4, et 189 étaient non-dépistés au cycle 4.

## 9. Pondération

Le présent chapitre décrit le processus de pondération longitudinale pour chacun des sous-ensembles de répondants définis à la section 7.5. 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, tel que décrit à la section 7.5, la définition d'une réponse longitudinale est différente d'une réponse transversale, ce qui nécessite des corrections différentes particulières 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. Les poids attribués aux membres du panel dans les fichiers transversaux étaient différents de ceux attribués aux membres du panel dans le fichier longitudinal et ne représentaient pas la bonne probabilité de sélectionner le membre longitudinal.

Pour le cycle 4, quatre ensembles de poids ont été créés : WT64LS, WT60LF, WT60LFE et WT60SLF. Le tableau 9.A relie chacun des sous-ensembles de répondants aux variables de poids et aux variables indicatrices correspondantes.

**Tableau 9.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes.**

<b>Sous-ensemble de répondants</b>	<b>Poids</b>	<b>Variables indicatrices</b>
Longitudinal carré	WT64LS	Aucune, tous les enregistrements
Longitudinal complet	WT60LF	WF60LF
Longitudinal complet C1 et C4	WT60LFE	WF60LFE
Longitudinal complet partagé	WT60SLF	WT60SLF

Les poids WT60LF, WT60LFE et WT60SLF ont été corrigés pour la non-réponse et les quatre poids ont été post-stratifiés selon des groupes d'âge-sexe pour chaque province, conformément aux estimations de la population de 1994-1995 basées sur le Recensement de 1996. La post-stratification 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.

### 9.1 Pondération longitudinale

La procédure pour la pondération longitudinale est basée essentiellement sur la pondération effectuée pour l'échantillon transversal du cycle 1 de l'ENSP. Quelques corrections qui ont été appliquées au poids transversaux du cycle 1 ne sont plus nécessaires pour les cycles subséquents. Ces corrections ont été enlevées pour créer un poids « épuré » pour chaque

membre du panel longitudinal. Ce poids épuré est le point de départ pour obtenir les poids longitudinaux.

### **9.1.1 Poids épurés du cycle 1 (1994-1995)**

Les poids épurés du cycle 1 sont obtenus 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 corrections sont apportées à ces 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. Toutes ces corrections effectuées au cycle 1 sont conservées 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 des fichiers de microdonnées à grande diffusion des cycles 2 et 3.

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

### **9.1.2 Poids longitudinaux**

#### **9.1.2.1 Poids longitudinal carré (WT64LS)**

Le poids longitudinal carré **WT64LS** doit être utilisé avec le sous-ensemble longitudinal carré. Il est calculé en post-stratifiant 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. La correction due à la post-stratification est donnée par :

$$\frac{\text{Estimation démographique dans une catégorie province-âge-sexe}}{\text{Somme des poids des membres longitudinaux dans une catégorie province-âge-sexe}}$$

#### **9.1.2.2 Poids longitudinal complet (WT60LF)**

Le sous-ensemble longitudinal complet inclut seulement les membres du panel qui ont fourni une réponse complète à chaque cycle ou qui sont décédés. Les poids des non-répondants doivent être redistribués afin de compenser pour la non-réponse.

Le point de départ est le poids épuré du cycle 1 et des corrections pour la non-réponse sont faites. Une correction différente est effectuée pour la non-réponse à chacun des cycles et ces corrections sont cumulatives d'un cycle à l'autre. Par exemple, pour obtenir le poids du cycle 4, les corrections pour la non-réponse aux cycles 2, 3 et 4 sont appliquées successivement aux poids épurés du cycle 1.

Les corrections nécessaires en vue d'obtenir le poids longitudinal complet du cycle 4 sont décrites ci-dessous.

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

La correction pour la non-réponse a été faite en utilisant l'approche des catégories de pondération. Les catégories de pondération 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 catégorie de pondération ils appartiennent. On définit les catégories à l'aide d'un algorithme de mise en grappe qui permet de disposer les unités d'échantillonnage en une structure arborescente en divisant successivement les ensembles de données en « branches » d'après les caractéristiques des unités. Chaque division a pour objet de répartir les unités existantes en au moins deux groupes 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 fort similaires). On peut se servir d'une autre caractéristique pour définir chaque division. Par exemple, on peut d'abord répartir les unités en un groupe de logements occupés par le propriétaire et un groupe de logements loués. Puis, on peut subdiviser le premier groupe en cinq sous-groupes, d'après le niveau de revenu du ménage, tandis qu'on peut subdiviser le deuxième groupe d'après l'âge du répondant. Chaque groupe nouvellement formé peut ensuite être divisé d'après d'autres caractéristiques et ainsi de suite. Les résultats des subdivisions finales correspondent aux catégories de pondération.

L'algorithme CHAID (« chi-square automatic interaction detection ») a été utilisé pour définir les catégories de pondération. Un minimum de 30 unités par catégorie de pondération a été utilisé afin de produire des corrections plus stables.

Des catégories de pondération ont été créées séparément pour chaque province. Notez que dans ce cas, le terme province se 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 considérées. 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 l'enquête dans l'analyse. Les 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 catégories de pondération pour corriger la non-réponse au cycle 2 sont énumérées ci-dessous. 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 9.B.

**Tableau 9.B : Variables pour la correction de la non-réponse au cycle 2 - WT60LF**

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, la correction qui suit est apportée aux poids des répondants :

$$\frac{\text{Somme des poids de tous les membres longitudinaux}}{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 et 2}}$$

Cette correction est apportée selon les catégories de pondération.

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

Les 15 670 enregistrements avec un statut de réponse complet pour les deux premiers cycles servent de point de départ. Le profil de réponse longitudinal (la variable LONGPAT) a été créé de telle sorte qu'à chaque cycle, on rajoute simplement un chiffre à la fin de cette variable pour indiquer le statut de ce cycle. Pour chaque cycle, le code utilisé dans la construction de cette variable reflète le statut du membre du panel. Les codes sont : 1=complet, 2=décédé, 3=en établissement de santé, 4=partiel et 5=non-réponse. Une « réponse longitudinale complète » après trois cycles est représentée par un des profils de réponse suivants : 111, 112, 113, 122, 131, 132 ou 133. Tous les autres profils de réponse forment la non-réponse (c.-à-d. 114, 115, 134 et 135). Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 (c.-à-d. profil 122) ou en établissement de santé depuis le cycle 2 (c.-à-d. profil 133) sont traités différemment des autres enregistrements. Pour ces enregistrements, aucune correction pour la non-réponse n'est faite puisque leur poids au cycle 2 a déjà été corrigé pour tenir compte du fait que certains non-répondants au cycle 2 étaient possiblement décédés ou en établissement de santé.

L'approche des catégories de pondération a été utilisée pour effectuer la correction pour la non-réponse. Des catégories distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour la correction 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 la correction 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 la correction pour la non-réponse au cycle 3.

Les variables choisies par l'algorithme CHAID afin de produire les catégories de pondération pour corriger la non-réponse au cycle 3 sont énumérées ci-dessous. La variable de plan du cycle 1 indiquant si le répondant provient de l'échantillon 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 9.C.

**Tableau 9.C : Variables pour la correction de la non-réponse au cycle 3 - WT60LF**

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	HWS_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, la correction suivante est apportée aux poids des membres répondants :

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

Cette correction est apportée selon les catégories de pondération. Seuls les enregistrements avec les profils de réponse longitudinale suivants ont été considérés : 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 santé depuis le cycle 2 ne font pas partie de cette correction.

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

Les 14 619 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 santé depuis le cycle 2 ou 3 sont traités différemment des autres enregistrements. Pour ces enregistrements, aucune correction pour la non-réponse n'est faite puisque leur poids au cycle 2 ou 3 a déjà été corrigé pour tenir compte du fait que certains non-répondants au cycle 2 ou 3 étaient possiblement décédés ou en établissement de santé.

Comme pour les cycles précédents, l'approche des catégories de pondération a été utilisée pour effectuer la correction pour la non-réponse. Des catégories distinctes ont été créées pour chaque province, c.-à-d. la province de résidence en 1994-1995. Pour la correction 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 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 la correction pour la non-réponse au cycle 4.

Les variables choisies par l'algorithme CHAID afin de produire les catégories de pondération pour corriger la non-réponse au cycle 4 sont énumérées ci-dessous. 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 9.D.

**Tableau 9.D : Variables pour la correction de la non-réponse au cycle 4 -WT60LF**

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
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, la correction suivante est apportée aux poids des membres 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

Cette correction est apportée selon les catégories de pondération. Les enregistrements pour lesquels le membre du panel était décédé au cycle 2 ou 3 ou en établissement de santé depuis le cycle 2 ou 3 ne font pas partie de cette correction.

#### **Correction 4 : Post-stratification**

Le poids des unités répondantes a été corrigé en le post-stratifiant 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. Cette correction de post-stratification est donnée par :

Estimation démographique dans une catégorie province-âge-sexe  
Somme des poids des membres longitudinaux répondants aux cycles 1 à 4  
dans une catégorie province-âge-sexe

Le poids final pour une réponse longitudinale complète **WT60LF** est calculé en prenant d'abord le poids épuré du cycle 1, et en le multipliant ensuite par les corrections 1 à 4.

#### **9.1.2.3 Poids longitudinal complet cycles 1 et 4 (WT60LFE)**

L'ENSP inclut maintenant, à partir du cycle 4, un poids additionnel à être utilisé avec le sous-ensemble complet C1 et C4. Ce nouveau sous-ensemble inclut les répondants faisant partie du sous-ensemble complet, auquel on rajoute les répondants ayant une réponse complète seulement aux cycles 1 et 4 (c'est-à-dire les non-répondants et les réponses partielles au cycle 2 ou

au cycle 3). Puisqu'un bon nombre de répondants faisant partie de ce nouveau sous-ensemble font aussi partie du sous-ensemble complet, la méthode de pondération utilisée pour ce dernier est à nouveau utilisée, avec quelques modifications.

Le point de départ est le poids longitudinal complet du cycle 4 juste avant l'étape de post-stratification. Tel que décrit dans les sections précédentes, ce poids a été corrigé pour la non-réponse à chacun des cycles. Cependant, certains non-répondants au cycle 2 ou au cycle 3, dont les poids avaient été répartis aux répondants du sous-ensemble complet, ont répondu de nouveau au cycle 4 et leurs poids doivent maintenant être récupérés. Le poids de départ qui est attribué à chaque enregistrement récupéré au cycle 4 est le poids que cet enregistrement avait au dernier cycle où il faisait partie du sous-ensemble complet. Ceci permet d'utiliser le poids le plus récent, qui tient compte des corrections pour la non-réponse aux cycles précédents. L'équivalent des poids ainsi attribués doit être enlevé des poids du sous-ensemble complet du cycle 4. Cette diminution se fait séparément pour chacune des catégories de pondération déterminées au cycle où les enregistrements récupérés étaient non-répondants pour la première fois. Il s'agit en fait de redonner le poids aux enregistrements de la même façon qu'ils l'avaient perdu. Ce poids doit cependant être corrigé afin que les enregistrements récupérés au cycle 4 représentent la même proportion dans le nouveau sous-ensemble complet C1 et C4, qu'ils représenteraient s'ils avaient fait partie du sous-ensemble complet à tous les cycles, c'est-à-dire s'ils avaient été impliqués dans toutes les corrections pour la non-réponse aux cycles subséquents. À nouveau cette correction se fait séparément à l'intérieur de chaque catégorie de pondération.

Finalement, comme pour les autres poids longitudinaux, le poids des unités faisant partie du sous-ensemble complet C1 et C4 a été corrigé en le post-stratifiant 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. Cette correction est donnée par :

Estimation démographique dans une catégorie province-âge-sexe  
Somme des poids des membres du sous-ensemble longitudinal complet  
C1 et C4, dans une catégorie province-âge-sexe

Le poids final pour une réponse complète aux cycles 1 et 4 est le poids WT60LFE. Un document technique donnant plus de détails sur la méthode de pondération du sous-ensemble complet C1 et C4 sera éventuellement disponible sur demande.

#### **9.1.2.4 Poids longitudinal complet partagé (WT60SLF)**

Dans certains cas, les répondants ont indiqué qu'ils n'acceptaient pas de partager l'information recueillie lors de tous les interviews menées dans le cadre de l'ENSP avec les ministères provinciaux de la santé et Santé Canada. 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.

Une seule correction simple est apportée au poids longitudinal complet pour créer le poids du sous-ensemble complet partagé (fichier partagé). Cette correction est donnée par :

$$\frac{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 à 4, dans une catégorie province / profil de réponse longitudinale / âge-sexe}}{\text{Somme des poids des membres longitudinaux répondants aux cycles 1 à 4 et ayant accepté de partager, dans une catégorie province / profil de réponse longitudinale / âge-sexe}}$$

Il est à noter qu'aux cycles 3 et 4, quelques profils de réponse longitudinale ont été regroupés afin de produire des corrections plus stables. Le regroupement a été fait pour quelques catégories province/âge-sexe qui n'avaient que quelques observations dans certains profils de réponse représentant les répondants décédés ou en établissement de santé. Dans chaque cas, le profil de réponse problématique a été regroupé avec un autre profil de la même catégorie province/âge-sexe de telle sorte que la somme des poids corresponde toujours aux totaux de la population. Le regroupement des profils de réponse a préservé la distribution pondérée des membres du panel demeurant dans un ménage, dans un établissement de santé ou décédés pour chacun des cycles. Le poids final pour le sous-ensemble complet partagé, (WT60SLF) est obtenu en multipliant le poids longitudinal complet (WT60LF) par cette correction. Puisque cette correction est faite en respectant les catégories utilisées lors de la correction pour la post-stratification, aucune post-stratification supplémentaire n'est nécessaire.

## 10. Qualité des données

### 10.1 Erreurs dans 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 fassent des erreurs en complétant le questionnaire, que les réponses soient mal saisies et 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*.

#### 10.1.1 Erreurs non dues à l'échantillonnage

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à des biais dans les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, une formation poussée sur les méthodes d'enquête et le questionnaire et l'observation des intervieweurs afin de déceler les problèmes. 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 eu de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Dans la plupart des cas, la non-réponse partielle à l'enquête était attribuable au fait que le répondant ne comprenait pas une question ou l'interprétait mal, refusait d'y répondre, ne se souvenait pas de l'information demandée ou n'était pas en mesure de fournir de l'information sur lui-même ou sur la personne pour laquelle il fournissait une réponse par

procuration. Il y a eu non-réponse totale lorsqu'il était impossible de dépister le répondant, qu'aucun membre du ménage ne pouvait fournir les renseignements voulus ou que le répondant refusait 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 9.2.2.2 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse.

### **10.1.2 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 due à l'échantillonnage est fondée sur l'écart type des estimations tirées des résultats 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 de l'estimation.

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 l'analyse des données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ENSP de 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 11.4.

#### **10.1.2.1 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 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, il a fallu appliquer 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 décrit à la sous-section suivante, 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 post-stratifiés 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 post-stratifier 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.

#### **10.1.2.2 Programme Bootvar pour l'estimation de la variance**

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 explique comment modifier et utiliser le programme selon les besoins de l'utilisateur.

## **11. 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.

### **11.1 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 dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou de rapports) devraient être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.
- e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

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

Le plan d'échantillonnage utilisé pour l'ENSP n'était pas auto-pondéré, c'est-à-dire que les personnes faisant partie l'échantillon n'ont pas toutes le même poids d'échantillonnage. Pour produire des estimations simples, y compris des tableaux statistiques, 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.

Il existe plusieurs progiciels permettant d'inclure les poids de sondage dans le calcul d'estimations provenant d'enquêtes complexes. Toutefois, ces progiciels utilisent différentes techniques pour tenir compte du poids de sondage. Les utilisateurs doivent tenir compte du fait que certains de ces progiciels ne permettront pas nécessairement d'obtenir des estimations qui correspondent exactement à celles de Statistique Canada.

### **11.2.1 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 catégorie 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.

#### ***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.

### **11.2.2 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 :

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

### **11.2.3 Totalisation d'estimations quantitatives**

On peut obtenir les estimations de quantités de la façon suivante :

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

## **11.3 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 poids plus raisonnable. Si l'on transforme les poids des enregistrements contenus dans le fichier de manière à ce que le poids moyen soit un (1), les résultats produits par les logiciels courants seront plus raisonnables et tiendront compte des probabilités inégales de sélection, bien qu'ils continueront à ne faire aucun cas de la stratification et de la répartition en grappes du plan d'échantillonnage. On peut effectuer cette transformation en employant dans l'analyse un poids égal au poids 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 plus 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 logiciels 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 différentes corrections appliquées aux poids.

#### **11.4 Lignes directrices pour la diffusion**

Avant de diffuser ou de publier des estimations tirées du fichier maître, les utilisateurs doivent d'abord déterminer le nombre de répondants qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à 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 11.A.

**Tableau 11.A : Lignes directrices relatives à la variabilité d'échantillonnage**

<b>Type d'estimation</b>	<b>CV (en %)</b>	<b>Lignes directrices</b>
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	<p>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 :</p> <p>«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».</p>

## **12. Utilisation du fichier maître longitudinal**

### **12.1 Utilisation des poids longitudinaux**

Dans les cycles antérieurs, quelques fichiers contenant un sous-ensemble des 17 276 membres du panel longitudinal ont été créés. Cette fois-ci, un seul fichier a été créé. Le fichier maître du cycle 4 contient quatre poids (section 9.1) correspondant à quatre sous-ensembles de répondants (section 7.5). Des variables indicatrices ont été créées afin d'identifier les enregistrements appartenant à chacun des sous-ensembles de répondants (tableau 9.A section 9). 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 carré » et s'applique au sous-ensemble « carré » qui comprend les 17 276 membres qui constituent l'échantillon longitudinal. Les utilisateurs qui désirent mener des études spécialisées sur le biais de non-réponse doivent utiliser ce poids.

Le poids WT60LF est appelé « poids longitudinal complet » et s'applique aux 13 582 enregistrements qui forment le sous-ensemble « complet ».

Le poids WT60LFE est appelé « poids longitudinal complet C1 et C4 » et s'applique aux 14 321 enregistrements qui forment le sous-ensemble « complet C1 et C4 ».

Le poids WT60SLF est appelé « poids longitudinal complet partagé » et s'applique aux 13 110 enregistrements du sous-ensemble « complet partagé ».

### **12.2 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 10.1.2.1 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 4, quatre ensembles différents de poids bootstrap ont été créés, soit le carré, le complet, le complet partagé et le complet C1 et C4. Le tableau 12.A présente les sous-ensembles de répondants et les fichiers de poids bootstrap correspondants.

**Tableau 12.A : Sous-ensembles de répondants et fichiers de poids Bootstrap correspondants**

<b>Sous-ensemble de répondants</b>	<b>Poids Bootstrap</b>
Longitudinal carré	B5long
Longitudinal complet	B5lngf
Longitudinal complet C1 et C4	B5lngfe
Longitudinal complet partagé	B5lngf (partagé)

En raison de la complexité du plan d'échantillonnage, les utilisateurs devraient utiliser le programme Bootvar pour calculer la variance. Les données de sortie 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 11 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 diverses corrections de pondération qui peuvent, dans certains cas, influencer considérablement sur les estimations de la variance.

### **12.3 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 diverses composantes / 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...) 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, en 1996-1997 et en 1998-1999. 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 santé et, auparavant, le volet des résidents du Nord et les suppléments.

#### **12.3.1 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 : Année / cycle durant lequel apparaît la variable

Position 5 : Type de variable

Positions 6 à 8 : Numéro / nom de la variable tiré du questionnaire



Par exemple : la signification des noms de variable DHC4\_AGE, DHC6\_AGE, DHC8\_AGE et DHC0\_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** : figure dans le cycle 1 (1994-1995), figure dans le cycle 2 (1996-1997), figure dans le cycle 3 (1998-1999), figure dans le cycle 4 (2000-2001);
- \_** : peut être trouvé sur le questionnaire ;
- AGE** : nom de la variable.

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

AD	Dépendance à l'égard de l'alcool	IS	Assurance
AL	Alcool	LF/LS	Population active
AM	Administration de l'enquête	MH	Santé mentale
AP	Opinions au sujet des parents	NU	Nutrition
BP	Tension artérielle	PA	Activités physiques
CC	Problèmes de santé chroniques	PC	Examen général
CE	Abandon du contact	PH	Santé préventive
CO	Sens de la débrouillardise (achat de contenu de l'Alberta, cycles 1 et 2)	PR	Province
DG	Consommation de médicaments	PY	Ressources psychologiques (estime de soi, contrôle, sentiment de cohésion)
DH	Démographie et ménage	RA	Limitation des activités
DV	Visites chez le dentiste	RP	Mouvements répétitifs
ED	Niveau de scolarité	RS	Sécurité routière
ES	Services d'urgence	RT	Rationalité (achat de contenu du Manitoba, cycle 1)
EX	Examen de la vue	SC	Soins personnels
FH	Antécédents médicaux de la famille	SD	Renseignements socio-démographiques
FI	Précarité alimentaire (achat de Développement des ressources humaines Canada au cycle 3)	SH	Santé sexuelle
FS	Vaccination antigrippale	SM	Usage du tabac
GE	Identificateurs géographiques	SP	Identificateurs d'échantillon
GH	État de santé général	SS	Soutien social
HC	Utilisation des soins de santé	ST	Stress
HI	Renseignements sur la santé	SV	Services de santé
HN	Numéro d'assurance-maladie	TA	Variante de l'usage du tabac
HS	État de santé	TU	Bronzage et exposition aux rayons UV
HV	VIH	TW	Incapacité au cours des deux dernières semaines
HW	Taille et poids	VS	Violence / sécurité personnelle
IJ	Blessures	WH	Santé des femmes : auto-examen des seins, examen des seins, mammographie, test de Papanicolaou et hystérectomie
IN	Revenu	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 12.B présente les variables qui n'apparaissent qu'une seule fois sur le fichier de données. Le nom de ces variables ne suit pas la convention.

**Tableau 12.B : Variables longitudinales « constantes »**

<b>Variables longitudinales</b>	<b>Concept</b>
SEX	Sexe
DOB	Jour de naissance
MOB	Mois de naissance
YOB	Année de naissance
COB	Pays de naissance
COBC	Code pour le pays de naissance
COBGC	Pays de naissance (7 groupes) - Groupé
IMM	Indicateur d'immigration pour le répondant
AOI	Âge au moment de l'immigration - Dérivé
HWB	Poids à la naissance
DOD	Jour de décès
MOD	Mois de décès
YOD	Année de décès
COD	Code pour la cause de décès

### **12.3.3 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 la précarité 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 particulières à un cycle qui ne sont pas répétées chaque fois (par ex. stress aux cycles 1 et 4, accès aux services de santé au cycle 2)
- 3 Variables d'administration de l'enquête pour le volet des ménages et la composante démographique (H03)
- 5 Variables d'administration de l'enquête pour la composante Générale (H05)
- 6 Variables d'administration de l'enquête pour la composante Santé (H06)

**12.3.4 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)
F	Cycle 11 (2014-2015)

**12.3.5 Position 5 : Type de variable**

_	Variable recueillie	Variable qui figure directement sur le questionnaire
C	Variable codée	Variable codée à partir d'une ou de plusieurs variables 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 un groupe (p. ex. groupes d'âge)
L	Variable longitudinale dérivée	Variable calculée d'après des variables provenant d'au moins deux cycles d'enquête

**12.3.6 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 ST\_4\_C1, la première question sur les facteurs de stress durant l'enfance et à l'âge adulte (traumatisme) a été nommée ST\_4\_T1. Un autre exemple survient dans le cas des questions sur l'état de santé général de l'Enquête sur la promotion de la santé. Ces questions ont été réparties en trois sections aux fins de leur intégration au questionnaire et les noms des variables correspondantes reflètent cette situation. Le caractère de la position 6 indique dans quelle section la question figure.

## 13. Accès aux données de l'ENSP

### 13.1 Fichiers de microdonnées

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 du cycle 4 (et l'accès aux fichiers maîtres de microdonnées transversales des trois premiers cycles de l'ENSP) est possible via le service de télé-accès 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'Unité d'accès aux données à [nphs-ensp@statcan.ca](mailto:nphs-ensp@statcan.ca).

L'accès direct sur place aux fichiers maîtres de microdonnées de l'ENSP est aussi possible dans les Centres de données de recherche (CDR) de Statistique Canada. Ces centres constitués en collaboration avec le Conseil de recherches en sciences humaines du Canada (CRSH) sont situés dans des endroits physiquement sécuritaires d'universités participantes. Ils sont en quelque sorte une extension de bureaux de Statistique Canada (SC) avec un employé de SC à plein temps dans chaque centre. Les chercheurs y poursuivent leur travail en vertu de la *Loi sur la statistique* comme tout employé de SC le ferait. Vous pouvez trouver de l'information supplémentaire sur le site du Conseil de recherches en sciences humaines du Canada à <http://www.sshrc.ca/rdc>.

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 par 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/edu/index.htm>. Les FMDG de l'ENSP des cycles 1,2 et 3 peuvent également être achetés. Veuillez communiquer par courriel avec le personnel du soutien technique de la Division de la statistique de la santé à [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca) ou dans un des bureaux régionaux de Statistique Canada.

### 13.2 Rapports analytiques et totalisations

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 imprimé sous le numéro de catalogue 82-003-XPF ou en format électronique dans le site de Statistique Canada sous le numéro de catalogue 82-003-XIF. Pour plus d'informations, visiter notre site Web [www.statcan.ca](http://www.statcan.ca) et sélectionner « Produits et services ».

Des tableaux personnalisés de l'ENSP peuvent être obtenus à frais recouvrables. Pour des estimations de faisabilité et des coûts, communiquer avec le soutien technique de la Division de la statistique de la santé à [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca).

Finalement, le Guide de la statistique de la santé (82-573-GIF) dans le site Web de Statistique Canada est un bon point de départ pour toute information reliée aux indicateurs de la santé provenant de différentes sources incluant l'ENSP. Le guide fait aussi le lien avec plusieurs documents relatifs à l'ENSP incluant les questionnaires. Pour consulter le Guide allez à [http://www.statcan.ca/francais/freepub/82-573-GIF/guide\\_f.htm](http://www.statcan.ca/francais/freepub/82-573-GIF/guide_f.htm).

**Annexe A : ENSP volet ménages, Contenu de la composante Santé au cours des 4 cycles**

<b>Contenu principal de l'ENSP : cycle 1 au cycle 4</b>				
	<b>Cycle 1 (1994-1995)</b>	<b>Cycle 2 (1996-1997)</b>	<b>Cycle 3 (1998-1999)</b>	<b>Cycle 4 (2000-2001)</b>
AL - Alcool	√	√	√	√
AM - Administration	√	√	√	√
CC - Problèmes de santé chroniques	√	√	√	√
DG - Consommation de médicaments	√	√	√	√
DH - Démographie et ménage	√	√	√	√
ED - Niveau de scolarité	√	√	√	√
GH - État de santé général	√	√	√	√
HC - Utilisation des soins de santé	√	√	√	√
HS - État de santé	√	√	√	√
HW - Taille et poids	√	√	√	√
IJ - Blessures	√	√	√	√
IN - Revenu	√	√	√	√
IS - Assurance		√	√	√
LF / LS - Population active	√	√	√	√
MH - Santé mentale	√	√	√	√
PA - Activités physiques	√	√	√	√
PH - Santé préventive	√		√	√
RA - Limitation des activités	√	√	√	√
RP - Mouvements répétitifs		√	√	√
SD - Renseignements socio-démographique	√	√	√	√
SM - Usage du tabac	√	√	√	√
SS - Soutien social	√	√	√	√
TU - Bronzage et exposition aux rayons UV				√
TW - Incapacité au cours des deux dernières semaines	√	√	√	

**ENSP Cycle 4 (2000-2001), Documentation longitudinale**

<b>Contenu thématique de l'ENSP : Cycles 1 à 4</b>				
	<b>Cycle 1 (1994- 1995)</b>	<b>Cycle 2 (1996-1997)</b>	<b>Cycle 3 (1998-1999)</b>	<b>Cycle 4 (2000-2001)</b>
AD - Dépendance à l'égard de l'alcool		√		
BP – Tension artérielle		√		
CC - Problèmes de santé chroniques (arthrite, maladie cardiaque, diabète)				√
DV – Visites chez le dentiste		√		
ES – Services d'urgence		√		
EX – Examen de la vue		√		
FH – Antécédents médicaux de la famille			√	
FS – Vaccination antigrippale		√		
GH – Allaitement		√		
NU - Nutrition			√	
PC – Examen général		√		
PY – Ressources psychologiques	√		√	
SC - Soins personnels			√	
SS - Stress (problèmes actuels, événements récents, facteur de stress chez les enfants et les adultes (« traumatismes »), stress au travail, estime de soi, contrôle)	√			√
WH - Santé préventive (examen des seins, auto-examen des seins (EPS), mammographie, test de Papanicolaou)		√		



<b>Contenu supplémentaire acheté, ENSP, Cycles 1 à 3</b>			
	<b>Cycle 1 (1994-1995)</b>	<b>Cycle 2 (1996-1997)</b>	<b>Cycle 3 (1998-1999)</b>
Enquête sur la promotion de la santé		√	
AL - Alcool		√	
FI - Insécurité alimentaire		√	√
GH - Allaitement		√	
GH - Grossesse		√	
HW - Taille et poids		√	
HV - VIH		√	
NU – Nutrition / diète		√	√
RS – Sécurité routière		√	
SH – Santé sexuelle		√	
SM - Usage du tabac		√	
TA – Variante du tabagisme			√
WH - Auto examen des seins		√	
Composante provinciale (achat de contenu)	√	√	
AP – Opinions au sujet des parents		√	
CO – Sens de la débrouillardise (Manitoba)	√		
CO - Sens de la débrouillardise (Alberta)	√	√	
HI – Renseignements sur la santé		√	
SH - Santé sexuelle		√	
SS - Soutien social		√	
SV – Services de santé		√	
SV - Services de santé pour les enfants		√	
TU - Bronzage et exposition aux rayons UV		√	
VS – Sécurité personnelle et violence		√	

**Annexe B : ENSP, volet ménages, Changements au questionnaire pour le cycle 4 (2000-2001)**

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

Pour ce qui suit, les noms de variables externes sur le questionnaire seront utilisés. Les noms des variables du fichier maître et du fichier partagé suivent la convention utilisée pour nommer les variables (voir la section 12.3).

**AM – Changement de résidence**

- Suppression : Questions sur le lieu de résidence habituel au cycle 1 (AMC8\_MV1, AMC8\_MV2, et AMC8\_MV3) – uniquement pour l'enquête longitudinale.

**AL – Alcool**

- Addition : Question sur la consommation d'alcool depuis la dernière interview (ALC0\_1A) – Si le répondant répond « non », l'application passe à la section suivante. S'il répond « oui », il doit répondre à la question « Avez-vous bu de l'alcool au cours des 12 derniers mois ? » (ALC0\_1). Si le répondant répond « non », on lui demande s'il a déjà pris un verre d'alcool (ALC0\_5B); la réponse devrait toujours être vraie parce que la réponse à (ALC0\_1A) est « oui ».

**AM – Administration**

- Suppression : Dans la question sur le consentement de partager les données (AM60\_SHA), « Développement des ressources humaines Canada (DRHC) » est enlevé de la liste des organismes avec lesquels il y a partage d'information.
- Addition : Question sur le numéro provincial d'assurance-maladie afin de faciliter le jumelage des données (AM60\_H3A).
- Addition : Question sur la province du numéro d'assurance-maladie (AM60\_H3B).
- Addition : Algorithmes pour éditer la validité des numéros d'assurance-maladie AM\_HN – introduits dans l'application informatique pour la plupart des provinces.

**CC – Problèmes de santé chroniques**

- Addition : Question sur la fibromyalgie (CCC0\_1X).
- Changement : Question sur l'arthrite et les rhumatismes (CCC0\_1D) – Ajouter « sauf la fibromyalgie ».
- Changement : Question sur les maux de dos (CCC0\_1E) –Ajouter- « autres que la fibromyalgie et l'arthrite».
- Addition : Questions sur l'âge au premier diagnostic pour chaque problème chronique, sauf « autre problème » (CCC0\_A3 à CCC0\_X3).
- Changement / clarification : Question sur la prise d'insuline pour le diabète (CCC0\_J5) - ajouter « actuellement » et « contre le diabète » est remplacé par «pour le diabète ».

- Changement : Question sur les autres traitements que l'insuline pour le diabète (CCC0\_J6) - « contre le diabète » remplacé par « pour le diabète ».
- Suppression : Questions sur fréquences de problèmes d'infections chroniques chez les enfants de moins de 3 ans (infections du nez ou de la gorge, et des oreilles) (CCK8\_1, CCK8\_2, et CCK8\_3).
- Suppression : Question sur les problèmes de sinusite (CCC8\_1I).
- Suppression : Questions sur le mois et l'année du diagnostic de certains problèmes de santé chroniques (CCC8\_C3M à CCC8\_O3M, et CCC8\_C3Y à CCC8\_O3Y) – remplacer par la question sur l'âge au premier diagnostic.
- Suppression : Questions qui confirment que certains problèmes chroniques existaient avant la dernière interview (CCC8\_C4 à CCC8\_O4) – enlever parce que le mois et l'année du diagnostic ne sont pas demandés.
- Suppression : Questions sur le traitement ou les médicaments (CCC8\_D5) et la question sur le choix de toutes les réponses appropriées pour le type de traitement ou de médicament pris (CCC8\_D6A à CCC8\_D6D) – remplacer par question thématique semblable (CC\_0\_D3 et CC\_0\_D3A à CC\_0\_D3H).
- Intégration : Pour le contenu thématique, questions sur le diagnostic et le contrôle - Voir « changements au contenu thématique ».

#### **DG – Consommation de médicaments**

- Changement : Question sur les exemples de pilules pour maigrir (DGC0\_1C) – ajouter – « tels que le Redux, le Ponderal ou le Fastin ».
- Changement : Question sur les exemples d'antidépresseurs (DGC0\_1D) – Ajouter – « tels que le Prozac, le Paxil ou l'Effexor ».
- Addition : Question sur l'utilisation d'autres produits de ce genre au cours des 2 derniers jours (DGC0\_4A).
- Addition : Question sur le nom exact du produit utilisé au cours des 2 derniers jours (DGC0C5A) et (DGC0C5B à DGC0C5L) – jusqu'à concurrence de 12 produits.
- Suppression : Question sur le nom exact du produit (aucune période de référence) (DGC8C5A à DGC8C5L) - jusqu'à concurrence de 12 produits.

#### **ED – Niveau de scolarité**

- Aucun changement.

#### **GH - État de santé général**

- Addition : Question sur l'évaluation personnelle du niveau de stress du répondant (STC0\_1).

### **HC - Utilisation des soins de santé**

- Suppression : Questions sur le lieu de la plus récente consultation avec un médecin de famille (HCC8\_3A), autre médecin (HCC8\_3C), ou infirmière (HCC8\_3D) - inclus dans l'ESCC .
- Suppression : Questions sur les soins de santé reçus aux États-Unis (HCC8\_11 et HCC8\_12) - inclus dans l'ESCC.
- Addition : Question sur les médecins de famille (TWC0\_5).

### **HS - État de santé**

- Aucun changement.

### **HW - Taille et poids**

- Suppression : Question sur le poids à la naissance (GHK8\_6) – aucun nouveau bébé dans l'échantillon .

### **IJ – Blessures**

#### **Révision majeure afin d'inclure le nouveau contenu et d'utiliser la Classification internationale des maladies, 10<sup>ième</sup> version (ICD-10)**

- Révision : Question sur le genre de blessure (IJC0\_3) - les catégories de réponses ont changé.
- Addition / révision : Question sur la partie du corps atteinte (IJC0\_4) - les catégories de réponses ont changé.
- Addition : Question sur la partie atteinte (IJC0\_4A) - détail de la blessure aux organes internes.
- Addition / révision : Question sur le lieu d'incident de la blessure (IJC0\_5) - les catégories de réponses ont changé.
- Addition / révision / suppression : Question sur le genre d'activité lors de la blessure (IJC0\_9), blessure à la suite d'une chute (IJC0\_10), manière à laquelle la chute s'est produite (IJC0\_10A) et la cause de la blessure (IJC0\_10B) – remplace (IJC8\_6).
- Addition : Question sur la réception de soins dans les 48 heures suivant la blessure (IJC0\_11) .
- Addition : Question sur le lieu où ont été reçus les soins (IJC0\_12).
- Addition : Question sur l'admission à l'hôpital (IJC0\_13).
- Addition : Question sur les blessures traitées sans limiter les activités ordinaires (IJC0\_14).
- Addition : Question sur le nombre de blessures traitées qui ne limitent pas les activités ordinaires (IJC0\_15).
- Suppression : Question sur les accidents de travail (IJC8\_7).
- Suppression : Question sur les précautions prise afin d'empêcher ce genre de blessure (IJC8\_8A à IJC8\_8G) – Choisir toutes les réponses appropriées.

**IN – Revenu**

- Aucun changement.

**IS – Assurance**

- Aucun changement.

**LF / LS - Population active**

**Révision majeure pour inclure le nouveau contenu et pour rendre les concepts semblables à ceux de l'Enquête sur la population active**

En général, le concept d'état (de l'activité) présent a été remplacé par l'état la semaine dernière. La possibilité d'avoir jusqu'à trois emplois a été remplacée par l'emploi principal et tous les autres emplois. La définition de l'emploi principal n'a pas changé. Les dates de début et de fin d'un à trois emplois (lesquelles permettent de dériver des variables sur les écarts entre les différents emplois occupés) ont été remplacées par le nombre de semaines travaillées et à chercher du travail et le nombre de semaines sans travail dans les 12 derniers mois. Une question sur le nombre de périodes de recherche d'emploi a aussi été ajoutée. Il y a plusieurs nouvelles questions comme celles sur la recherche d'emploi et la raison principale pour le type d'heures travaillées.

- Addition / révision : Question sur l'emploi la semaine dernière (LSC0\_1) remplace les questions sur le travail actuel (LFC8\_61, LFC8\_62, et LFC8\_63) – changement de l'expression « a travaillé contre rémunération ou en vue d'un bénéfice » par « travaille à un emploi ou une entreprise ».
- Addition : Question sur les absences d'un emploi ou d'une entreprise la semaine dernière (LSC0\_2).
- Addition : Question concernant le fait d'avoir plus d'un emploi la semaine dernière (LSC0\_3) - remplace (LFC8\_111 et LFC8\_112).
- Addition : Question sur la recherche de travail au cours des 4 dernières semaines (LSC0\_11).
- Addition : Question sur la date de début d'emploi futur (connue la semaine dernière) (LSC0\_12).
- Addition / révision : Question sur la raison de ne pas travailler présentement - (LSC0\_13) - les catégories de réponses ont changé; remplace (LFC8\_17B).
- Addition / révision : Question sur le travail à un emploi ou une entreprise au cours des 12 derniers mois (LSC0\_21) – remplace (LFC8\_2).
- Addition : Question sur la recherche de travail au cours des 12 derniers mois (LSC0\_22).
- Addition : Question sur l'existence de plus d'un emploi ou entreprise au cours des 12 derniers mois (LSC0\_23) – ceux qui ont travaillé la semaine dernière sont exclus de l'univers.

- Addition / révision / suppression : Question sur l'état du travail (employé ou à votre compte) (LSC0\_31) - remplace (LFC8\_16), changement de la catégorie de réponse « travailleur familial non-rémunéré » par « travail non-rémunéré dans une entreprise familiale ».
- Addition / révision : Question sur le travail à son compte « nom de l'entreprise » (LSC0F32 et LSC0F33) - remplace (LFC8FE1).
- Addition / révision : Question sur le « nom de l'employeur » de l'employé (LSC0F33 et LSC0F32) - remplace (LFC8FE1).
- Addition / révision : Question sur le « genre d'entreprise » (LSC0F34) petit changement de terme - remplace (LFC8F13).
- Addition / révision : Question sur le « genre de travail » (LSC0F35) petit changement de terme - remplace (LFC8F14).
- Addition / révision : Question sur les « activités ou fonctions » (LSC0F36) petit changement de terme - remplace (LFC8F15).
- Addition / révision : Question sur la raison de l'absence de l'emploi ou de l'entreprise la semaine dernière (LSC0\_41).
- Addition / révision : Nombre d'heures habituelles au travail (LSC0\_42) voir aussi (LSC0\_53) - remplace (LFC8\_81, LFC8\_82, et LFC8\_83) pour chaque emploi.
- Addition : Question sur le nombre d'heures préféré et la rémunération (LSC0\_43).
- Addition / révision : Question sur l'horaire de travail habituel (LSC0\_44) - remplace (LFC8\_91, LFC8\_92, et LFC8\_93) pour chaque emploi.
- Addition / révision : Question sur la raison principale pour l'horaire de travail habituel (LSC0\_45).
- Addition / révision : Question sur le travail les fins de semaine (LSC0\_46) voir aussi (LSC0\_54) - remplace (LFC8\_101, LFC8\_102, et LFC8\_103) pour chaque emploi.
- Addition / révision : Question sur le nombre de semaines travaillées à plus d'un emploi au cours des 12 derniers mois (LSC0\_51).
- Addition / révision : Question sur la raison principale de travailler à plus d'un emploi (LSC0\_52).
- Addition / révision : Question sur le nombre d'heures habituelles à l'autre emploi (LSC0\_53) voir aussi (LSC0\_42) - remplace (LFC8\_81, LFC8\_82, et LFC8\_83).
- Addition / révision : Question sur le travail les fins de semaine à l'autre emploi (LSC0\_54) voir aussi (LSC0\_46) remplace (LFC8\_101, LFC8\_102, et LFC8\_103) pour chaque emploi.
- Addition / révision : Question sur le nombre de semaines d'emploi au cours des 12 derniers mois (LSC0\_61).
- Addition / révision : Question sur le nombre de semaines à chercher du travail au cours des 12 derniers mois (LSC0\_71).
- Addition / révision : Question sur le nombre de semaines sans travailler et sans chercher du travail au cours des 12 derniers mois (LSC0\_72).
- Addition / révision : Question sur la raison de ne pas chercher du travail (LSC0\_73).
- Addition / révision : Question sur le nombre de périodes passé à chercher du travail (LSC0\_74)

- Suppression : Question sur la date du début et de la cessation pour chaque emploi (LFC8\_71D, LFC8\_71M, LFC8\_71Y, LFC8\_72D, LFC8\_72M, LFC8\_72Y, LFC8\_73D, LFC8\_73M, LFC8\_73Y).
- Suppression : Question qui confirme que le répondant travaille pour le même employeur que lors de la dernière interview - (LFC8\_2B).

#### **MH - Santé mentale**

- Aucun changement.

#### **PA - Activités physiques**

- Aucun changement.

#### **PH - Santé préventive**

- Suppression /nouveau saut de question : Suppression des questions pour confirmer des incohérences à propos de la tension artérielle (BPC8\_10A), le test PAP (WHC8\_20A) et la mammographie (WHC8\_30A) car la réponse au cycle 3 constitue dorénavant la confirmation.
- Addition : Question sur la raison de passer une mammographie, une nouvelle catégorie de réponse a été ajoutée : « problème mammaire » (WHC0\_33H) – Cette question a été demandée pour la dernière fois au cycle 2.
- Addition / changement : Question sur la raison d'avoir subi une hystérectomie (WHC0\_5BA à WHC0\_5BG) – le genre de question est passé de « cocher une réponse seulement » à « choisissez toutes les réponses appropriées ».

#### **RA - Limitation des activités**

- Aucun changement.

#### **RP - Mouvements répétitifs**

- Addition / révision : Question sur la partie du corps atteinte / mouvements répétitifs (RPC0\_3) - les catégories de réponse sont passé de 9 à 13. La liste ressemble plus à celle de la version révisée de la question sur la partie du corps qui a été blessée (IJC0\_4)
- Addition : Question sur le genre d'activité / mouvement répétitif (RPC0\_5A à RPC0\_5F) - choisissez toutes les réponses appropriées.
- Suppression : Question sur le lieu où la blessure a été causée par mouvement répétitif (RPC8\_4A à RPC8\_4D) – choisissez toutes les réponses appropriées.

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

- Suppression : Questions sur le pays de naissance (SDC8\_1) et l'année de l'immigration au Canada (SDC8\_3).

### **SM - Usage du tabac**

- Addition : Nouveau saut de question SM\_C103 L'âge auquel les fumeurs réguliers (qui étaient fumeurs réguliers au cycle précédent) ont commencé à fumer tous les jours n'est plus demandé. L'information a déjà été recueillie.
- Addition : Nouveau saut de question SM\_C105D On ne demande plus aux personnes qui avaient déjà fumé tous les jours s'ils ont déjà fumé tous les jours. L'information a déjà été recueillie.

### **SS - Soutien social**

- Aucun changement.

### **TU - Bronzage et exposition aux rayons UV**

- Addition : Question sur les coups de soleil au cours des 12 derniers mois (TUC0\_3).

### **TW - Incapacité au cours des deux dernières semaines**

- Suppression : Questions sur les jours alités et les jours limités (TWC8\_1 à TWC8\_4) – inclus dans l'ESCC.
- Déplacement : Question sur le médecin de famille (TWC8\_5) - fait maintenant partie de la section Utilisation des soins de santé.

### **Changements au contenu thématique**

Les sections thématiques du cycle 3 sur les soins personnels, les antécédents médicaux de la famille, la nutrition et le sentiment de cohésion ont été éliminées. Le contenu thématique suivant est inclus au cycle 4 :

### **CC – Problèmes de santé chroniques (diagnostic et contrôle)**

#### **Arthrite**

- Addition : Question sur le genre d'arthrite ou de rhumatisme - genre (CC\_0\_D1).
- Addition : Question sur les interventions chirurgicales à cause de l'arthrite (CC\_0\_D2).
- Addition : Question sur le type d'opération chirurgicale pour l'arthrite (CC\_0\_D2A à CC\_0\_D2D) – Choisissez toutes les réponses appropriées.
- Addition / révision : Question sur le traitement pour soulager la douleur due à l'arthrite - (CC\_0\_D3).
- Addition / révision : Questions sur le type de traitement pour l'arthrite (CC\_0\_D3A à CC\_0\_D3H) - Choisissez toutes les réponses appropriées.
- Addition : Question sur le type de médicament pour l'arthrite (CC\_0\_D4A à CC\_0\_D4C) - Choisissez toutes les réponses appropriées.



## **Diabète**

- Addition : Question afin de savoir si le diagnostic a été posé durant la grossesse (CC\_0\_J3A).
- Addition : Question sur le diagnostic à un moment autre que durant la grossesse (CC\_0\_J3B).
- Addition : Question sur le moment quand a commencé à prendre de l'insuline (CC\_0\_J3C).
- Addition : Question sur la prise de l'insuline (CC\_0\_J4).
- Addition : Question sur le nombre de fois par jour que la personne prend de l'insuline (CC\_0\_J4A).
- Addition : Question sur le nombre d'injection d'insuline par jour (CC\_0\_J4B).
- Addition : Question afin de savoir si le répondant a suivi un cours sur la façon de contrôler le diabète (CC\_0\_J5A).
- Addition : Question visant à savoir si un membre du ménage a suivi un cours sur la façon de contrôler le diabète (CC\_0\_J5B).
- Addition : Question visant à savoir si un professionnel de la santé a donné des renseignements sur la diète (CC\_0\_J5C).
- Addition : Question sur le type de professionnel qui a donné des renseignements sur la diète (CC\_0\_J6A à CC\_0\_J6E) - Choisissez toutes les réponses appropriées.
- Addition : Question sur la vérification de la concentration de sucre (CC\_0\_J7A).
- Addition : Question visant à savoir si un membre du ménage a appris à vérifier la concentration de sucre (CC\_0\_J7B).
- Addition : Question sur la fréquence de vérification de la concentration de sucre (CC\_0\_J8A).
- Addition : Question sur la fréquence de l'examen des pieds (CC\_0\_J8B).

## **Maladie cardiaque**

- Addition : Question sur les crises cardiaques (CC\_0\_L1A).
- Addition : Question sur le nombre de crises cardiaques (CC\_0\_L1B).
- Addition : Question sur l'âge à la première crise (CC\_0\_L1C).
- Addition : Question sur l'âge à la crise la plus récente (CC\_0\_L1D).
- Addition : Question sur l'hospitalisation (CC\_0\_L1E).
- Addition : Question sur les traitements pour les crises cardiaques (CC\_0\_L4A à CC\_0\_L4E) - Choisissez toutes les réponses appropriées.
- Addition : Question sur le programme de réadaptation (CC\_0\_L5A).
- Addition : Question sur la participation à un programme de réadaptation - (CC\_0\_L5B).
- Addition : Question afin de savoir si le répondant a achevé un programme de réadaptation (CC\_0\_L5C).
- Addition : Question sur l'angine (CC\_0\_L6).
- Addition : Question sur l'insuffisance cardiaque globale (CC\_0\_L7).

## **ST - Stress**

Les questions sont les mêmes que celles du cycle 1 avec les changements suivants :

### **Problèmes actuels**

- Aucun changement.

### **Événements récents**

- Aucun changement.

### **Facteurs de stress chez les enfants et les adultes (“traumatismes”).**

- Révision de l’univers : Questions posées aux répondants qui avaient moins de 18 ans au cycle 1 et qui ont maintenant 18 ans et plus. L’information a déjà été recueillie pour les répondants qui avaient 18 ans et plus lors du cycle.

### **Stress au travail**

- Révision de l’univers : Les questions sont dorénavant posées aux personnes âgées entre 15 et 75 ans et non aux personnes âgées de 15 ans et plus. C’est la même population que celle qui répond aux questions sur la population active.
- Addition : Question visant à savoir si le répondant a un emploi (ST\_0\_W1).

### **Estime de soi**

- Aucun changement.

### **Contrôle**

- Aucun changement.

### **Changements à l’achat de contenu**

Il n’y a pas d’achat de contenu au cycle 4. Les questions suivantes provenant du contenu supplémentaire du cycle 3 ont donc été éliminées :

### **FI – Insécurité alimentaire**

- Suppression : Question sur l’inquiétude vis-à-vis un manque d'argent pour la nourriture.
- Suppression : Question sur le manque de nourriture à cause d'un manque d'argent (FIC8\_2).
- Suppression : Question afin de savoir si le répondant n'a pas mangé des aliments de la qualité ou de la variété désirées à cause d'un manque d'argent (FIC8\_3).

**Annexe C : ENSP volet ménages, Exemples de rétroaction des données du cycle 3 (1998-1999) et de questions de suivi**

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 du cycle 4, 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 ?» était posée.
Limitation des activités	On s'est servi des données indiquant si le répondant souffrait ou non d'une incapacité au cycle 3. En cas de changement de situation, on a posé des questions pour obtenir une explication.
Problèmes de santé chroniques	Pour chaque répondant, on a utilisé les réponses du cycle précédent concernant certains problèmes de santé chroniques (asthme, hypertension, migraine, diabète, épilepsie, ulcères de l'estomac ou de l'intestin et les séquelles d'un accident cérébro-vasculaire) pour essayer d'obtenir des explications concernant le changement. Si le problème était nouveau, on a enregistré la date de sa manifestation.
Usage du tabac	Si un fumeur quotidien avait déclaré lors de la dernière interview l'âge auquel il ou elle avait commencé à fumer tous les jours, cette réponse était utilisée au cycle 4. En ce qui concerne le fumeur occasionnel ou le non-fumeur au cycle 4 qui avait déclaré fumer (ou avoir déjà fumé) tous les jours lors d'une interview antérieure, on a utilisé un indicateur d'usage quotidien. En cas de changement de situation, on a posé des questions pour obtenir une explication.
Renseignements socio-démographiques	Pour chaque répondant, on a utilisé un indicateur montrant que les renseignements concernant le pays de naissance ont été recueillis. On a de nouveau recueilli les renseignements pour les variables origine ethnique et race. 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 du cycle 3. 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 et que le répondant avait donné le même numéro que celui enregistré lors du dernier cycle, on a demandé au répondant de donner son numéro d'assurance-maladie à nouveau.