



**STATISTIQUE CANADA**  
**ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION**  
**VOLET ÉTABLISSEMENTS DE SOINS DE SANTÉ**  
**CYCLE 5 (2002-2003)**  
**DOCUMENTATION LONGITUDINALE**  
**Janvier 2007**



Statistique  
Canada

Statistics  
Canada

Canada

## NOTE AUX UTILISATEURS

Depuis la collecte du cycle 3 en 1998-1999, l'Enquête nationale sur la santé de la population (ENSP) – volet établissements de soins de santé 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 5 du volet établissements de soins de santé de l'ENSP. Ce fichier contient les 2 287 membres du panel longitudinal de l'ENSP – volet établissements de soins de santé peu importe le type de réponse obtenu aux cycles précédents. Ce fichier maître a trois sous-ensembles de répondants avec les poids de sondage correspondants ainsi que des variables indicatrices rendant leur identification plus facile.

La documentation longitudinale du cycle 5 de l'ENSP fournit un large éventail d'information sur l'enquête : objectifs, contenu, plan d'échantillonnage, collecte, traitement, qualité des données, pondération, lignes directrices pour la totalisation et accès aux données. Le chapitre 12 donne 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, c'est-à-dire les ministères provinciaux de la santé, Santé Canada et l'Agence de santé publique du Canada. Le fichier partagé inclut les répondants qui ont accepté de partager leurs données au cycle 5 de même que le poids d'échantillonnage correspondant. Ce groupe de répondants est un sous-ensemble du fichier maître. 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

Cycle 5 = 2002-2003

## Table des matières

<b>1. Introduction .....</b>	<b>1</b>
<b>2. Contexte .....</b>	<b>2</b>
<b>3. Objectifs .....</b>	<b>3</b>
<b>4. Contenu de l'enquête .....</b>	<b>4</b>
4.1. Critères.....	4
4.2. Révisions du contenu pour le cycle 5 (2002-2003) .....	5
<b>5. Plan d'échantillonnage.....</b>	<b>6</b>
5.1. Stratification et sélection des établissements au cycle 1 (1994-1995) .....	6
5.2. Sélection des résidents au cycle 1 (1994-1995).....	7
5.3. Échantillon longitudinal .....	8
<b>6. Collecte des données.....</b>	<b>9</b>
6.1. Conception du questionnaire et collecte des données .....	9
6.2. Non-réponse à l'ENSP.....	9
<b>7. Traitement des données .....</b>	<b>10</b>
7.1. Saisie des données et vérification .....	10
7.2. Codage .....	10
7.3. Création de variables dérivées .....	10
7.4. Estimation et pondération .....	11
7.5. Sous-ensembles de répondants .....	11
7.6. Définition du profil de réponse longitudinal .....	12
<b>8. Qualité des données.....</b>	<b>13</b>
8.1. Taux de réponse longitudinaux.....	13
8.1.1. Taux de réponse du cycle 1 (1994-1995).....	13
8.1.2. Taux de réponse des cycle 2 (1996-1997), cycle 3 (1998-1999), cycle 4 (2000-2001) et cycle 5 (2002-2003).....	14
8.2. Taux d'érosion .....	15
8.3. Erreurs relatives à l'enquête .....	16
8.3.1. Erreurs d'échantillonnage .....	16
8.3.2. Erreurs non dues à l'échantillonnage .....	17
8.4. Imputation.....	18
<b>9. Lignes directrices pour la totalisation, l'analyse et la diffusion .....</b>	<b>20</b>
9.1. Lignes directrices pour l'arrondissement .....	20
9.2. Lignes directrices pour la pondération de l'échantillon en vue de la totalisation .....	21
9.2.1. Définitions des catégories d'estimation : nominale par opposition à quantitative ..	21
9.2.2. Totalisation d'estimations de type nominal .....	22
9.2.3. Totalisation d'estimations quantitatives .....	22
9.3. Lignes directrices pour l'analyse statistique .....	23
9.4. Lignes directrices pour la diffusion .....	24

<b>10. Pondération.....</b>	<b>26</b>
10.1. Poids longitudinal carré (WTI4LS) .....	26
10.2. Poids longitudinal complet (WTI2LF) .....	28
10.3. Poids longitudinal complet partagé (WTI2SLF) .....	30
<b>11. Calcul de la variance.....</b>	<b>32</b>
11.1. Méthode du bootstrap .....	32
11.2. Estimation de la variance avec le programme BOOTVAR.SAS (.SPS) .....	33
<b>12. Utilisation du fichier maître longitudinal .....</b>	<b>34</b>
12.1. Utilisation des poids longitudinaux .....	34
12.2. Assurance de la fiabilité des estimations avec l'utilisation des poids bootstrap .....	35
12.3. Convention utilisée pour nommer les variables .....	35
12.3.1. Structure des noms des variables .....	36
12.3.2. Positions 1 et 2 : Nom de la variable / section du questionnaire .....	36
12.3.3. Position 3 : Type d'enquête / volet .....	37
12.3.4. Position 4 : Variable de l'année de référence / du cycle .....	37
12.3.5. Position 5 : Type de variable .....	38
12.3.6. Positions 6 à 8 : Nom de la variable .....	38
12.4. Accès aux fichiers maîtres .....	39
12.4.1. Fichiers de microdonnées .....	39
12.4.2. Totalisations personnalisées .....	39

## Liste des tableaux

- Tableau 5.A : Taille du sous-ensemble longitudinal complet, pour chaque cycle
- Tableau 7.A : Sous ensembles de répondants
- Tableau 8.A : Taux de réponse des établissements
- Tableau 8.B : Taux de réponse individuel
- Tableau 9.A : Lignes directrices relatives à la variabilité d'échantillonnage
- Tableau 10.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes
- Tableau 12.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes

## **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 volet établissements de soins de santé de l'ENSP, est la première enquête nationale longitudinale menée auprès des résidents des établissements de soins de santé au Canada. Le premier cycle de l'enquête a eu lieu en 1994-1995. L'ENSP a recueilli des renseignements de nature transversale et longitudinale durant les deux premiers cycles, mais depuis le cycle 3 (1998-1999) l'ENSP, volet établissements de soins de santé est devenue uniquement une enquête longitudinale.

Le volet établissements de soins de santé de l'ENSP, comprend les résidents des établissements de soins de longue durée (plus de 6 mois) de toutes les provinces, sauf ceux des territoires, des réserves indiennes et des bases des Forces canadiennes. Les établissements de soins de santé de longue durée considérés devaient avoir un minimum de 4 lits et les résidents ne devaient pas être autonomes. Cinq cycles de collecte ont été complétés pour le volet établissements de soins de santé : l'ENSP cycle 1 (1994-1995), l'ENSP cycle 2 (1996-1997), l'ENSP cycle 3 (1998-1999), l'ENSP cycle 4 (2000-2001) et l'ENSP cycle 5 (2002-2003).

Après 5 cycles de collecte (1994-1995 à 2002-2003), le volet établissements de soins de santé de l'ENSP se termine à cause du grand nombre de décès dans l'échantillon. Les répondants du volet ménages, qui déménagent dans un établissement de soins de santé, continuent à être suivis et font toujours partie de l'échantillon du volet ménages.

Le volet établissements de soins de santé de l'ENSP cycle 5, a recueilli des renseignements détaillés sur la santé des répondants longitudinaux sélectionnés au hasard lors du premier cycle. Tout comme au cycle 4, la collecte de données des personnes faisant partie de l'échantillon des établissements de soins de santé qui sont retournées en ménages privés a été faite par le volet des établissements de soins de santé et non le volet des ménages. Les questions posées aux résidents des ménages n'étaient pas les mêmes que celles posées aux résidents des établissements, mais plus de 90 % des questions étaient identiques.

Ce document a été produit pour faciliter l'utilisation des fichiers maître et partagé du cycle 5 (2002-2003) du volet établissements de soins de santé de l'ENSP. Ces fichiers sont décrits de façon plus détaillée dans les chapitres suivants. Pour toute question au sujet des données ou de leur utilisation, veuillez vous adresser aux :

### **Services d'information et d'accès aux données**

Division de la statistique de la santé

Requêtes, accès (Centres de données de recherche, Accès à distance)

Téléphone : 1-613-951-1653, Courriel : [nphs-ensp@statcan.ca](mailto:nphs-ensp@statcan.ca),

Télécopieur : 1-613-951-0792

Totalisations personnalisées ou renseignement généraux sur les données

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

Télécopieur : 1-613-951-0792

## **2. Contexte**

À l'automne 1991, le Conseil national d'information sur la santé (CNIS) a recommandé que l'on réalise une enquête nationale permanente sur la santé de la population. Cette recommandation découlait des pressions économiques et financières exercées sur le régime de santé et du besoin correspondant de renseignements permettant d'améliorer l'état de santé de la population canadienne. Or, les sources existantes de données sur la santé ne suffisaient pas pour brosser un tableau complet de l'état de santé de la population ni de la multitude de facteurs qui ont un effet sur la santé.

En avril 1992, Statistique Canada a obtenu les fonds nécessaires pour élaborer l'Enquête nationale sur la santé de la population. Cette enquête a été conçue pour être flexible et dans le but de produire des données valides, fiables et d'actualité. Il fallait aussi que l'enquête puisse répondre à l'évolution des besoins, des intérêts et des politiques.

On a ajouté à l'enquête un volet spécial, qui couvre les résidents des établissements de soins de santé, parce que cette population fait rarement partie du champ d'observation des enquêtes nationales et que les caractéristiques de leur état de santé diffèrent vraisemblablement de celles de la population générale.

### **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 permettant de réaliser des études analytiques qui aideront à comprendre les déterminants de la santé;
- recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques et environnementaux;
- permettre de mieux comprendre la relation entre l'état de santé et l'utilisation des services de santé;
- 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, et pour déterminer quels facteurs influencent le placement en établissement;
- donner aux provinces, aux territoires et à d'autres clients une capacité d'enquête sur la santé qui leur permettra d'augmenter le contenu ou l'échantillon;
- permettre de coupler les données de l'enquête à des données administratives courantes, telles que les statistiques de l'état civil, les mesures environnementales, les variables liées aux collectivités, ainsi que l'utilisation des services de santé.

## **4. Contenu de l'enquête**

### **4.1. Critères**

Le contenu du volet établissements de soins de santé de l'ENSP a été sélectionné conformément aux critères suivants :

- 1) l'enquête permettra de recueillir des données sur l'état de santé des Canadiens qui résident dans un établissement de soins de santé;
- 2) les données recueillies pourront être comparées à celles sur la population résidant dans des ménages privés, dans la mesure du possible;
- 3) l'enquête permettra de mieux comprendre les circonstances du placement en établissement;
- 4) les renseignements tirés de l'enquête permettront d'étudier, au fil du temps, la transition de la vie dans un ménage au placement en établissement et inversement;
- 5) l'enquête permettra de recueillir des données à l'échelle nationale.

Les répondants ont été sélectionnés au hasard dans certains établissements de soins de santé. Deux questionnaires ont été utilisés pour la collecte des données; le Formulaire de contrôle pour l'établissement (FCE) et le questionnaire pour le répondant (QR). Le FCE a servi à recueillir de l'information concernant les politiques des établissements et le QR portait sur l'état de santé, les facteurs de risque, le soutien social, les contacts avec les professionnels de la santé, ainsi que la situation démographique et socio-économique. Par exemple, pour mesurer l'état de santé, on a posé des questions sur la perception personnelle de l'état de santé, la capacité fonctionnelle, les problèmes de santé chroniques et la limitation des activités. Les questions sur les comportements à risque portaient sur l'usage du tabac et la consommation d'alcool. Pour évaluer le niveau de soutien social, on a déterminé la fréquence des contacts avec des amis et des parents dans l'établissement et à l'extérieur de celui-ci. Les renseignements de nature démographique et socio-économique incluent l'âge, le sexe, le niveau de scolarité, l'origine ethnique et le revenu personnel.



#### **4.2. Révisions du contenu pour le cycle 5 (2002-2003)**

Il y a eu peu de changements apportés au contenu du cycle 5. Les items suivants ont été modifiés :

- L'année prédéfinie pour la date de l'interview sur la page couverture du questionnaire du répondant a été changée à 2003.
- La question sur la permission de jumelage a été modifiée.
- La question sur la permission de partage a été modifiée.
- Les codes d'état final sur la page couverture du questionnaire du répondant ont été changés pour suivre les codes standards utilisés par l'ensemble des enquêtes de Statistique Canada.

## 5. Plan d'échantillonnage

La population cible du volet établissements de soins de santé de l'ENSP comprend les résidents des établissements de soins de longue durée (plus de 6 mois) de toutes les provinces, sauf ceux des territoires, des réserves indiennes et des bases des Forces canadiennes. Les établissements de soins de santé de longue durée considérés devaient avoir un minimum de 4 lits et les résidents ne devaient pas être autonomes. Les établissements qui ne font pas partie du système de santé, c'est-à-dire les établissements correctionnels, les prisons, les établissements pour jeunes délinquants, ainsi que les foyers collectifs pour enfants (orphelinats) et les établissements religieux ne sont pas inclus dans la base de sondage des établissements de soins de santé.

L'échantillon longitudinal du volet établissements de soins de santé de l'ENSP est composé des répondants longitudinaux choisis au cycle 1 (1994-1995). Au cycle 1, on a fixé la taille de cet échantillon à 2 600 résidents. Avec un taux de réponse hypothétique de 85 %, un échantillon de cette taille est suffisant pour produire des estimations nationales dont le coefficient de variation (c.v.) est de 10 % pour les caractéristiques présentées par au moins 10 % de la population.

À la suite de la collecte du cycle 1, l'échantillon comprenait 2 287 répondants qui habitaient dans un établissement de soins de santé en 1994-1995.

La sélection des répondants a été faite en deux étapes. Premièrement les établissements de soins de santé ont été sélectionnés, puis des résidents ont été sélectionnés à l'intérieur de ces établissements.

### 5.1. Stratification et sélection des établissements au cycle 1 (1994-1995)

On a créé une liste d'établissements de soins de santé (longue durée, un minimum de 4 lits et résidents non autonomes). Au départ, on a stratifié cette liste d'établissements selon la région géographique (strates géographiques), puis selon la catégorie d'établissement (strates de catégorie) et le nombre de lits (strates de taille).

Les strates géographiques correspondent à cinq régions : les provinces de l'Atlantique, le Québec, l'Ontario, les provinces des Prairies et la Colombie-Britannique. Dans chaque strate géographique, on a défini trois strates de catégorie :

Établissements pour personnes âgées	y compris les établissements de soins accueillant des personnes âgées et les hôpitaux de soins prolongés;
Établissements de soins cognitifs	y compris les établissements de soins accueillant des enfants présentant des troubles affectifs et des personnes présentant des troubles mentaux ou des troubles de retard du développement et les hôpitaux psychiatriques;

Autres établissements de réadaptation y compris les centres de réadaptation, les hôpitaux pédiatriques et d'autres hôpitaux spécialisés, les hôpitaux généraux comprenant des unités de soins de longue durée, ainsi que les établissements de soins accueillant des personnes physiquement handicapées.

Dans chacune de ces strates région géographique-catégorie, on a formé des strates de taille en regroupant les établissements comptant un nombre de lits comparable. Le nombre de strates de taille dépend du nombre total de lits dans la strate région géographique-catégorie. Une fois le nombre de strates de taille déterminé, on a fixé les limites des diverses strates de taille en appliquant la règle  $Cum \sqrt{f(y)}$  où  $f(y)$  représente le nombre de lits.

Au premier cycle, le nombre d'établissements sélectionnés à partir d'une strate de taille dépendait de la taille de l'échantillon attribué à la strate (voir la section 5.2) et de la taille des établissements compris dans cette dernière. Pour les strates formées de grands établissements, on a sélectionné un plus grand échantillon de résidents dans chaque établissement, ce qui a réduit le nombre total d'établissements visités. Une fois qu'on a déterminé le nombre d'établissements à sélectionner dans chaque strate de taille, on a procédé à un échantillonnage systématique d'établissements figurant sur la liste de la strate avec probabilité de sélection proportionnelle à la taille (PPT). La taille était fonction du nombre de lits de soins de longue durée.

Un siège social regroupant plusieurs petits établissements pouvait se trouver sur la liste. Le cas échéant, on a obtenu la liste de tous les établissements relevant de ce siège social et on en a sélectionné deux, à savoir le plus grand (d'après le nombre de lits) et un second sélectionné au hasard selon la méthode d'échantillonnage PPT.

## **5.2. Sélection des résidents au cycle 1 (1994-1995)**

Une fois les établissements sélectionnés, des résidents de ces établissements ont été sélectionnés. L'échantillon total de 2 600 résidents a été réparti proportionnellement entre les strates de taille, en fonction du nombre de lits dans chaque strate. Un minimum de 30 résidents par strate a toutefois été fixé.

À la suite de la collecte du cycle 1, l'échantillon comprenait 2 287 répondants qui habitaient dans un établissement de soins de santé en 1994-1995.

### 5.3. Échantillon longitudinal

L'échantillon longitudinal, aussi appelé panel longitudinal ou simplement panel, est composé des 2 287 personnes sélectionnées au cycle 1 et qui ont répondu complètement ou partiellement au questionnaire du cycle 1. C'est cet échantillon longitudinal qui a été suivi aux cycles 2, 3, 4 et 5. À cause du grand nombre de décès dans le panel, la composante des établissements de soins de santé se termine avec le cycle 5.

L'échantillon longitudinal n'est pas renouvelé au fil du temps. Aucun membre du panel n'a été et ne sera classé hors du champ de l'enquête. La taille de l'échantillon longitudinal demeurera donc égale à 2 287 pour tous les cycles.

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 (par exemple les refus ou les personnes qui n'ont pu être retracées). Malgré l'érosion, l'échantillon longitudinal est toujours représentatif de la population cible de 1994-1995. L'érosion étant très faible (voir la section 8.2), elle ne devrait pas entraîner de fortes augmentations de la variance des estimations. Il est à noter que les membres du panel du volet établissements qui sont décédés ou qui ont déménagé dans un ménage privé font toujours partie de l'échantillon longitudinal du volet établissements. De plus, les membres décédés sont considérés comme répondants. Ces personnes ne contribuent donc pas à l'érosion du panel longitudinal.

Le tableau 5.A présente la taille de l'échantillon longitudinal tel que défini en 1994-1995 et montre également le nombre de personnes ayant répondu partiellement ou complètement (faisant ainsi partie du sous-ensemble longitudinal complet, voir section 7.5) pour chacun des cycles. Le tableau contient aussi le nombre de personnes décédées parmi ce sous-ensemble.

**Tableau 5.A : Taille du sous-ensemble longitudinal complet, pour chaque cycle**

<b>Cycle</b>	<b>Sous-ensemble longitudinal complet</b>	<b>Personnes décédées</b> <i>(Parmi le sous-ensemble longitudinal complet)</i>
1	2 287	--
2	2 192	721
3	2 178	1 250
4	2 143	1 524
5	2 131	1 708

## **6. Collecte des données**

### **6.1. Conception du questionnaire et collecte des données**

Les questions du volet établissements de soins de santé de l'ENSP sont conçues pour être posées lors d'une entrevue sur place par la méthode papier et crayon. L'entrevue téléphonique est permise dans les cas où une personne répondant par procuration ne peut être interrogée sur place. Pour chaque établissement, l'administrateur ou une personne-ressource a déterminé pour lesquels des résidents sélectionnés les réponses devaient être recueillies par procuration. Cette décision était basée sur l'état de santé des résidents sélectionnés. La personne répondant par procuration pouvait être un parent, un membre du personnel ou un bénévole de l'établissement. En tout, 72 % des entrevues ont eu lieu par procuration (dans 34 % des cas, la personne qui a répondu par procuration était un parent du résident). Un membre du personnel de l'établissement a fourni pour chaque résident sélectionné, des renseignements sur la consommation de médicaments et les contacts avec les professionnels de la santé.

La collecte des données a eu lieu d'avril à juin 2003. Les entrevues ont été menées par des intervieweurs de Statistique Canada. Pour commencer, des intervieweurs ont communiqué avec chaque établissement pour organiser une rencontre entre un intervieweur et l'administrateur ou une personne-ressource de l'établissement. Durant cette visite de liaison, l'intervieweur a fait remplir un court questionnaire sur les politiques de l'établissement. La liste des résidents pour lesquels les réponses devaient être recueillies par procuration a également été dressée à ce moment-là et l'intervieweur a obtenu pour chacun de ces résidents le nom et le numéro de téléphone d'un proche parent. Puis, on a téléphoné aux proches parents auxquels on a offert d'être interviewés eux-mêmes ou d'autoriser un membre informé du personnel de l'établissement à répondre aux questions de l'enquête.

Tous les intervieweurs travaillent sous la supervision d'intervieweurs principaux. Ces derniers sont chargés de veiller à ce que les intervieweurs connaissent bien les concepts et les procédures de l'enquête. Ils contrôlent périodiquement le travail des intervieweurs et revoient les documents que ceux-ci remplissent. Les intervieweurs principaux sont, à leur tour, sous la surveillance des gestionnaires de projet postés dans les divers Bureaux régionaux de Statistique Canada.

### **6.2. Non-réponse à l'ENSP**

Les intervieweurs ont reçu pour instructions de faire tous les efforts raisonnables en vue d'interviewer les résidents sélectionnés ou la personne désignée le cas échéant. Les refus de la part des établissements ont fait l'objet d'un suivi effectué par des intervieweurs principaux, des gestionnaires de projet ou d'autres intervieweurs pour tenter de convaincre ces établissements de participer à l'enquête. Au cycle 5, il n'y a pas eu de refus au niveau des établissements.

## **7. Traitement des données**

### **7.1. Saisie des données et vérification**

À la fin de chaque entrevue, l'intervieweur a examiné le questionnaire pour s'assurer d'avoir suivi correctement l'enchaînement des questions. Les bureaux régionaux ont fait une vérification supplémentaire des questionnaires pour s'assurer que les réponses étaient complètes, lisibles et cohérentes. Il a donc été possible, au besoin, de faire un suivi immédiat auprès du répondant.

Le questionnaire du répondant ainsi que le formulaire de contrôle pour les établissements ont été saisis au Bureau central en utilisant le système EP90 (« Entry Point 90 »). Les programmes de saisie de données des questionnaires sont conçus pour éviter la saisie de valeurs aberrantes. Cette information a été vérifiée à 100 % dû au petit nombre de données.

Après la saisie des données, on s'est assuré que l'enchaînement des questions était correct et on a procédé à des contrôles de cohérence sur certaines zones des enregistrements. Sauf dans le cas de l'Indice de l'état de santé (HUI3), aucune imputation n'a été effectuée (voir la section 8.4).

### **7.2. Codage**

Les problèmes de santé causant une limitation des activités ont été codés conformément à la Classification internationale des maladies, 9<sup>e</sup> édition (CIM-9) ou au Système de codage supplémentaire pour les incapacités ostéo-articulaires et musculaires établi au Canada aux fins de l'Enquête sur la santé et les limitations d'activités (ESLA). Pour le cycle 5, les problèmes de santé causant une limitation des activités ont aussi été codés conformément à la Classification statistique internationale des maladies et des problèmes de santé connexes, 10<sup>e</sup> révision (CIM-10). Les deux codes (CIM-9 et CIM-10) sont inclus sur le fichier longitudinal. La cause de décès (lorsqu'applicable) a aussi été codée avec la CIM-9 et la CIM-10.

La classification des médicaments se fonde sur la Classification anatomique, thérapeutique chimique (ATC) développée par l'Organisation mondiale de la santé tel que contenue dans la Base de données sur les produits pharmaceutiques (BDPP) de Santé Canada en date de septembre 2003. Une révision complète des codes de médicaments a été faite pour tous les répondants longitudinaux de l'ENSP pour le cycle 5 (2002-2003) et pour tous les cycles précédents.

### **7.3. Création de variables dérivées**

Pour simplifier l'analyse des données, on a dérivé plusieurs variables du fichier d'après des réponses au questionnaire du volet établissements de soins de santé de l'ENSP. La

lettre « D » en cinquième position dans le nom d'une variable indique que celle-ci est une variable dérivée. Les détails sur la création de ces variables figurent dans la documentation des variables dérivées.

#### **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 10.

#### **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 5 de l'ENSP, volet établissements de soins de santé. Ce fichier contient les 2 287 membres du panel longitudinal de l'ENSP peu importe le type de réponse obtenu aux cycles 1 à 5. Dans ce fichier maître, trois 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 10. Pour l'utilisation des poids longitudinaux veuillez consulter la section 12.1. Le tableau 7.A donne une description de ces trois sous-ensembles de répondants en fonction du type de réponse.

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

Sous-ensembles de répondants	Type de réponse	Variabes indicatrices	Nombre de répondants
Longitudinal carré	Panel entier : tous les membres du panel peu importe le type de réponse aux cycles 1, 2, 3, 4 et 5.	Aucune, tous les enregistrements	2 287
Longitudinal complet	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1, 2, 3, 4 et 5.	<b>WFI2LF</b>	2 131
Longitudinal complet partagé	Tous les membres du panel qui ont fourni une réponse complète aux cycles 1, 2, 3, 4, et 5 et qui ont autorisé le partage des données.	<b>WFI2SLF</b>	2 067

### 7.6. Définition du profil de réponse longitudinal

À chaque cycle, chacun des membres du panel se voit assigner un statut de réponse. Dépendamment du résultat de l'interview, une personne se voit assigner un des quatre statuts suivants : ménage (1), décédé (2), institutionnalisé (3) ou non-réponse (5). Il existe une série de règles pour distinguer ces statuts. Un membre du panel qui a fourni une réponse complète ou partiel à l'interview a un statut « institutionnalisé » ou « ménage » suivant qu'il demeure dans un établissement ou dans un ménage. Si le membre longitudinal n'est pas décédé, institutionnalisé ou en ménage, il a le statut « *non-réponse* ». Pour le panel longitudinal de l'ENSP établissements de soins de santé, une réponse complète comprend les statuts « ménage », « *décédé* » et « institutionnalisé ».

Au fil des cycles, les statuts de réponse d'un même répondant sont concaténés en une seule variable appelée le « patron de réponse longitudinal » (LONGPAT). Cette variable est disponible dans le fichier maître de microdonnées et permet de connaître rapidement le profil de réponse d'un membre du panel. Cette variable est aussi utilisée pour définir les différents sous-ensembles analytiques tels que décrits à la section 7.5.



## 8. Qualité des données

### 8.1. Taux de réponse longitudinaux

Deux taux de réponse distincts peuvent être calculés pour le fichier longitudinal du volet établissements de soins de santé de l'ENSP, à savoir le taux de réponse des établissements et le taux de réponse individuel. 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 2 444 résidents faisant partie du champ d'enquête et choisis en vue de former le panel, alors que les taux de réponse des cycles subséquents sont basés sur les 2 287 résidents formant le panel.

#### 8.1.1. Taux de réponse du cycle 1 (1994-1995)

##### Taux de réponse des établissements

Le taux de réponse des établissements correspond au pourcentage d'établissements faisant partie du champ d'observation de l'enquête qui ont accepté que l'enquête soit menée auprès de leurs résidents<sup>1</sup>. Ces derniers ne pouvaient être interrogés sans l'autorisation de l'établissement. Le calcul du taux de réponse des établissements est le suivant :

$$\frac{\text{Nombre d'établissements sélectionnés qui ont accepté de participer}}{\text{Nombre total d'établissements où logeaient des personnes faisant partie du panel}} \times 100$$

$$= \frac{214}{224} \times 100 = 95,5 \%$$

##### Taux de réponse individuel

Le taux de réponse individuel correspond au pourcentage de résidents sélectionnés dans les établissements répondants qui ont participé à l'interview. Le calcul de ce taux est le suivant :

$$\frac{\text{Nombre de résidents qui ont participé complètement ou partiellement à l'interview}}{\text{Nombre total de résidents sélectionnés dans les établissements participants}} \times 100$$

$$= \frac{2\,287}{2\,444} \times 100 = 93,6 \%$$

NOTA : la multiplication des deux taux entre eux ne donne aucun résultat valable car le nombre de résidents sélectionnés varie selon l'établissement.

<sup>1</sup> Les résidents qui faisaient partie des 8 établissements hors du champ de l'enquête ont été exclus des calculs des taux de réponse du cycle 1.

### 8.1.2. Taux de réponse des cycle 2 (1996-1997), cycle 3 (1998-1999), cycle 4 (2000-2001) et cycle 5 (2002-2003)

#### Taux de réponse des établissements

Le taux de réponse des établissements est basé sur le nombre d'établissements total où habitaient les personnes faisant partie du panel (soient les établissements qui ont participé au premier cycle de l'enquête et qui étaient encore en exploitation, et tous les établissements nouvellement couverts par l'enquête) et est calculé selon la formule suivante :

$$\frac{\text{Nombre d'établissements sélectionnés qui ont accepté de participer}}{\text{Nombre total d'établissements où logeaient des personnes faisant partie du panel}} \times 100$$

Le taux de réponse des établissements étaient de 100 % aux cycles 2, 3 et 5, et de 99,3 % au cycle 4. Ces résultats sont présentés au tableau 8.A.

**Tableau 8.A : Taux de réponse des établissements**

Cycle	Nombre d'établissement où logeaient les membres du panel		Taux de réponse (1) / (2)
	Qui ont accepté de participer (1)	Total (2)	
2	314	314	100,0 %
3	352	352	100,0 %
4	287	289	99,3 %
5	270	270	100,0 %

#### Taux de réponse individuel

Le taux de réponse individuel des cycles 2, 3, 4 et 5 sont obtenu à l'aide des 2 287 membres du panel longitudinal. Les personnes décédées, de même que celles déménagées dans un ménage sont considérées comme des unités répondantes aux fins longitudinales. Le taux de réponse individuel est calculé à l'aide de la formule suivante :

$$\frac{\text{Nombre de résidents qui ont participé complètement ou partiellement à l'interview}}{\text{Nombre total de résidents sélectionnés dans les établissements participants}} \times 100$$

Les taux de réponse individuel du panel étaient de 95,9 % au cycle 2, 98,4 % au cycle 3, 96,9 % au cycle 4 et 97,7 % au cycle 5. Ces taux de réponse individuel sont présentés au tableau 8.B.

**Tableau 8.B : Taux de réponse individuel**

Cycle	Nombre de résidents qui ont participé complètement ou partiellement à l'interview				Taux de réponse = $\frac{(1) + (2) + (3)}{2\ 287}$
	En établissement (1)	Dans un ménage (2)	Morts (3)	Total (1) + (2) + (3)	
2	1 427	44	722	2 193	95,9 %
3	907	73	1 271	2 251	98,4 %
4	595	50	1 572	2 217	96,9 %
5	421	41	1 773	2 235	97,7 %

NOTA : la multiplication des taux de réponse des établissements et individuel entre eux ne donne aucun résultat valable car le nombre de résidents sélectionnés varie selon l'établissement.

## 8.2. Taux d'érosion

L'érosion est une réduction de la taille d'un échantillon attribuable aux non-répondants. Pour le volet établissements de soins de santé de l'ENSP, l'érosion est très faible étant donné le taux de réponse élevé. Il est important de rappeler que les personnes décédées font partie du sous-ensemble longitudinal complet et sont considérées comme « répondants ».

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 (voir la section 5.3).

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

Nombre de membres du panel de l'échantillon longitudinal : 2 287  
 Nombre de personnes dans le sous-ensemble longitudinal complet du cycle 2 : 2 192  
 Nombre de personnes dans le sous-ensemble longitudinal complet du cycle 3 : 2 178  
 Nombre de personnes dans le sous-ensemble longitudinal complet du cycle 4 : 2 143  
 Nombre de personnes dans le sous-ensemble longitudinal complet du cycle 5 : 2 131

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

Cycle 2 (1996-1997) :	$\frac{2\ 287 - 2\ 192}{2\ 287}$	=	$\frac{95}{2\ 287}$	=	4,2 %
Cycle 3 (1998-1999) :	$\frac{2\ 192 - 2\ 178}{2\ 192}$	=	$\frac{14}{2\ 192}$	=	0,6 %
Cycle 4 (2000-2001) :	$\frac{2\ 178 - 2\ 143}{2\ 178}$	=	$\frac{35}{2\ 178}$	=	1,6 %
Cycle 5 (2002-2003) :	$\frac{2\ 143 - 2\ 131}{2\ 143}$	=	$\frac{12}{2\ 143}$	=	0,6 %

**Taux d'érosion cumulatifs**

Cycle 2 (1996-1997) :	$\frac{2\ 287 - 2\ 192}{2\ 287}$	=	$\frac{95}{2\ 287}$	=	4,2 %
Cycle 3 (1998-1999) :	$\frac{2\ 287 - 2\ 178}{2\ 287}$	=	$\frac{109}{2\ 287}$	=	4,8 %
Cycle 4 (2000-2001) :	$\frac{2\ 287 - 2\ 143}{2\ 287}$	=	$\frac{144}{2\ 287}$	=	6,3 %
Cycle 5 (2002-2003) :	$\frac{2\ 287 - 2\ 131}{2\ 287}$	=	$\frac{156}{2\ 287}$	=	6,8 %

**8.3. Erreurs relatives à l'enquête**

**8.3.1. Erreurs d'échantillonnage**

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

Les estimations d'une enquête par sondage comportent une erreur d'échantillonnage. De bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. Cette partie du guide décrit les *mesures de l'erreur d'échantillonnage* que Statistique Canada utilise couramment et dont il conseille

vivement l'emploi aux utilisateurs qui produisent des estimations à partir de ce fichier maître.

L'évaluation de l'importance éventuelle des erreurs d'échantillonnage se fonde sur l'erreur-type des estimations calculées d'après les résultats de l'enquête. Étant donné la diversité des estimations que l'on peut tirer d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (c.v.) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation (égale à la racine carrée de la variance de l'estimation) par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons qu'à partir des résultats de l'enquête, on estime que 10,4 % de résidents des établissements de soins pour bénéficiaires internes fument des cigarettes tous les jours et que l'on établit à 0,0094 l'erreur-type de cette estimation. Le coefficient de variation se calcule comme suit :

$$\left( \frac{0,0094}{0,104} \right) \times 100 \% = 9,04 \%$$

Le chapitre 11 contient plus de détails sur le calcul de la variance pour cette enquête. Pour l'interprétation des c.v. et les lignes directrices pour la diffusion, consulter la section 9.4.

### **8.3.2. Erreurs non dues à l'échantillonnage**

Des erreurs qui ne se rapportent pas à l'échantillonnage peuvent se produire à presque toutes les étapes d'une enquête. Il se peut que les instructions pour les intervieweurs ne soient pas suffisamment claires ou soient mal interprétées, que les répondants fassent des erreurs en répondant aux questions, que des réponses soient mal saisies sur le questionnaire ou que des erreurs se produisent au moment du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'*erreurs non dues à l'échantillonnage*.

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

L'effet de la *non-réponse* constitue une source importante d'erreur non due à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. Il y a non-réponse partielle lorsque le répondant refuse de répondre à une question ou qu'il ne peut se souvenir des renseignements demandés. Il y a non-réponse totale lorsque l'intervieweur a été incapable de communiquer avec la personne chargée de répondre par procuration ou que le répondant sélectionné a refusé de participer à l'enquête. Dans le cas du volet établissements de soins de santé de l'ENSP, la non-réponse tant partielle que totale est faible. On traite les cas de non-réponse totale en corrigeant le poids des résidents qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. La méthode employée pour calculer ces poids est décrite au chapitre 10.

#### 8.4. Imputation

On a eu recours à l'imputation pour calculer les valeurs manquantes de l'une des variables du volet établissements de soins de santé de l'ENSP. La variable HSI2DHSI représente l'Indice de l'état de santé (Health Utility Index – HUI3<sup>2</sup>) d'un répondant. Cette mesure de l'état de santé global englobe l'évaluation de la vue, de l'ouïe, de l'élocution, de la mobilité (capacité à se déplacer), de la dextérité (mouvement des mains et des doigts), des sentiments, de la cognition (mémoire et pensée) et de la douleur. La cote globale de l'HUI3, qui peut varier de -0,360 à 1,000, est calculée d'après les réponses à une série de questions sur l'état de santé. Or, il est impossible de calculer cette cote globale si la réponse concernant l'une ou plusieurs de ces composantes est manquante, situation qui s'est produite pour environ 10 % des répondants. Il a été décidé de procéder à l'imputation des valeurs manquantes pour le calcul de l'HUI3 pour le cycle 5 (2002-2003) du volet établissements de soins de santé. On a donc recouru à une forme d'**imputation par la méthode hot deck** pour imputer des valeurs aux composantes manquantes, afin de pouvoir calculer l'HUI3 global pour les personnes concernées. Il convient de souligner que la même méthode a été utilisée aux cycles précédents.

L'HUI3 a été calculé d'après les réponses aux questions de huit attributs de la section sur l'état de santé du questionnaire. On a calculé une cote partielle pour chacun des attributs, puis on a poursuivi les calculs sur ces cotes partielles pour obtenir la cote globale de l'HUI3. L'imputation a été exécutée au niveau des huit cotes partielles plutôt qu'au niveau des questions. Après l'imputation, on a modifié légèrement le programme pour le calcul de la variable dérivée de l'HUI3 afin qu'il choisisse comme données d'entrée les huit valeurs imputées pour la vue, l'ouïe, l'élocution, la mobilité, les sentiments, la cognition, la dextérité et la douleur.

---

<sup>1</sup> Pour plus de renseignements sur le calcul du HUI3, voir la documentation sur les variables dérivées.

L'imputation a été effectuée en deux étapes.

- La première étape correspond à une imputation déterministe. Dans certains cas, même si la personne n'avait pas répondu à la question fournissant la cote partielle, on possédait suffisamment de renseignements pour déduire cette cote avec certitude. On a donc attribué la cote partielle d'après ces renseignements partiels dans tous les cas où il a été jugé approprié de le faire.
- La deuxième étape correspond à une imputation hot deck par donneur pour attribuer les cotes partielles manquantes. La méthode du plus proche voisin a été utilisée pour identifier les donneurs. Pour identifier un plus proche voisin, un HUI3 temporaire a été calculé en ne tenant compte que des cotes partielles ne contenant pas de données manquantes.

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

Cette section du guide décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent des données provenant des fichiers de données de l'enquête. Ces lignes directrices devraient leur permettre de reproduire les nombres publiés par Statistique Canada et de produire aussi des nombres non encore publiés conformes aux lignes directrices établies.

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

Voici des lignes directrices à suivre pour l'arrondissement des estimations tirées du fichier de données :

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



- f) Des estimations non arrondies ne doivent être publiées ni diffusées de toute autre façon en aucune circonstance. Des estimations non arrondies donnent l'impression d'être beaucoup plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour le volet établissements de soins de santé de l'ENSP n'est pas autopondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour tous les répondants. Même pour produire des estimations simples, y compris des tableaux statistiques standard, l'utilisateur doit employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir du présent fichier ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles qui pourraient être produites par Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue que certains progiciels ne prennent pas les poids en considération, ce qui empêche les utilisateurs d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

### **9.2.1. Définitions des catégories d'estimation : nominale par opposition à quantitative**

Deux principaux genres d'estimations ponctuelles des caractéristiques de la population peuvent être produites d'après le fichier de données du volet des établissements de soins de santé de l'ENSP.

#### Estimations de type nominal :

Les estimations de type nominal (aussi appelées estimations d'un agrégat) sont des estimations du nombre ou du pourcentage de personnes qui, dans la population observée, possèdent certaines caractéristiques ou font partie d'une catégorie particulière. Le nombre de personnes qui fument tous les jours est un exemple de ce genre d'estimation.

#### Exemple de question de type nominal :

SMI2\_1 : Actuellement, fumez-vous (...fume-t-il / elle) des cigarettes tous les jours, à l'occasion ou jamais?

- Tous les jours
- À l'occasion
- Jamais

Estimations quantitatives :

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

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

Exemple de question quantitative :

SMI2\_3 : Actuellement, combien de cigarettes fumez-vous (...fume-t-il / elle) chaque jour?

|\_|\_| Nombre de cigarettes

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

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

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

**9.2.3. Totalisation d'estimations quantitatives**

On peut obtenir des estimations de quantités à partir du fichier de données en multipliant la valeur de la variable voulue par le poids de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements concernés. Par exemple, pour obtenir une estimation du nombre *total* de cigarettes que fument chaque jour les personnes qui fument tous les jours, on multiplie la valeur déclarée

à la question SMI2\_3 par le poids de l'enregistrement (WTI2LF), puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMI2\_1 est « tous les jours ».

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

- a) estimer le nombre total de cigarettes que fument par jour les personnes qui fument tous les jours, selon la méthode susmentionnée;
- b) estimer le nombre de personnes qui fument tous les jours en additionnant les poids de tous les enregistrements dont la réponse à la question SMI2\_1 est « tous les jours »;
- c) diviser l'estimation du numérateur (obtenue en « a ») par celle du dénominateur (calculée en « b »).

### 9.3. Lignes directrices pour l'analyse statistique

Le volet établissements de soins de santé de l'ENSP se fonde sur un plan d'échantillonnage à deux degrés, où les établissements sont sélectionnés sans remise. L'utilisation des données provenant de ce type d'enquête pose des difficultés aux analystes, car le choix des méthodes d'estimation et de calcul de la variance dépend du plan d'échantillonnage et des probabilités de sélection.

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

Pour calculer la variance des estimations obtenues avec les données de l'ENSP, il est recommandé d'utiliser la méthode du bootstrap et d'utiliser le programme Bootvar fourni avec les données (voir le chapitre 11).

Dans le cas des différents progiciels statistiques, il est possible, pour de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de variance), de rendre leur application un peu plus juste. Si l'on transforme les poids des enregistrements de façon à ce que le poids moyen soit égal à un (1), les résultats produits par les progiciels courants sont ainsi plus raisonnables. Même s'ils ne reflètent toujours pas la stratification du plan d'échantillonnage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un

poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (personnes) qui contribuent à l'estimateur en question.

#### 9.4. Lignes directrices pour la diffusion

Avant de diffuser ou de publier des estimations tirées du fichier maître ou d'un des sous-ensembles, 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, quel 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 9.A.

**Tableau 9.A : Lignes directrices relatives à la variabilité d'échantillonnage**

Fiabilité de l'estimation	c.v. (%)	Lignes directrices
<b>Acceptable</b>	$0,0 \leq c.v. \leq 16,5$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
<b>Médiocre</b>	$16,5 < c.v. \leq 33,3$	On peut envisager une diffusion générale non restreinte des estimations, avec une mise en garde aux utilisateurs quant à la forte variabilité d'échantillonnage qui y est associée. Les estimations de ce genre doivent être marquées de la <b>lettre E</b> (ou d'une manière similaire).
<b>Inacceptable</b>	$c.v. > 33,3$	Statistique Canada recommande de ne pas publier d'estimations dont la qualité est inacceptable. L'utilisateur qui choisit malgré tout de le faire doit les marquer de la <b>lettre F</b> (ou tout autre identificateur) et y joindre l'avertissement suivant :  « Nous avisons l'utilisateur que ... (préciser 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 que l'on pourrait en tirer ne doivent pas être publiées. L'utilisateur qui choisit de les publier est tenu de publier également le présent avertissement. »

Par définition, on calcule le c.v. en multipliant l'erreur-type (égal à la racine carrée de l'estimation de la variance) par 100 et en divisant le résultat par l'estimation. Consulter le chapitre 11 pour obtenir des renseignements supplémentaires sur le calcul de la variance.

## 10. Pondération

Le présent chapitre décrit le processus de pondération longitudinale pour chacun des sous-ensembles de répondants décrits 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, la définition d'une réponse longitudinale est différente d'une réponse transversale, ce qui nécessite des ajustements différents particuliers à chaque type de non-réponse. Les analystes doivent toujours utiliser les poids longitudinaux pour les études faites à partir des sous-ensembles de répondants. Les poids longitudinaux ont été conçus spécifiquement pour représenter la population cible de 1994-1995. Aux cycles 1 et 2, des fichiers transversaux et longitudinaux ont été produits. Malgré le fait que les membres du panel faisaient partie à la fois des fichiers transversaux et longitudinaux, leurs poids n'étaient pas identiques dans ces deux types de fichier, mais plutôt ajustés afin de bien représenter la population d'intérêt.

Pour le cycle 5, trois ensembles de poids ont été créés : WTI4LS, WTI2LF et WTI2SLF. Le tableau 10.A présente les variables de poids et les variables indicatrices associées avec chaque sous-ensemble. Un membre du panel fait partie d'un sous-ensemble donné lorsque la variable indicatrice de celui-ci est égale à 1.

**Tableau 10.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes**

Sous-ensemble de répondants	Poids	Variable indicatrice
Longitudinal carré	WTI4LS	Aucune, tous les enregistrements
Longitudinal complet	WTI2LF	WFI2LF
Longitudinal complet partagé	WTI2SLF	WFI2SLF

La procédure utilisée pour la pondération longitudinale est en grande partie basée sur la pondération du cycle 1. On trouvera une description complète des méthodes de pondération utilisées en 1994-1995 dans le guide du fichier de microdonnées à grande diffusion de l'ENSP de 1994-1995 : Établissements de soins de santé.

### 10.1. Poids longitudinal carré (WTI4LS)

Le poids longitudinal carré **WTI4LS** est en grande partie basée sur la pondération du cycle 1.

**Probabilité de sélection des établissements au cycle 1 (1994-1995)**

Notation :

- $M_h$  = nombre de lits dans la strate  $h$  (d'après la liste des hôpitaux et des établissements de soins pour bénéficiaires internes)
- $M_{h,i}$  = nombre de lits dans l'établissement  $i$  de la strate  $h$  (d'après la liste des hôpitaux et des établissements de soins pour bénéficiaires internes)
- $n_h$  = nombre d'établissements à sélectionner dans la strate (de taille)  $h$

On a échantillonné les établissements à partir de la base de sondage de 1994 avec probabilité proportionnelle au nombre de lits. Par conséquent, dans la plupart des cas, la probabilité de sélectionner un établissement  $i$  était :

$$n_h \times \frac{M_{h,i}}{M_h}$$

En cas de sélection d'un siège social (voir la section 5.1 pour plus de détails), la probabilité était :

$$n_h \times \frac{M_{h,i}}{M_h} \times P_{h,i,j}$$

où  $P_{h,i,j}$  représente la probabilité qu'un établissement  $j$  relevant du siège social  $i$  soit sélectionné dans la strate  $h$ . Dans le cas du plus grand établissement relevant de  $i$ ,  $P_{h,i,j} = 1$ .

Dans le cas des autres établissements  $j$  :

$$P_{h,i,j} = \frac{M_{h,i,j}}{\sum_{j \in i'} M_{h,i,j}}$$

où  $i'$  représente tous les établissements relevant du siège social  $i$ , à l'exclusion du plus grand.

**Calcul des poids de base des établissements au cycle 1 (1994-1995) et correction pour la non-réponse au niveau des établissements**

Le poids d'un établissement correspond au nombre d'établissements que représente l'établissement échantillonné. Le **poids de base de l'établissement** est égal à l'inverse de la probabilité de sélectionner cet établissement. Cependant, comme la non-réponse est possible à ce niveau, une correction est nécessaire pour tenir compte des établissements qui refusent de participer. Dans le cas où des interviews n'ont pu être menées dans un établissement sélectionné faisant partie du champ de l'enquête, on apporte une correction aux poids des autres établissements appartenant à la même strate de taille pour tenir compte de l'établissement non-répondant. Cette correction correspond à :

$$\frac{\text{nombre d'établissements répondants et non – répondants}}{\text{nombre d'établissements répondants}}$$

La multiplication du poids de base de l'établissement par ce facteur de correction donne le **poids final de l'établissement au cycle 1**.

**Probabilité de sélection des résidents au cycle 1 (1994-1995)**

Notation :

$L_{h,i}$  = nombre réel de résidents de longue durée dans la strate  $h$ , établissement  $i$  (obtenu au moment de la première visite)

$r_{h,i}$  = nombre de résidents répartis à sélectionner dans la strate  $h$ , établissement  $i$

Une fois un établissement sélectionné, chaque résident de cet établissement avait une probabilité égale d'être sélectionné, probabilité définie par :

$$\left\{ \begin{array}{ll} \frac{r_{h,i}}{L_{h,i}} & \text{si } L_{h,i} \geq r_{h,i} \\ 1 & \text{si } L_{h,i} < r_{h,i} \end{array} \right.$$

**Calcul des Poids de base des résidents au cycle 1 (1994-1995) et correction pour la non-réponse au niveau des résidents**

Pour calculer le poids de base applicable à un résident, on multiplie le poids final de l'établissement par l'inverse de la probabilité de sélection d'un résident dans l'établissement. Encore une fois, comme la non-réponse est aussi possible à ce niveau, des corrections sont nécessaires pour tenir compte des résidents qui refusent de répondre (au cycle 1). La correction supplémentaire est apportée au poids de base des personnes pour tenir compte de la non-réponse des résidents :

$$\frac{\text{somme des poids des résidents répondants et non – répondants}}{\text{somme des poids des résidents répondants}}$$

La multiplication du poids de base du résident par cet ajustement pour la non-réponse donne le **poids longitudinal carré (WTI4SL)**.

**10.2. Poids longitudinal complet (WTI2LF)**

Le sous-ensemble longitudinal complet inclut seulement les membres du panel qui ont fourni une réponse complète, c'est-à-dire les membres qui ont eu un statut « institutionnalisé », « décédé » ou « ménage », à chaque cycle. Les membres du panel



exclus de ce sous-ensemble ont donc été non-répondants, c'est-à-dire qu'ils ont eu un statut « non-réponse », à un certain point durant les cinq premiers cycles de l'enquête, et leur poids doit donc être redistribué pour compenser pour la non-réponse.<sup>3</sup>

Le point de départ est le poids longitudinal carré (WTI4LS) et des ajustements pour la non-réponse sont faits. Un ajustement différent est effectué pour la non-réponse à chacun des cycles et ces ajustements sont cumulatifs d'un cycle à l'autre. Par exemple, pour obtenir le poids du cycle 5, les ajustements pour la non-réponse aux cycles 2, 3, 4 et 5 sont appliqués successivement au poids WTI4LS.

Les ajustements nécessaires en vue d'obtenir le poids longitudinal complet du cycle 5 sont décrits ci-dessous.

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

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

$$\frac{\text{somme des poids des résidents répondants et non – répondants}}{\text{somme des poids des résidents répondants}}$$

Cet ajustement est fait au niveau des classes de non-réponse, à l'échelle du Canada. Les classes sont déterminées à l'aide de l'algorithme CHAID (« Chi-Square Automatic Interaction Detector »). Cet algorithme est offert avec le logiciel *Knowledge Seeker* (développé par ANGOSS Software International Limited).

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

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

$$\frac{\text{somme des poids des résidents répondants et non – répondants}}{\text{somme des poids des résidents répondants}}$$

Cette correction est apportée pour chaque profil de réponse longitudinal (la variable LONGPAT).

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

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

---

<sup>3</sup> Voir la section 7.6 pour la définition du profil de réponse longitudinal et celle du statut de réponse complet pour l'ENSP.

$$\frac{\text{somme des poids des résidents répondants et non – répondants}}{\text{somme des poids des résidents répondants}}$$

Cette correction est encore une fois apportée pour chaque profil de réponse longitudinal (la variable LONGPAT).

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

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

$$\frac{\text{somme des poids des résidents répondants et non – répondants}}{\text{somme des poids des résidents répondants}}$$

Cette correction est encore une fois apportée pour chaque profil possible de réponse longitudinale (la variable LONGPAT).

**Ajustement 5 : Post-stratification**

Puisque le nombre total de personnes au Canada vivant en établissements de soins de santé est inconnu (selon la définition d'établissement de l'ENSP), il est impossible d'effectuer une post-stratification en se basant sur ces totaux. Cependant, une post-stratification est faite en utilisant les poids longitudinal carré (WTI4LS). La post-stratification est faite en deux étapes : premièrement à l'échelle de cinq régions, puis pour chaque catégorie type établissement/âge-sexe.

Le poids final pour une réponse longitudinale complète, **WTI2LF**, est calculé en prenant d'abord le poids longitudinal carré (**WTI4LS**) et en le multipliant ensuite par les ajustements 1 à 5.

**10.3. Poids longitudinal complet partagé (WTI2SLF)**

Le sous-ensemble complet partagé inclut les membres du panel qui ont accepté de partager l'information recueillie lors de toutes les interviews menées dans le cadre de l'ENSP avec les ministères provinciaux de la santé et Santé Canada et ont fourni une réponse complète aux cycles 1, 2, 3, 4 et 5. 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 en établissement de soins de santé.

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

*Somme des poids des membres longitudinaux répondants aux cycles 1 à 5,  
dans une classe type d'établissement / profil de réponse longitudinal / âge-sexe*  
*Somme des poids des membres longitudinaux répondants aux cycles 1 à 5 et ayant  
accepté de partager, dans une classe type d'établissement / profil de réponse  
longitudinal / âge-sexe*

Il est à noter que quelques profils de réponse longitudinaux ont été regroupés afin de produire des ajustements plus stables. Le regroupement a été fait pour quelques classes type d'établissement/âge-sexe qui n'avaient que quelques observations dans certains profils de réponse. Dans chaque cas, le profil de réponse problématique a été regroupé avec un autre profil de la même classe type d'établissement /âge-sexe de telle sorte que la somme des poids corresponde toujours, sauf exception, aux totaux de la population. Le regroupement a aussi été fait en respectant le plus possible le statut de réponse au cycle 5. Par exemple, si le cas problématique avait un statut de réponse « décédé » au cycle 5, il a été regroupé avec un profil de réponse longitudinal qui indiquait également un statut de réponse « décédé » au cycle 5. Pour quelques cas, un tel regroupement n'a pu être fait alors, plutôt que de regrouper des profils de réponse, on a regroupé des groupes d'âge-sexe.

Le poids final pour le sous-ensemble complet partagé, **WTI2SLF**, est obtenu en multipliant le poids longitudinal complet, **WTI2LF**, par cet ajustement. En général, puisque cet ajustement est fait en respectant les classes utilisées lors de l'ajustement pour la post-stratification, aucune post-stratification supplémentaire n'est nécessaire.

## **11. Calcul de la variance**

La méthode utilisée pour calculer la variance des estimations au cycle 5 est la même que celle utilisée au cycle 3 et 4 (et diffère de celle employée au cycle 2). Depuis le cycle 3, la méthode du bootstrap est utilisée. Il s'agit de la même méthode employée pour le volet ménages de l'ENSP. Cette méthode est expliquée dans cette section.

Un programme de calcul de variance, rédigé en SAS ou SPSS, est fourni avec le fichier de données. Il peut être utilisé pour obtenir des estimations de variance précises pour des statistiques telles que des totaux et ratios, et pour des analyses plus complexes comme des régressions. Un guide d'utilisation est aussi fourni avec le programme.

### **11.1. Méthode du bootstrap**

Les plans d'échantillonnage pour les enquêtes sur la santé sont complexes. Comme le calcul de la variance pour de tels plans ne peut être fait à l'aide de simples formules, on a recours à une méthode de rééchantillonnage pour calculer des estimations de la variance. La méthode du bootstrap consiste à sous échantillonner l'échantillon initial. À l'intérieur de chaque strate, on choisit un échantillon aléatoire simple (ÉAS), avec remise, de  $n-1$  grappes parmi les  $n$  grappes de la strate. Un certain nombre  $B$  de nouveaux échantillons (ou répétitions) sont ainsi créés. La même estimation est ensuite calculée pour chacun des  $B$  échantillons, ce qui permet d'obtenir  $B$  estimations différentes. L'estimation de la variance correspond à la simple variance entre les  $B$  estimations. Afin d'obtenir chacune des  $B$  estimations, un poids spécifique à chaque échantillon est nécessaire. Pour chaque échantillon ÉAS, le facteur de pondération est donc recalculé pour chaque enregistrement dans la strate. Ces  $B$  poids, les poids bootstrap, ont été produits et sont disponibles avec les données (pour le sous-ensemble longitudinal complet seulement, en raison de la faible différence entre le nombre d'observations des sous-ensembles longitudinaux complet et carré).

En résumé, la méthode du bootstrap consiste à :

- A) Calculer d'abord une estimation (total, rapport, etc.) en utilisant le poids final inclus dans le fichier de données. Cette estimation est l'estimation ponctuelle.
- B) Calculer ensuite cette même estimation en utilisant cette fois chacun des  $B$  poids bootstrap contenus dans le fichier de poids bootstrap. On obtient ainsi  $B$  estimations (du total, du rapport, etc.).
- C) Finalement, calculer la variance de ces  $B$  estimations. Cette variance est l'estimation de la variance de l'estimation ponctuelle calculée en A.

Les règles de confidentialité et seuils de diffusion s'appliquent aussi pour les estimations de variance provenant de la méthode du bootstrap.

## **11.2. Estimation de la variance avec le programme BOOTVAR.SAS (.SPS)**

Le programme BOOTVAR.SAS (.SPS) permet d'estimer la variance. Ce programme est fourni avec le fichier de données. Il permet de calculer des estimations de variance pour des totaux, des ratios, des différences entre des ratios, des paramètres de régression linéaire et logistique et des modèles linéaires généralisés. L'utilisateur doit s'assurer que les références au nom des fichiers sont cohérentes lors de l'utilisation du programme. Pour plus de détails sur l'utilisation de BOOTVAR.SAS (.SPS), consulter le guide d'utilisation fourni avec le programme.

## 12. Utilisation du fichier maître longitudinal

### 12.1. Utilisation des poids longitudinaux

Dans les trois premiers cycles, des fichiers correspondant à un sous-ensemble des 2 287 membres du panel longitudinal ont été créés. Au cycle 5, tout comme au cycle 4, un seul fichier a été créé. Le fichier maître du cycle 5 contient trois poids correspondant à trois sous-ensembles de répondants. Des variables indicatrices ont été créées afin d'identifier les enregistrements appartenant à chacun des sous-ensembles de répondants.

**Tableau 12.A : Sous-ensembles de répondants, variables de poids et variables indicatrices correspondantes**

Sous-ensemble de répondants	Poids	Variable indicatrice
Longitudinal carré	WTI4LS Poids maître - longitudinal	Aucune, tous les enregistrements
Longitudinal complet	WTI2LF Poids maître – réponse longitudinale complète	WFI2LF
Longitudinal complet partagé	WTI2SLF Poids partagé – réponse longitudinale complète	WFI2SLF

Les enregistrements qui ne font pas partie d'un sous-ensemble de répondants particulier ont une variable indicatrice égale à 0 et un poids laissé vide pour ce sous-ensemble de répondants particulier. Pour constituer le sous-ensemble de répondants d'intérêt, sélectionner les enregistrements où la variable indicatrice appropriée est égale à 1.

Le poids WTI4LS est appelé le poids maître – longitudinal, aussi appelé « poids carré » et s'applique au sous-ensemble « carré » qui comprend les 2 287 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 WTI2LF est appelé le poids maître – réponse longitudinale complète, aussi appelé « poids longitudinal complet » et s'applique aux 2 131 enregistrements qui forment le sous-ensemble « complet ».

Le poids WTI2SLF est appelé le poids partagé – réponse longitudinale complète, aussi appelé « poids longitudinal complet partagé » et s'applique aux 2 067 enregistrements du sous-ensemble « complet partagé ».

**NOTA** : pour l'ENSP – volet établissements de soins de santé, la définition de “complet” inclut les réponses partiellement complétées ainsi que les réponses entièrement complétées.

## **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. Le chapitre 11 contient de l'information sur la méthode bootstrap à appliquer pour le calcul de l'estimation de la variance. Des poids sont disponibles pour les sous-ensembles complets (maître et partagé) seulement.

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 9 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**

En 1996-1997, on a adopté une convention pour nommer les variables de l'ENSP qui permet aux utilisateurs des données de faire facilement référence à des données similaires provenant d'autres périodes de collecte et aux divers volets de l'ENSP. Les exigences qui suivent ont été appliquées : limiter les noms des variables à huit caractères au plus pour qu'il soit facile de les utiliser avec les logiciels d'analyse ; identifier le cycle de l'enquête (1994-1995, 1996-1997, 1998-1999, 2000-2001, 2002-2003) dans le nom et permettre de repérer facilement les variables conceptuellement identiques d'un cycle à l'autre. Par exemple, des données conceptuellement identiques sur l'usage du tabac ont été recueillies en 1994-1995 et en 1996-1997. Les noms des variables sur l'usage du tabac ne devraient différer qu'en ce qui concerne la position réservée dans le nom à l'année de référence particulière durant laquelle les données ont été recueillies. Cette convention sera appliquée pendant toute la durée de l'enquête longitudinale et sera adoptée pour tous les volets de l'ENSP, à savoir le volet ménages, le volet établissements de soins de santé, le volet nord et les suppléments.

### 12.3.1. Structure 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 / volet
- Position 4 : Année/cycle durant lequel apparaît la variable
- Position 5 : Type de variable (c.-à-d. questionnaire, codée, dérivée, etc.)
- Positions 6 et 8 : Numéro/nom de la variable tiré du questionnaire

Par exemple, la signification du nom de variable DHI2DAGE est :

- DH : présent dans la section du questionnaire sur la démographie et la composition du ménage;
- I : questions qui font partie du volet établissements de soins de santé;
- 2 : figure dans le cycle 5 (2002-2003);
- D : variable dérivée; et
- AGE : nom de la variable.

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

On se sert des valeurs suivantes pour le nom de la section du questionnaire :

AL	Alcool	HW	Taille et poids
AM	Administration de l'enquête	IN	Revenu
CC	Problèmes de santé chroniques	IP	Politiques de l'établissement
DG	Consommation de médicaments	RA	Limitation des activités
DH	Démographie et ménage	SD	Renseignements socio-démographiques
ED	Niveau de scolarité	SM	Usage du tabac
FL	Équilibre et chutes	SP	Identificateurs d'échantillon (méthodologie)
GH	État de santé général	SS	Soutien social
HC	Utilisation des soins de santé	WT	Poids
HS	État de santé		



### 12.3.3. Position 3 : Type d'enquête / volet

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

### 12.3.4. 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)

### 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., code de la Classification type des industries (CTI)).
D	Variable dérivée	Variable calculée d'après une ou plusieurs variables recueillies ou codées, ordinairement pendant le traitement au Bureau central (p. ex., indice de l'état de santé).
F	Variable indicatrice	Variable calculée à partir d'une ou de plusieurs variables recueillies (comme une variable dérivée), mais ordinairement par l'application informatique, aux fins de son utilisation ultérieure durant l'entrevue (p. ex., indicateur de travail).
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. Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent au nom qui figure sur le questionnaire. On utilise des nombres dans la mesure du possible : Q1 devient 1. Dans le cas des questions de type « Cochez toutes les réponses pertinentes », on se sert des lettres correspondant à chaque catégorie possible de réponse : Q1 (cochez toutes les réponses pertinentes) devient 1A, 1B, 1C, etc. On représente les variables démographiques utilisées fréquemment par les analystes par trois lettres indicatrices plutôt que par un numéro de question; par exemple, « âge » est représenté par DHI2DAGE en 2002-2003. 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. À titre d'exemple, mentionnons le cas des questions sur l'état de santé général de l'Enquête sur la promotion de la santé. Ces questions ont été réparties en trois sections aux fins de leur intégration au questionnaire et les noms des variables correspondantes reflètent cette situation, le caractère de la position 6 indiquant dans quelle section la question figure.

## **12.4. Accès aux fichiers maîtres**

### **12.4.1. Fichiers de microdonnées**

Le souci de protéger la confidentialité des données longitudinales écarte la possibilité de diffuser un fichier de microdonnées à grande diffusion (FMGD) de nature longitudinale. Cependant, l'accès direct sur place aux fichiers maîtres de microdonnées du volet établissements de soins de santé de l'ENSP est 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 avec un employé de Statistique Canada à temps plein dans chaque centre. Les chercheurs y poursuivent leur travail en vertu de la Loi sur la statistique comme tout employé de Statistique Canada le ferait. Vous pouvez trouver de l'information supplémentaire sur le site du Programme des centres de données de recherche au :

[http://www.statcan.ca/francais/rdc/index\\_f.htm](http://www.statcan.ca/francais/rdc/index_f.htm).

Des FMGD transversaux sont disponibles pour les deux premiers cycles du volet établissements de soins de santé de l'ENSP et procurent un accès élargi aux fichiers transversaux de l'enquête. Les FMGD de l'ENSP – volet établissements de soins de santé 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 FMGD transversaux du volet établissements de soins de santé de l'ENSP des cycles 1 et 2 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.

### **12.4.2. Totalisations personnalisées**

Afin d'accéder aux fichiers maîtres de l'enquête, une approche offerte à tous les clients est de demander au personnel spécialisé des Services personnalisés à la clientèle de la Division de la statistique de la santé de produire des totalisations personnalisées. Ce service permet aux utilisateurs qui n'ont pas la connaissance des progiciels statistiques d'obtenir des totalisations personnalisées. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Ce service est offert contre remboursement des frais. Vous pouvez avoir accès à ce service en écrivant à [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca), en téléphonant au 1-613-951-1746 ou via télécopie au 1-613-951-0792.