

**ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION  
COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE 1996-1997  
TABLE DES MATIÈRES**

<b>1. INTRODUCTION.....</b>	<b>3</b>
<b>2. CONTEXTE .....</b>	<b>4</b>
<b>3. OBJECTIFS .....</b>	<b>5</b>
<b>4. CONTENU DE L'ENQUÊTE.....</b>	<b>6</b>
4.1 CRITÈRES .....	6
4.2 RÉVISIONS DU CONTENU POUR L'ENQUÊTE DE 1996-1997.....	6
<b>5. PLAN D'ÉCHANTILLONNAGE .....</b>	<b>8</b>
5.1 PLAN D'ÉCHANTILLONNAGE DE 1996-1997.....	8
5.1.1 <i>Suivi du panel longitudinal complet.....</i>	<i>8</i>
5.2 BASE DE SONDAGE DE 1996.....	11
5.2.1 <i>Identification des nouveaux établissements dans la base de sondage de 1996.....</i>	<i>12</i>
5.3 STRATIFICATION ET RÉPARTITION.....	12
5.3.1 <i>Stratification et répartition de l'échantillon de 1994-1995 .....</i>	<i>12</i>
5.3.2 <i>Stratification et répartition de l'échantillon de 1996-1997 .....</i>	<i>13</i>
5.4 SÉLECTION DE L'ÉCHANTILLON .....	14
5.4.1 <i>Sélection de l'échantillon de 1994-1995.....</i>	<i>14</i>
5.4.2 <i>Sélection de l'échantillon de 1996-1997 pour les établissements du premier cycle .....</i>	<i>14</i>
5.4.3 <i>Sélection de l'échantillon de 1996-1997 de nouveaux établissements au deuxième cycle .....</i>	<i>15</i>
<b>6. COLLECTE DES DONNÉES.....</b>	<b>16</b>
6.1 CONCEPTION DU QUESTIONNAIRE ET MÉTHODE DE COLLECTE DES DONNÉES.....	16
6.2 INTERVIEWS.....	16
6.3 SURVEILLANCE ET CONTRÔLE .....	16
6.4 NON-RÉPONSE À L'ENSP.....	17
<b>7. TRAITEMENT DES DONNÉES.....</b>	<b>18</b>
7.1 SAISIE DES DONNÉES .....	18
7.2 VÉRIFICATION .....	18
7.3 CODAGE.....	18
7.4 CRÉATION DE VARIABLES DÉRIVÉES .....	19
7.5 PONDÉRATION .....	19
7.6 SUPPRESSION DE RENSEIGNEMENTS CONFIDENTIELS .....	19
<b>8. QUALITÉ DES DONNÉES .....</b>	<b>21</b>
8.1 TAUX DE RÉPONSE .....	21
8.2 ERREURS RELATIVES À L'ENQUÊTE .....	22
8.2.1 <i>Erreurs d'échantillonnage .....</i>	<i>22</i>
8.2.2 <i>Erreurs non dues à l'échantillonnage.....</i>	<i>23</i>
8.3 IMPUTATION .....	24

<b>9.</b>	<b>LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION.....</b>	<b>26</b>
9.1	LIGNES DIRECTRICES POUR L'ARRONDISSEMENT .....	26
9.2	LIGNES DIRECTRICES POUR LA PONDÉRATION DE L'ÉCHANTILLON EN VUE DE LA TOTALISATION .....	27
9.2.1	<i>Définitions des catégories d'estimation : nominale par opposition à quantitative.....</i>	27
9.2.2	<i>Totalisation d'estimations de type nominal.....</i>	28
9.2.3	<i>Totalisation d'estimations quantitatives.....</i>	29
9.3	LIGNES DIRECTRICES POUR L'ANALYSE STATISTIQUE .....	29
9.4	LIGNES DIRECTRICES POUR LA DIFFUSION.....	30
<b>10.</b>	<b>PONDÉRATION.....</b>	<b>32</b>
10.1	PROBABILITÉ DE SÉLECTION DES ÉTABLISSEMENTS EN 1994-1995 .....	32
10.1.1	<i>Calcul des poids de 1994-1995 pour les établissements et corrections.....</i>	33
10.1.2	<i>Poids de base de 1994-1995 pour les établissements.....</i>	33
10.1.3	<i>Correction pour la non-réponse au niveau des établissements de 1994-1995.....</i>	33
10.2	PROBABILITÉ DE SÉLECTION D'UN NOUVEL ÉTABLISSEMENT EN 1996-1997 .....	33
10.3	PROBABILITÉ DE SÉLECTION D'UN RÉSIDENT.....	34
10.3.1	<i>Probabilité de sélectionner un résident dans les établissements sélectionnés en 1994-1995.....</i>	34
10.3.2	<i>Probabilité de sélectionner un résident dans les établissements nouvellement sélectionnés.....</i>	35
10.3.3	<i>Poids de base pour les personnes.....</i>	36
10.3.4	<i>Correction pour les décès.....</i>	36
10.3.5	<i>Correction pour la non-réponse au niveau des personnes.....</i>	38
<b>11.</b>	<b>CALCUL DE LA VARIANCE .....</b>	<b>40</b>
11.1	EXÉCUTION DES PROGRAMMES DE CALCUL DE LA VARIANCE .....	41
11.1.1	<i>Exécution du programme SAS de calcul de la variance.....</i>	41
11.1.2	<i>Exécution du programme SPSS de calcul de la variance.....</i>	43
<b>12.</b>	<b>UTILISATION DU FICHIER .....</b>	<b>45</b>
12.1	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES .....	45
12.1.1	<i>Structure élémentaire des noms des variables.....</i>	45
12.1.2	<i>Positions 1 et 2 : Nom de la variable / section du questionnaire .....</i>	46
12.1.3	<i>Position 3 : Type d'enquête / composante.....</i>	46
12.1.4	<i>Position 4 : Variable de l'année de référence / du cycle.....</i>	47
12.1.5	<i>Position 5 : Type de variable.....</i>	47
12.1.6	<i>Positions 6 à 8 : Nom de la variable .....</i>	48
12.2	ACCÈS AUX FICHIERS MAÎTRES .....	48

**Annexe A: Questionnaire de 1996-1997**

**Annexe B: Cliché d'enregistrement**

**Annexe C: Dictionnaire des données**

**Annexe D: Variables dérivées**

## **1. Introduction**

La composante des établissements de santé de l'Enquête nationale sur la santé de la population (ENSP) est la première enquête nationale longitudinale menée auprès des résidents des établissements de santé au Canada. Le deuxième cycle de collecte des données a débuté à l'automne 1996 par le suivi des résidents interrogés en 1994-1995 lors du premier cycle de l'enquête. Pour remédier au taux d'érosion élevé dû au décès de certains membres de l'échantillon de départ, on a sélectionné un échantillon supplémentaire de résidents dans les établissements qui avaient participé à l'enquête en 1994, ainsi qu'un échantillon de nouveaux établissements. Cette mesure permet de produire des estimations transversales représentatives en plus des estimations longitudinales.

Le premier suivi des membres du panel longitudinal a eu lieu en octobre et novembre 1996. Les interviews ont été menées de décembre 1996 à avril 1997, et le dépistage des personnes qui avaient déménagé s'est poursuivi jusqu'à juin 1997. Le présent guide a pour but de faciliter l'utilisation du fichier de microdonnées transversales de 1996-1997 provenant des résultats de l'enquête.

Toute question au sujet de l'ensemble de données ou de son utilisation doit être adressée à :

Soutien technique/renseignements généraux sur les données :

Service de renseignements sur les produits électroniques 1-800-949-9491

Totalisations spéciales/renseignements généraux sur les données : 1-613-951-1746

Services personnalisés à la clientèle,  
Division des statistiques sur la santé

Internet : [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca)

Renseignements sur le téléaccès aux fichiers maîtres :

1-613-951-1653

Colette Koeune, ENSP,

Division des statistiques sur la santé

Internet : [nphs@statcan.ca](mailto:nphs@statcan.ca)

Télécopieur : 1-613-951-4198

Renseignements sur le contenu de l'enquête :

Mario Bédard, ENSP,

Division des statistiques sur la santé

1-613-951-8933

Internet : [bedardm@statcan.ca](mailto:bedardm@statcan.ca)

Télécopieur : 1-613-951-4198

## **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 broser 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 une composante spéciale, qui couvre les résidents des établissements 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, professionnels 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 de la composante des établissements 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 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 santé. Le questionnaire 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 l'enquête de 1996-1997**

Le questionnaire a subi peu de changements de 1994-1995 à 1996-1997. Voici la liste des questions qui ont été modifiées, ajoutées ou supprimées :

- Une étiquette portant les renseignements démographiques a été apposée sur le questionnaire en vue de la confirmation de ces renseignements durant l'interview.

- Les questions sur la politique de l'établissement en ce qui concerne le partage d'une chambre par des époux et les visites d'animaux familiers ont été supprimées. Des questions sur le genre d'activités organisées à l'intention des résidents, sur celles organisées à l'intention des membres de leur famille et sur la préparation d'un testament biologique ont été ajoutées au questionnaire de 1996-1997.
- Une question permettant d'évaluer l'agilité de la personne interrogée a été ajoutée à la section sur l'état de santé.
- La liste des problèmes de santé chroniques a été révisée afin de la rendre plus comparable à la liste du questionnaire de la composante des ménages et qu'elle reflète mieux les problèmes de santé qui, d'après les résultats de l'enquête de 1994-1995, affectent les résidents des établissements de santé.
- Une question visant à déterminer si la personne interrogée a un téléphone dans sa chambre a été ajoutée pour le deuxième cycle.
- La section sur la limitation des activités a été modifiée afin de la rendre plus comparable à celle du questionnaire de la composante des ménages. Au premier cycle, on a recueilli et codé des renseignements sur quatre états entraînant une limitation des activités; au deuxième cycle, la liste a été réduite à deux.
- La section sur le soutien social a subi quelques révisions; des questions ont été ajoutées pour déterminer si le résident a pris part à des activités personnalisées avec un membre du personnel et à quelles activités il a participé à l'extérieur de l'établissement. On a aussi posé des questions sur la fréquence des contacts téléphoniques avec des amis ou des parents. La question concernant la souplesse de l'horaire quotidien a été supprimée.
- Les catégories de la question sur le niveau de scolarité ont été modifiées pour les rendre plus comparables aux catégories de la variable dérivée établies pour la composante des ménages.
- En plus d'obtenir auprès de l'établissement des renseignements sur les médicaments consommés, on a recueilli des renseignements sur la taille et le poids des résidents qui ont participé à l'enquête (données qui ne figurent pas dans le fichier de microdonnées à grande diffusion) et on a déterminé s'ils avaient été transférés dans un établissement de soins de courte durée pour une période de moins de 21 jours au cours des 12 mois qui ont précédé l'interview.

## **5. Plan d'échantillonnage**

La population cible de la composante des établissements de santé de l'ENSP comprend les résidents des établissements de soins de longue durée de toutes les provinces, sauf ceux des territoires, des réserves indiennes et des bases des Forces canadiennes. On a créé une liste d'établissements de soins de longue durée que l'on a stratifiée selon la région géographique, la catégorie d'établissement et la taille de l'établissement. À partir de cette liste, on a sélectionné un échantillon d'établissements, puis, dans chaque établissement, un échantillon subséquent de résidents. 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 santé.

### **5.1 Plan d'échantillonnage de 1996-1997**

Pour le premier cycle de la composante des établissements de santé de l'ENSP, on a créé l'échantillon en commençant par sélectionner des établissements, puis des résidents dans ces établissements. Pour le deuxième cycle, on fait la distinction entre l'échantillon sélectionné aux fins de l'étude longitudinale et celui sélectionné aux fins de l'étude transversale.

L'échantillon transversal comprend les membres de l'échantillon longitudinal suivis au deuxième cycle, ainsi que la sélection d'un échantillon transversal supplémentaire. L'échantillon longitudinal de 1996-1997 comprend tous les répondants longitudinaux choisis au cours du premier cycle qui ont répondu au questionnaire en 1994-1995. On s'attend à ce que l'érosion de l'échantillon soit beaucoup plus forte pour les personnes vivant en établissement que pour celles qui font partie de la population résidant dans des ménages (voir la section 5.1.1 plus bas). L'érosion de l'échantillon de résidents des établissements de santé est due principalement au décès d'une partie des membres du panel longitudinal. Comme la réduction de l'effectif de l'échantillon est considérable, la variance des estimations pourrait augmenter fortement si on ne sélectionnait aucun échantillon supplémentaire. De plus, il faut sélectionner un échantillon des personnes nouvellement admises dans les établissements qui ont participé au premier cycle de l'enquête, ainsi qu'un échantillon supplémentaire de nouveaux établissements pour assurer la représentativité de l'échantillon transversal. Les procédures de suivi du panel longitudinal sont décrites ci-dessous.

#### **5.1.1 Suivi du panel longitudinal complet**

Au deuxième cycle, les données ont été recueillies en deux phases. Lors de la première phase, les intervieweurs se sont de nouveau rendus dans les



établissements de santé sélectionnés au premier cycle afin de déterminer si les personnes choisies pour faire partie du panel longitudinal vivaient encore dans l'établissement ou si elles avaient déménagé ou étaient décédées. Les administrateurs ou les personnes-ressources de tous les établissements de santé sélectionnés dans l'échantillon du premier cycle ont été contactés en octobre et en novembre 1996. L'intervieweur leur a demandé de remplir un questionnaire papier sur la situation des membres du panel longitudinal (pour obtenir la date de décès des personnes décédées ou des renseignements permettant de dépister celles qui avaient déménagé) et de dresser la liste de tous les résidents de l'établissement bénéficiant de soins de longue durée en indiquant la date de leur admission. Puis, l'intervieweur a rempli un formulaire de renseignements de sélection donnant le nombre de résidents recevant des soins de longue durée admis avant le 1<sup>er</sup> avril 1995, le nombre admis le 1<sup>er</sup> avril 1995 ou après cette date, le nombre de membres du panel longitudinal qui vivaient encore dans l'établissement et le nombre qui avait quitté l'établissement. Cette date limite a été choisie parce que le 31 mars 1995 était le dernier jour de la collecte des données du premier cycle. Ces renseignements ont permis d'identifier les résidents nouvellement admis. Les résultats de la première phase ont montré que plus du tiers des personnes sélectionnées au départ pour faire partie du panel longitudinal ne vivaient plus dans l'établissement où elles résidaient au premier cycle. De celles-ci, presque les trois quarts étaient décédées et le reste avait déménagé (en fait, on s'est aperçu plus tard, durant le dépistage, que certaines personnes qui avaient déménagé étaient aussi des membres du panel qui étaient décédés). Tous les renseignements ont été envoyés au bureau central pour permettre la sélection de l'échantillon supplémentaire pour le fichier transversal de 1996-1997.

L'échantillon transversal de 1996-1997 comprend les établissements observés au premier cycle encore en exploitation, ainsi qu'un échantillon supplémentaire de nouveaux établissements sélectionnés au deuxième cycle. Le tableau qui suit énonce les règles appliquées pour déterminer si un membre du panel longitudinal faisait partie ou non du fichier transversal de 1996-1997.

**Tableau 1**

<b>Situation du membre du panel longitudinal</b>	<b>Mesure prise</b>
Vivait encore dans l'établissement du cycle 1	Les membres du panel longitudinal qui vivaient encore dans l'établissement où ils résidaient au premier cycle ont été interviewés et sont inclus dans le fichier transversal.
Avait déménagé dans un autre établissement	Les membres du panel longitudinal qui avaient déménagé dans un autre établissement ont été interviewés (sauf s'ils avaient déménagé dans un établissement correctionnel), mais ils ne sont pas inclus dans le fichier transversal. Les données à leur sujet apparaissent uniquement dans le fichier longitudinal de la composante des établissements de santé.
Avait déménagé dans un ménage privé	Les membres du panel longitudinal qui avaient déménagé dans un ménage privé ont été interviewés dans le cadre de la composante des ménages de l'ENSP. Ils ne sont pas inclus dans le fichier transversal, mais les données qui les concernent figurent dans le fichier longitudinal de la composante des établissements de santé.
Décédé	On a confirmé, par couplage avec la Base canadienne de données sur la mortalité, le décès de tout membre du panel longitudinal reconnu comme étant décédé au moment de l'interview du deuxième cycle. Les membres décédés du panel longitudinal font partie du fichier longitudinal de la composante des établissements de santé, mais ne figurent pas dans le fichier transversal.

Les autres catégories pour un membre du panel longitudinal sont les suivantes : déménagé dans les Territoires du Nord-Ouest ou au Yukon, déménagé dans une réserve indienne ou une base des Forces canadiennes et déménagé à l'extérieur du Canada (de façon temporaire ou permanente). Pour l'échantillon de la composante des établissements de santé, aucune de ces situations ne s'est présentée au deuxième cycle. Ces genres de déménagements sont généralement observés pour les membres de la population résidant dans des ménages privés plutôt que pour les personnes résidant en établissement.

## **5.2 Base de sondage de 1996**

On a créé la base de sondage de 1996 d'après la liste de 1995 des établissements de soins pour bénéficiaires internes établie par l'Institut canadien d'information sur la santé (ICIS) et d'après la liste de 1993 des hôpitaux tenue à jour par la Division des statistiques sur la santé (DSS) de Statistique Canada. On a demandé aux ministères provinciaux de la Santé de vérifier ces listes et de les mettre à jour au besoin afin de s'assurer de leur exactitude. Les établissements ont été classés d'après la catégorie principale de soins qu'ils fournissent et on n'a retenu que ceux prodiguant des soins de longue durée<sup>1</sup>. Sur la liste des établissements de soins pour bénéficiaires internes, on a retenu uniquement les établissements qui fournissent des soins de longue durée aux personnes âgées, aux enfants présentant des troubles affectifs, ainsi qu'aux personnes présentant des troubles du développement, un handicap physique ou des troubles mentaux. Les établissements sélectionnés sur la liste des hôpitaux incluent les hôpitaux généraux comprenant des unités de soins de longue durée, les établissements de soins prolongés et les centres de réadaptation, ainsi que les hôpitaux spécialisés comprenant des unités de soins de longue durée, comme les hôpitaux pédiatriques et les hôpitaux psychiatriques. Le nombre de lits de soins de longue durée était connu pour chaque établissement.

On a limité la population étudiée aux établissements comptant au moins quatre lits et offrant des soins de longue durée aux personnes qui ont des problèmes de santé. La liste d'établissements de soins pour bénéficiaires internes dressée par l'ICIS ne comprenait aucun établissement comptant moins de quatre lits. On a également classé les établissements de soins pour bénéficiaires internes d'après le niveau de soins fournis par l'établissement. Les résidents des établissements qui fournissent des soins de niveau inférieur au type 1 sont considérés comme autonomes et ces établissements sont par conséquent exclus de la base de sondage. Par définition, les soins de type 1 sont les soins qu'exige une personne qui est ambulatoire et (ou) capable de se déplacer seule, dont les facultés physiques et (ou) mentales sont affaiblies et qui exige une supervision et (ou) de l'aide principalement pour vaquer aux activités de la vie quotidienne. À noter que, au Québec, la variable niveau de soins n'est pas disponible. Après consultations auprès de représentants du ministère de la Santé du Québec, il a été établi que, pour cette province, la liste des établissements de soins pour bénéficiaires internes ne comprend généralement pas d'établissements dont les résidents sont autonomes.

---

<sup>1</sup> Les établissements offrant uniquement des soins de courte durée, comme les centres de traitement de la toxicomanie, sont exclus, car les résidents de ces établissements font partie de la population cible de la composante des ménages de l'ENSP.

### **5.2.1 Identification des nouveaux établissements dans la base de sondage de 1996**

La création de la base de sondage de 1996 se fonde sur les mêmes critères et les mêmes variables de stratification (voir la section 5.3 plus bas) que ceux appliqués pour la base de sondage de 1994. Les listes de 1994 et de 1996 d'établissements provinciaux de soins de longue durée ont été appariées afin d'identifier les nouveaux établissements. Puis, on a sélectionné un nouvel échantillon parmi ces établissements (voir la section 5.4 pour la sélection de l'échantillon). Il convient de souligner que l'expression « nouvel établissement » ne désigne pas nécessairement un établissement nouvellement établi depuis la dernière enquête, mais plutôt tous les établissements inclus dans la base de sondage de 1996 qui ne figurent pas dans celle de 1994. Certains établissements sont aujourd'hui dans le champ de l'enquête parce qu'ils offraient un niveau plus élevé de soins en 1996 (habituellement passage du niveau 3 au niveau 4 (ou type 1)).

## **5.3 Stratification et répartition**

La répartition de l'échantillon adoptée pour le deuxième cycle dépend en grande partie de celle adoptée pour le premier cycle. La section qui suit décrit la stratification et la répartition de l'échantillon au premier cycle.

### **5.3.1 Stratification et répartition de l'échantillon de 1994-1995**

On a fixé la taille de l'échantillon transversal à 2 600 résidents. Pour 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.

Au départ, on a stratifié la liste des établissements de santé selon la région géographique (strate géographique), puis selon la catégorie d'établissement (strate de catégorie) et le nombre de lits (strate de taille).

Les strates géographiques correspondent à cinq régions, à savoir 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, à savoir :

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. 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. Dans le cas où l'effectif de base de l'échantillon d'une strate était inférieur à 30 résidents, l'effectif de cet échantillon a été augmenté à 30 résidents.

### **5.3.2 Stratification et répartition de l'échantillon de 1996-1997**

Comme on l'a mentionné à la section 5.2, on s'est servi du plan de stratification appliqué au premier cycle (région géographique, catégorie d'établissement, nombre de lits) pour classifier les nouveaux établissements de santé. D'après les renseignements recueillis à l'étape du suivi des membres du panel longitudinal, on a décidé du nombre supplémentaire de résidents qu'il fallait interviewer dans chaque établissement du premier cycle et on a sélectionné au hasard un échantillon supplémentaire de résidents. La taille de l'échantillon supplémentaire de chaque établissement ayant participé au premier cycle dépend de l'érosion de l'échantillon pour l'établissement en question, du nombre de résidents recevant des soins de longue durée en 1996 et de la fraction de sondage correspondante en 1994.

Pour chaque nouvel établissement, on a d'abord fixé la taille de l'échantillon comme étant égal au nombre de résidents sélectionnés par établissement en 1994 pour la strate observée, puis on a rajusté la taille de l'échantillon de façon à tenir

compte de la non-réponse au niveau individuel. La correction se fondait sur le taux de réponse individuel au premier cycle, c'est-à-dire 93,6 %. Le tableau 2 montre la répartition finale de l'échantillon de 1996-1997 selon la catégorie d'établissement.

<b>Tableau 2</b>			
<b>Taille de l'échantillon selon la catégorie d'établissement</b>			
<b>Personnes âgées</b>	<b>Soins cognitifs</b>	<b>Autres établissements de réadaptation</b>	<b>Total</b>
1 816	339	238	2 393

#### **5.4 Sélection de l'échantillon**

##### **5.4.1 Sélection de l'échantillon de 1994-1995**

Au premier cycle, le nombre d'établissements sélectionnés à partir d'une strate de taille dépendait de l'échantillon attribué à la strate 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 établi 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.

##### **5.4.2 Sélection de l'échantillon de 1996-1997 pour les établissements du premier cycle**

D'après la liste des résidents bénéficiant de soins de longue durée du deuxième cycle, on a produit pour chaque établissement, deux sous-listes de résidents en fonction de la date d'admission. La première comprend tous les résidents qui vivaient dans l'établissement au moment du premier cycle de l'enquête (c.-à-d.

admis avant le 1<sup>er</sup> avril 1995). Les résidents longitudinaux qui vivaient encore dans l'établissement au moment du deuxième cycle ont été supprimés de la liste avant cette étape. La deuxième liste comprend les noms de toutes les personnes nouvellement admises (c.-à-d. admises le 1<sup>er</sup> avril 1995 ou après cette date). Un échantillon systématique de résidents a été sélectionné à partir de la liste des personnes nouvellement admises. Pour certains établissements, on a sélectionné un autre échantillon systématique de résidents d'après la liste des personnes qui vivaient dans l'établissement au moment du premier cycle de l'enquête. Puis, l'échantillon supplémentaire sélectionné par le bureau central a été transmis aux bureaux régionaux afin qu'ils procèdent aux interviews.

#### **5.4.3 Sélection de l'échantillon de 1996-1997 de nouveaux établissements au deuxième cycle**

Comme on l'a mentionné à la section 5.4.1, au premier cycle, on a tiré un échantillon systématique d'établissements dans chaque strate par échantillonnage avec probabilité proportionnelle au nombre de lits. Étant donné que les établissements ont été sélectionnés systématiquement, il était possible de sélectionner un échantillon de nouveaux établissements en ne modifiant ni le plan de sondage original ni la méthode d'estimation de la variance. Pour cela, on a trié les nouveaux établissements de façon aléatoire et on les a ajoutés au bas de la liste de la strate appropriée de la base de sondage de 1994 comme s'ils faisaient partie de la liste originale. Puis, de l'origine choisie au hasard et du pas de sondage de chaque strate en 1994, on a pu échantillonner les nouveaux établissements comme s'ils avaient été sélectionnés dans l'échantillon original. Cependant, par cette méthode, on sélectionne un moins grand nombre d'établissements que si les nouveaux établissements étaient stratifiés différemment. On a jugé que les avantages d'une modification de la stratification, donc de la sélection d'un plus grand nombre d'établissements, étaient faibles et que le changement de méthode aurait nécessité des modifications importantes au calcul de la variance.

Pour les nouveaux établissements sélectionnés lors du deuxième cycle, on a tiré un échantillon systématique de résidents d'après la liste de tous les résidents recevant des soins de longue durée.

## **6. Collecte des données**

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

Les questions de la composante des établissements de santé de l'ENSP sont conçues pour être posées lors d'une interview sur place par la méthode papier et crayon. L'interview 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 parce qu'ils étaient malades ou souffraient d'une incapacité. La personne répondant par procuration pouvait être un parent, un membre du personnel ou un bénévole de l'établissement. En tout, 59,1 % des interviews ont eu lieu par procuration (dans 72,8 % des cas, la personne qui a répondu 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, le poids et la taille et les contacts avec les professionnels de la santé.

### **6.2 Interviews**

La collecte des données a eu lieu de décembre 1996 à avril 1997 et le dépistage des personnes qui avaient déménagé s'est poursuivie jusqu'en juin 1997. Les interviews ont été menées par des intervieweurs de Statistique Canada.

Pour commencer, des intervieweurs principaux ont communiqué avec chaque établissement pour organiser une rencontre entre un interviewer 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 une personne interposée 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.

La plupart des interviews ont eu lieu sur place. L'interview complète a duré en moyenne 37 minutes dans le cas de l'interview sans procuration et 32 minutes dans le cas de l'interview par procuration.

### **6.3 Surveillance et contrôle**

Tous les intervieweurs travaillent sous la supervision d'une équipe 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 programme postés dans les divers bureaux régionaux de Statistique Canada.

#### **6.4 Non-réponse à l'ENSP**

Les intervieweurs ont reçu pour instructions de faire tous les efforts raisonnables en vue d'interviewer les résidents sélectionnés. Les refus de la part des établissements ont fait l'objet d'un suivi effectué par des intervieweurs principaux, des chargés de projet ou d'autres intervieweurs pour tenter de convaincre ces établissements de participer à l'enquête.

## **7. Traitement des données**

### **7.1 Saisie des données**

Les réponses au questionnaire s'adressant aux résidents des établissements (formulaire 6) ont été saisies par la méthode de reconnaissance de textes par codage optique (RTCO) qui consiste à balayer les questionnaires à l'aide d'un lecteur optique. Si au moins 3 % de la surface d'une case cochée est remplie, le système optique saisit l'information contenue dans la zone. Dans les cas où plus d'une case était cochée pour une question pour laquelle il ne fallait cocher qu'une seule réponse, un opérateur a examiné l'image optique pour déterminer quelle valeur était la bonne. Dans les cas où la valeur correcte n'était pas évidente, on a choisi la valeur correspondant à la deuxième réponse. Les réponses en toutes lettres dont le scanner n'était pas sûr à 90 % ont été vérifiées par un opérateur qui a examiné manuellement les images balayées sur un écran d'ordinateur. L'identificateur unique et le nom des médicaments ont été vérifiés entièrement par l'opérateur.

Les réponses au questionnaire sur les politiques de l'établissement (formulaire 5) ont été saisies manuellement. Les renseignements ont été vérifiés complètement, car le nombre d'enregistrements était petit.

Les programmes de saisie des données des questionnaires sont conçus pour éviter la saisie de valeurs aberrantes. La vérification de l'enchaînement des questions a été effectuée après la saisie des données.

### **7.2 Vérification**

À la fin de chaque interview, 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, si bien qu'il a été possible de faire un suivi immédiat.

Le bureau central a vérifié les variables démographiques et les codes de réponse avant la saisie des données. Après celle-ci, 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 (IES), aucune imputation n'a été effectuée (voir la section 8.3).

### **7.3 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). Les médicaments ont été codés d'après une version révisée du Système de classification anatomique, thérapeutique et chimique (ATC).

#### **7.4 Création de variables dérivées**

Pour simplifier l'analyse des données, on a calculé plusieurs variables du fichier d'après des réponses au questionnaire de la composante des établissements 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 de la dérivation de ces variables figurent à l'annexe D.

#### **7.5 Pondération**

Le principe sur lequel s'appuie l'estimation dans le cas d'un échantillon probabiliste comme celui de l'ENSP veut que, à part elle-même, chaque personne qui fait partie de l'échantillon en « représente » plusieurs autres qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple 2 % de la population, chaque personne représente 50 personnes de la population. Conformément à la terminologie utilisé ici, 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. Le poids figure dans le fichier de microdonnées et doit être utilisé pour produire des estimations valables à partir des résultats de l'enquête. Par exemple, pour estimer le nombre de personnes qui fument tous les jours (voir la question SMI6\_1 à la section 9.2.1), on choisit les enregistrements se rapportant aux personnes de l'échantillon qui possèdent cette caractéristique et l'on fait la somme des poids inscrits sur ces enregistrements.

La méthode employée pour calculer ces poids est décrite à la section 10.

#### **7.6 Suppression de renseignements confidentiels**

Le fichier de microdonnées « à grande diffusion » décrit ci-dessus diffère à bien des égards du fichier « principal » de l'enquête conservé par Statistique Canada. Ces différences sont dues aux mesures prises, tel que le groupement des variables, pour protéger l'anonymat des répondants. Par exemple, le fichier principal contient l'âge exacte du répondant tandis que le fichier de microdonnées contient l'âge groupé. Les utilisateurs qui désirent obtenir des renseignements exclus du fichier de microdonnées peuvent obtenir des totalisations personnalisées (voir section 12.2). Les estimations produites seront communiquées à

***ENSP DE 1996-1997 : COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ  
GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION***

---

l'utilisateur, à condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites à la section 9 de ce guide.

## 8. Qualité des données

### 8.1 Taux de réponse

Deux taux de réponse distincts peuvent être calculés pour le fichier transversal de la composante des établissements de santé de l'ENSP, à savoir le taux de réponse des établissements et le taux de réponse individuel.

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. Ces derniers ne pouvaient être interrogés sans l'autorisation de l'établissement. Tous 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 ont accepté que celle-ci soit menée auprès de leurs résidents. Le calcul du taux de réponse des établissements est le suivant :

Le taux de réponse individuel correspond au pourcentage de résidents sélectionnés dans

$$\frac{\text{Nombre d'établissements sélectionnés qui ont accepté de participer}}{\text{Nombre total d'établissements faisant partie du champ de l'enquête}} \times 100$$

$$= \frac{213}{213} \times 100$$

$$= 100 \%$$

les établissements répondants qui ont participé à une interview. Le calcul de ce taux est le suivant :

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

$$= \frac{2118}{2383} \times 100$$

$$= 88,9 \%$$

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

## **8.2 Erreurs relatives à l'enquête**

### **8.2.1 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 et aux mêmes intervieweurs, surveillants, méthodes de traitement, etc., que pour cette 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.

Étant donné que les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. Cette partie du guide décrit les *mesures de l'erreur d'échantillonnage* que Statistique Canada utilise couramment et dont il conseille vivement l'emploi aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées.

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 :

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

La section 11 du guide 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.

### **8.2.2 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 intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en répondant aux questions, que des réponses soient mal entrées dans l'ordinateur 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 du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, donné à ceux-ci une formation poussée sur les méthodes d'enquête et le questionnaire, observé les intervieweurs afin de déceler les problèmes et adopté des procédures pour veiller à réduire au minimum les erreurs de collecte de données.

L'effet de la *non-réponse* sur les résultats de l'enquête constitue une source importante d'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. Dans le cas de la composante des établissements de santé de l'ENSP, la non-réponse tant partielle que totale est faible. 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. 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.

### **8.3 Imputation**

On a eu recours à l'imputation pour calculer les valeurs de l'une des variables de la composante des établissements de santé de l'ENSP. La variable HSI6DHSI représente la cote obtenue par une personne particulière sur l'Indice de l'état de santé (IES). 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'IES, qui peut varier de 0 à 1, est calculée d'après les réponses à une série de questions sur l'état de santé<sup>2</sup>. Or, il est impossible de calculer cette cote globale si la réponse concernant l'une ou plusieurs de ces composantes manque, situation qui s'est produite pour 26,5 % des répondants. Étant donné ce taux élevé de non-réponse aux questions sur l'état de santé, il a été décidé de procéder à l'imputation des valeurs manquantes pour le calcul de l'IES pour le cycle de 1996-1997 de la composante des établissements 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'IES global pour les personnes concernées.

Il convient de souligner que la méthode utilisée n'était pas la même que celle appliquée en 1994. Cependant, pour assurer la cohérence des données, on a appliqué la méthode d'imputation choisie pour 1996-1997 aux données du fichier transversal de 1996 ainsi qu'à celles du fichier longitudinal de 1994-1996.

L'IES 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'IES. 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 de calcul de la variable dérivée de l'IES 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.

L'imputation a été effectuée en trois étapes.

- La première 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 la cote partielle avec certitude. On a

---

<sup>2</sup> Pour plus de renseignements sur le calcul de l'IES, voir l'annexe D : variables dérivées.



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 enregistrement donneur pour attribuer les cotes partielles manquantes. On a analysé à l'aide de modèles de régression les enregistrements contenant des valeurs valides pour chacune des huit cotes partielles afin de déterminer quelles autres variables du questionnaire pourraient expliquer la valeur des cotes partielles et pourraient donc être utilisées pour prédire la valeur des cotes partielles manquantes. On s'est servi de ces variables explicatives comme variables d'appariement pour repérer des enregistrements donneurs en vue de l'imputation dans les enregistrements où manquaient les valeurs de certaines cotes partielles.
- Enfin, dans certains cas, la valeur imputée d'une cote partielle était en contradiction avec les renseignements partiels figurant sur l'enregistrement receveur. Par exemple, les renseignements partiels figurant sur un enregistrement pourraient indiquer que la cote partielle pour la composante de l'ouïe devait être comprise entre 2 et 3. Quand la valeur imputée se situait en dehors de cet intervalle, on a choisi pour l'imputation la valeur la plus proche de la valeur imputée tombant dans l'étendue acceptable. Dans l'exemple choisi ici, si la valeur imputée était supérieure à 3, on l'a modifiée pour qu'elle soit égale à 3 et si elle était inférieure à 2, pour qu'elle soit égale à 2.

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

Cette section du guide décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent de quelque autre façon des données provenant des fichiers de microdonnées 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 microdonnées :

- a) Les estimations qui figurent dans le corps d'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 dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à retenir ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) le dernier chiffre à retenir.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la centaine près (ou à la décimale près) selon la méthode d'arrondissement classique.

- e) Si, en raison de contraintes d'ordre technique ou autre, on applique une autre méthode que l'arrondissement classique, si bien que les estimations qui seront publiées ou diffusées de toute autre façon diffèrent des estimations correspondantes publiées par Statistique Canada, il est vivement conseillé à l'utilisateur d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.
- f) Des estimations non arrondies ne doivent 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 la composante des établissements de santé de l'ENSP n'est pas autopondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour toutes les personnes qui font partie de l'échantillon. Même pour produire des estimations simples, y compris des tableaux statistiques ordinaires, l'utilisateur doit employer le poids d'échantillonnage approprié.

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

L'utilisateur ne doit pas non plus perdre de vue que certains logiciels 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 microdonnées de la composante des établissements 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*

SMI6\_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*

SMI6\_3: Actuellement, combien de cigarettes est-ce que ... 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 microdonnées 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 par celle du dénominateur.

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

On peut obtenir des estimations de quantités à partir du fichier de microdonnées 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 SMI6\_3 par le poids de l'enregistrement, puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMI6\_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 SMI6\_1 est « tous les jours »;
- c) diviser l'estimation obtenue en « a » par celle calculée en « b ».

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

La composante des établissements 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.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de variance), on peut rendre l'application des progiciels courants plus significative. 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 puisque, même s'ils ne reflètent toujours pas la stratification du plan de sondage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (personnes) qui contribuent à l'estimateur en question.

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

Avant de diffuser et /ou de publier des estimations tirées de ce fichier, l'utilisateur doit déterminer le nombre de personnes qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à **30**, l'estimation pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons d'au moins 30 personnes, l'utilisateur doit calculer le coefficient de variation de l'estimation au moyen du programme SAS ou SPSS d'estimation de la variance fourni (voir la section 11.1) et suivre les lignes directrices qui suivent.

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

<b>Fiabilité de l'estimation</b>	<b>c.v. (%)</b>	<b>Lignes directrices</b>
1. Acceptable	0,0 à 16,5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Médiocre	16,6 à 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 lettre M (ou d'une manière similaire).
3. Inacceptable	Supérieur à 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 lettre I (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 la section 11 pour des renseignements supplémentaires sur le calcul de la variance.

## **10. Pondération**

Les poids attribués aux unités répondantes de la composante des établissements de santé de l'ENSP se fondent sur la probabilité de sélectionner la personne corrigée pour tenir compte des résidents sélectionnés qui étaient décédés au moment de la collecte des données du deuxième cycle et de la non-réponse au niveau individuel. La méthode de pondération de 1996 est fondée en grande partie sur celle de 1994. Pour les établissements qui ont participé au premier cycle de l'enquête, on prend comme point de départ les poids finals de 1994 pour les établissements. Pour les établissements nouvellement sélectionnés, il faut calculer cette probabilité de sélection. La section qui suit donne une brève description des méthodes de pondération de 1994 qui restent applicables pour la pondération de 1996. On trouvera une description complète des méthodes de pondération utilisées en 1994 dans le guide du fichier de microdonnées à grande diffusion de l'ENSP de 1994-1995 : Établissements de santé.

### **10.1 Probabilité de sélection des établissements en 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); et  
 $n_h$  = 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 5.4.3 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 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.

### **10.1.1 Calcul des poids de 1994-1995 pour les établissements et corrections**

À ce stade, on peut calculer le poids de base au niveau de l'établissement. Cependant, comme la non-réponse est possible à ce niveau, des corrections sont nécessaires pour tenir compte des établissements qui refusent de répondre.

### **10.1.2 Poids de base de 1994-1995 pour les 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.

### **10.1.3 Correction pour la non-réponse au niveau des établissements de 1994-1995**

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**.

## **10.2 Probabilité de sélection d'un nouvel établissement en 1996-1997**

Comme on l'a décrit à la section 5.4.3, les nouveaux établissements ont été sélectionnés conformément au plan de sondage original. Par conséquent, la probabilité de sélection d'un nouvel établissement  $i$  correspondait à :

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

où, ici,  $M_{h,i}$  représente le nombre de lits dans l'établissement  $i$  de la strate  $h$  provenant de la base de sondage de 1996. Huit nouveaux établissements ont été sélectionnés au cycle 2 conformément à cette méthode de sélection. De ces établissements, deux ont été jugés hors du champ de l'enquête (parce qu'ils ne fournissaient que des soins de courte durée); les autres ont accepté de participer à l'enquête. L'inverse de la probabilité de sélection de l'établissement donne le **poids final de l'établissement**. On notera que les établissements nouvellement sélectionnés n'ont pas été regroupés avec ceux du premier cycle lorsque l'on a fait la correction pour tenir compte de la non-réponse au niveau des établissements du premier cycle.

### **10.3 Probabilité de sélection d'un résident**

#### **10.3.1 Probabilité de sélectionner un résident dans les établissements sélectionnés en 1994-1995**

La sélection du nouvel échantillon de résidents doit répondre à deux grands objectifs, à savoir sélectionner un échantillon représentatif et l'efficacité de la méthode d'échantillonnage. Par représentativité, on entend que la probabilité de sélection doit être non-nulle pour chaque résident, y compris les nouveaux résidents. Pour que l'échantillon soit efficace, autrement dit que la variance des estimations soit faible, il est souhaitable que les poids d'échantillonnage appliqués aux résidents sélectionnés dans un établissement soient aussi semblables que possible. De surcroît, l'échantillon devrait être représentatif de la distribution de la population de l'établissement en ce qui concerne les résidents nouvellement admis par opposition aux résidents qui vivaient déjà dans l'établissement au moment de l'enquête du premier cycle. Remplacer simplement les résidents sélectionnés au premier cycle qui ne vivent plus dans l'établissement par un échantillon de résidents nouvellement admis peut produire des écarts importants entre les poids d'échantillonnage. En outre, si le taux d'érosion de la population d'un établissement diffère considérablement de celui observé pour l'échantillon prélevé dans cet établissement, il se peut que l'on doive sélectionner certains résidents qui vivaient déjà dans cet établissement au moment du premier cycle. Pour chaque établissement, on a appliqué la fraction de sondage de 1994 aux deux sous-populations pour déterminer combien de résidents supplémentaires il fallait interviewer. De cette façon, on a rendu les poids d'échantillonnage des résidents sélectionnés dans un établissement aussi semblables que possible.

Notation :

$R_{h,i,b}$  = nombre réel de résidents recevant des soins de longue durée dans l'établissement  $i$  de la strate  $h$  admis avant le 1<sup>er</sup> avril 1995;

$r_{h,i,b}$  = nombre de résidents de l'établissement  $i$  de la strate  $h$  admis avant le 1<sup>er</sup> avril 1995 et désignés pour être sélectionnés (sauf les membres du panel longitudinal);

$R_{h,i,a}$  = nombre réel de résidents recevant des soins de longue durée dans l'établissement  $i$  de la strate  $h$  admis le 1<sup>er</sup> avril 1995 ou après cette date;

$r_{h,i,a}$  = nombre de résidents de l'établissement  $i$  de la strate  $h$  admis le 1<sup>er</sup> avril 1995 ou après cette date et désignés pour être sélectionnés; et

$r_{panel}$  = nombre de membres du panel longitudinal vivant encore dans l'établissement  $i$  de la strate  $h$ .

La probabilité de sélection d'un résident est égale à :

$$\frac{r_{h,i,b} + r_{panel}}{R_{h,i,b}} \quad \text{s'il s'agit d'un résident sélectionné admis avant le 1er avril 1995 ou d'un membre du panel longitudinal}$$

ou

$$\frac{r_{h,i,a}}{R_{h,i,a}} \quad \text{s'il s'agit d'un résident sélectionné admis le 1er avril 1995 ou après cette date.}$$

### **10.3.2 Probabilité de sélectionner un résident dans les établissements nouvellement sélectionnés**

Notation :

$R_{h,i}$  = nombre réel de résidents recevant des soins de longue durée dans l'établissement  $i$  de la strate  $h$ ; et

$r_{h,i}$  = nombre de résidents de l'établissement  $i$  de la strate  $h$  désignés pour être sélectionnés.

Pour les nouveaux établissements, la probabilité de sélection est la même pour tous les résidents et est égale à :

$$\frac{r_{hi}}{R_{hi}}$$

### **10.3.3 Poids de base pour les personnes**

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.

### **10.3.4 Correction pour les décès**

Comme, dans certains cas, il y a eu un décalage important entre le moment du tirage de l'échantillon et le début de l'enquête, le nombre de décès observés durant l'enquête a été beaucoup plus élevé que le nombre prévu. Puisque le profil des décès chez les résidents des établissements de santé n'est pas aléatoire (les personnes les moins bien portantes étant plus susceptibles que les autres de mourir), les estimations risquent d'être biaisées si aucune correction n'est faite pour remédier à ce problème. Pour tenir compte du fait que certains résidents sélectionnés sont décédés durant la période de collecte, on a calculé un facteur de correction pour le décès que l'on a appliqué aux poids de base des unités répondantes.

Le calcul de ce facteur de correction pour le décès comprend de nombreuses étapes. En premier lieu, en se servant du panel longitudinal, on a déterminé les caractéristiques des personnes qui sont décédées dans les six mois qui ont suivi l'interview du premier cycle. Puis, on a formé des catégories, soit des groupements d'unités (c.-à-d. de résidents) affichant le même risque d'être décédés dans les six mois après l'interview du premier cycle. On s'est servi, pour former ces catégories, d'un algorithme de mise en grappes qui dispose les unités d'échantillonnage selon une structure arborescente par subdivision successive de l'ensemble des données en « branches » d'après les caractéristiques des unités. On s'est servi du logiciel *KnowledgeSeeker IV for Windows*, mis au point par ANGOSS Software International Limited, pour produire la structure arborescente. On a utilisé l'algorithme CHAID (Chi-square Automatic Interaction Detection) du logiciel *KnowledgeSeeker* pour repérer, à chaque nœud, la caractéristique qui permettait le mieux de subdiviser l'échantillon en groupes dissimilaires en ce qui concerne l'indicateur « décédé dans les six mois »/« non décédé dans les six mois ». Pour

chaque catégorie, on a dérivé un facteur défini comme étant le ratio du nombre de résidents décédés dans les six mois qui ont suivi l'interview du premier cycle au nombre de résidents qui ne sont pas décédés durant ce six mois. Ce facteur est donné par :

*∑ des poids de 1994 pour les membres du panel décédés six mois après l'interview du premier cycle dans la catégorie*

*∑ des poids de 1994 des membres du panel non décédés six mois après l'interview du premier cycle dans la catégorie*

Puisque l'on a défini les catégories d'après des caractéristiques observées au premier cycle, on se sert des données existantes sur les caractéristiques correspondantes des répondants du deuxième cycle pour définir l'appartenance à une catégorie, l'hypothèse étant que les ratios pour les données du deuxième cycle devraient être fort semblables à ceux calculés d'après les données du premier cycle. Idéalement, on devrait redistribuer dans chaque strate/catégorie la somme des poids des résidents sélectionnés qui sont décédés durant la période de collecte. Cependant, cette méthode produirait des corrections instables étant donné le petit nombre d'enregistrements dans certaines combinaisons strate-catégorie. Pour respecter autant que possible la stratification originale et produire des corrections plus stables, on a subdivisé l'échantillon en six strates poststrates, à savoir quatre pour les établissements pour personnes âgées, une pour les établissements de soins cognitifs et une pour les autres établissements de réadaptation. Dans le cas des établissements pour les personnes âgées, chaque région comprend quatre à cinq strates de taille. Les quatre poststrates a posteriori ont été formées de façon à ce qu'elles soient à peu près de même taille (en ce qui a trait au total des poids et au nombre d'enregistrements) et correspondent, dans l'ensemble, aux établissements de petite, moyenne, grande et très grande taille. On notera que la majorité des résidents sélectionnés qui sont décédés durant la période de collecte vivaient dans des établissements pour personnes âgées.

Ensuite, on a subdivisé l'échantillon en groupes « poststrate-catégorie » pour chaque répondant du deuxième cycle. Puis, on a calculé la somme des poids de base des répondants pour chaque groupe « poststrate-catégorie ». En multipliant cette somme par le ratio de personnes « décédées dans les six mois »/« non décédées dans les six mois » (tel que défini plus haut) déterminé pour la catégorie en question, on obtient le nombre initial pondéré de résidents sélectionnés décédés qu'il convient d'affecter au groupe. Ensuite, on procède au calage de cette répartition initiale sur la somme correspondante des poids de base des résidents sélectionnés qui sont décédés dans cette poststrate. Le facteur final de correction pour décès à appliquer à un groupe « poststrate-catégorie » est donné par :

*(somme des poids de base des répondants classés dans le groupe « poststrate - catégorie »*

+

*∑ (somme des poids de base des résidents sélectionnés décédés affectés au groupe « poststrate - catégorie »)*

*∑ (somme des poids de base des répondants classés dans le groupe « poststrate – catégorie »*

Le facteur de correction pour décès ainsi calculé a été appliqué au poids de base des répondants classés dans le groupe poststrate-catégorie en question. Après cette étape, on n'a maintenu dans les fichiers que les unités répondantes et les unités non répondantes.

### **10.3.5 Correction pour la non-réponse au niveau des personnes**

Comme certains résidents sélectionnés n'ont pas répondu au questionnaire, une correction supplémentaire a dû être apportée au poids de base des personnes pour tenir compte de la non-réponse. Comme on l'a décrit à la section 5.1.1, l'échantillon du deuxième cycle comprend les membres du panel longitudinal ainsi que des résidents nouvellement sélectionnés. Puisque l'on dispose de renseignements supplémentaires, tels que la date de naissance et le sexe, pour tous les membres du panel longitudinal, la correction pour la non-réponse d'une personne n'est pas la même pour cette partie de l'échantillon (qui représente environ 60 % de l'échantillon total). Pour les résidents nouvellement sélectionnés lors du deuxième cycle qui ont refusé de répondre, on ne possède de renseignements que sur le sexe (on notera que la valeur de cette variable a été imputée dans cinq enregistrements).

Pour les membres du panel, la correction pour la non-réponse du poids de la personne est donnée par :

*∑ (somme des poids des résidents répondants et non répondants d'une catégorie genre d'établissement/âge/sexe)*

*∑ (somme des poids des résidents répondants d'une catégorie genre d'établissement/âge/sexe)*

Ici, le poids correspond au poids de base des personnes après correction pour les décès. Le groupe âge-sexe correspond à celui utilisé à l'heure actuelle dans le fichier de microdonnées à grande diffusion. Il convient de souligner que certains groupes originaux âge-sexe ont été regroupés pour les établissements de soins cognitifs et pour les autres établissements de réadaptation. L'objet du regroupement était de produire une correction plus stable.

**ENSP DE 1996-1997 : COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ**  
**GUIDE DU FICHIER DE MICRODONNÉES À GRANDE DIFFUSION**

---

Pour l'échantillon de remise à niveau, la correction pour la non-réponse des poids des personnes est donnée par :

$$\frac{\text{somme des poids de résidents répondants et non répondants d' une catégorie genre d' établissement/sexe}}{\text{somme des poids des résidents répondants d' une catégorie genre d' établissement/sexe}}$$

Ici, le poids correspond au poids de base des personnes après correction pour les décès.

La multiplication du poids après correction pour les décès par la correction pour la non-réponse au niveau des personnes donne le **poids final de la personne** qui figure dans le fichier. Comme l'enquête se concentre sur les personnes plutôt que sur les établissements, le poids final de l'établissement ne figure pas dans le fichier.

## 11. Calcul de la variance

Dans le cas de la composante des établissements de santé de l'ENSP, on se sert d'une formule simple et bien connue pour calculer la variance et le coefficient de variation (c.v.) des estimations. Cette formule s'appuie sur l'hypothèse selon laquelle les établissements sont sélectionnés avec probabilité inégale et avec remise. En réalité, les établissements ont été sélectionnés sans remise; autrement dit, une fois sélectionné, un établissement ne pouvait être choisi de nouveau. Cette mesure a été prise pour des raisons d'ordre opérationnel plutôt que pour améliorer la variance, si bien que l'hypothèse d'un échantillon avec remise devrait avoir un effet négligeable.

Un programme de calcul de la variance, rédigé pour les systèmes SAS et SPSS, est fourni avec le présent fichier de microdonnées. Il peut être utilisé pour calculer la variance des moyennes et des totaux. Les formules utilisées pour calculer la variance d'un total  $Y$  ou d'un rapport  $R=Y/X$  sont :

$$V(\hat{Y}_{total}) = \frac{1}{N} \sum_{h=1}^N \frac{\sum_{i=1}^{n_h} (\hat{Y}_{h,i} - \hat{Y}_h)^2}{n_h(n_h - 1)} \quad \text{et} \quad V(\hat{R}) = \frac{1}{\hat{X}^2} \sum_{h=1}^N \frac{\sum_{i=1}^{n_h} (\hat{Y}_{h,i} - \hat{R}_h \hat{X}_{h,i})^2}{n_h(n_h - 1)}$$

où :

$\hat{Y}_h$  est l'estimation, pour la strate  $h$ , de la variable de réponse  $Y$  basée sur tous les répondants dans la strate  $h$ ;

$\hat{Y}_{h,i}$  est l'estimation, pour la strate  $h$ , de la variable de réponse  $Y$  basée sur l'ensemble des répondants de l'établissement  $i$  de la strate  $h$ ;

$N$  est le nombre de strates;

$n_h$  est le nombre d'établissements échantillonnés dans la strate  $h$ ;

$\hat{X}_h$  est l'estimation, pour la strate  $h$ , de la variable  $X$  du dénominateur du rapport basée sur l'ensemble des répondants dans la strate  $h$ ;

$\hat{X}_{h,i}$  est l'estimation, pour la strate  $h$ , de la variable  $X$  du dénominateur du rapport basée sur l'ensemble des répondants de l'établissement  $i$  de la strate  $h$ ;



$\hat{X}$  est l'estimation globale de la variable  $X$  du dénominateur du rapport; et

$\hat{R}_h$  est le rapport  $\hat{Y}_h / \hat{X}_h$ .

## **11.1 Exécution des programmes de calcul de la variance**

Les deux programmes (SAS et SPSS) qui accompagnent le fichier de microdonnées permettent à l'utilisateur de calculer des totaux et des rapports sans trop de difficulté. Les programmes utilisent des fichiers périphériques SAS et SPSS, qui sont aussi fournis. L'utilisateur doit s'assurer que les références au nom des fichiers sont cohérents lors de l'utilisation de ces programmes et fichiers.

### **11.1.1 Exécution du programme SAS de calcul de la variance**

Les étapes du programme SAS sont les suivantes :

ÉTAPE 1 : À l'étape des données, sous « Étape 1 », l'utilisateur identifie les variables pour lesquelles il veut calculer des totaux et /ou des rapports. Pour chaque variable, une valeur doit être attribuée pour les données manquantes (p. ex. valeurs pour Ne sait pas, Refus ou Non déclaré). Consulter le dictionnaire des données pour les valeurs à traiter comme manquantes. Dans le cas des totaux, une variable 0/1 est attribuée à chaque caractéristique voulue. De même, dans le cas des rapports, une variable 0/1 est attribuée au numérateur et au dénominateur du rapport. Dans l'exemple, on veut calculer trois totaux et les variables 0/1 sont désignées par tot1-tot3. Dans l'exemple de rapport, les variables sont désignées par num1-num4 et denom1-denom4, où num1/denom1 représente un rapport nécessitant une estimation de la variance. Il faut respecter la convention de désignation des variables, c'est-à-dire tot1-tot $n$ , num1-num $m$  et denom1-denom $m$ , car le programme utilise automatiquement ces désignations. Dans l'instruction « keep » à la fin de la présente étape, modifier les variables tot $n$ , num $m$  et denom $m$  de façon à indiquer le nombre de totaux ou de rapports à calculer.

Les estimations quantitatives se calculent de la même façon. La seule différence tient au fait que la variable 0/1 est remplacée par une variable de quantité, où la quantité représente la valeur de la caractéristique observée pour le répondant.

ÉTAPE 2 : Dans l'instruction « proc format », l'utilisateur peut attribuer des noms descriptifs à la place des noms des totaux et des rapports créés par le programme. « Totfrmt » désigne les totaux et « ratfrmt » les rapports. Par exemple, dans totfrmt, 2=Popn 65+, car tot2 sert à calculer l'estimation et la variance de la population totale âgée de 65 ans et plus.

ÉTAPE 3 : À cette étape, il suffit de changer les désignations des vecteurs de manière à afficher le nombre exact de totaux, de numérateurs et de dénominateurs. Par exemple, si l'utilisateur indique deux totaux et trois rapports à l'ÉTAPE 1, les instructions « array » se liront comme suit :

```
array totarray{*}tot1-tot2;  
array numarray{*}num1-num3;  
array denarray{*}denom1-denom3;
```

Le programme peut ensuite être exécuté comme tout autre programme SAS.

Nota : Le programme exécute des sous-routines différentes selon qu'il s'agit de calculer des totaux ou des rapports. Si l'on ne veut calculer que des totaux ou que des rapports, il n'est pas nécessaire d'exécuter les deux sous-routines. À la fin du programme, il y a les deux lignes suivantes:

```
%totals;  
%rates;
```

Pour gagner du temps, il est possible de supprimer la sous-routine inutile en plaçant l'instruction pertinente entre /\* et \*/. Par exemple, si l'on ne veut calculer que des totaux, il faut indiquer :

```
%totals;  
/* %rates; */
```

Seule la sous-routine de calcul de la variance des totaux sera exécutée.

Dans cet exemple, il n'est pas nécessaire d'indiquer les variables num1-numm et denom1-denomm à l'ÉTAPE 1, car il n'y a aucun taux à calculer. De même, si la ligne %totals est supprimée, il n'est pas nécessaire d'indiquer les variables tot1-totn. Il suffit de supprimer le nom de ces variables dans l'instruction « keep » à l'ÉTAPE 1.

### **11.1.2 Exécution du programme SPSS de calcul de la variance**

Le programme SPSS de calcul de la variance a été rédigé de façon similaire au programme SAS. Il comprend les trois mêmes étapes que celles définies pour le programme SAS. Cependant, pour des raisons techniques, l'ordre séquentiel de ces étapes n'est pas le même pour le programme SPSS que pour le programme SAS. Mais chaque étape est identifiée correctement dans le programme. Les étapes du programme SPSS sont les suivantes :

ÉTAPE 1 : À l'étape des données, sous « ÉTAPE 1 », l'utilisateur identifie les variables pour lesquelles il veut calculer des totaux et(ou) des rapports. Pour chaque variable, une valeur doit être attribuée pour les données manquantes (p. ex. valeurs pour Ne sait pas, Refus ou Non déclaré). Consulter le dictionnaire des données pour les valeurs à traiter comme manquantes. Dans le cas des totaux, une variable indicatrice est attribuée à chaque caractéristique voulue. De même, dans le cas des rapports, une variable indicatrice est attribuée au numérateur et au dénominateur du rapport. Dans l'exemple, on veut calculer trois totaux et les variables indicatrices sont désignées par tot1-tot3. Dans l'exemple de rapport, les variables sont désignées par num1-num4 et den1-den4, où num1/den1 représente un rapport nécessitant une estimation de la variance. Il faut respecter la convention de désignation des variables, c'est-à-dire tot1-tot $n$ , num1-num $m$  et den1-den $m$ , car le programme utilise automatiquement ces désignations.

Les estimations quantitatives se calculent de la même façon. La seule différence tient au fait que la variable indicatrice est remplacée par une variable de quantité, où la quantité représente la valeur de la caractéristique que possède le répondant.

ÉTAPE 2 : L'utilisateur peut attribuer des étiquettes pour remplacer les noms de totaux et de rapports créés par le programme. Les instructions commentaires /\*<<< ÉTIQUETTES DES TOTAUX <<< et /\*<<< ÉTIQUETTES DES RAPPORTS <<< indiquent l'emplacement dans le programme où les étiquettes de totaux et de rapports doivent être précisées.

**ENSP DE 1996-1997 : COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ  
GUIDE DU FICHER DE MICRODONNÉES À GRANDE DIFFUSION**

---

ÉTAPE 3 : À cette étape, l'utilisateur doit simplement modifier la valeur des macrovariables !ntotal, !nrates et !dim de sorte que soit affiché le nombre correct de totaux et de rapports. Par exemple, si l'utilisateur a défini deux totaux et trois rapports, alors les instructions SPSS se liront :

```
!!let !ntotal=2.
!!let !nrates=3.
!!let !dim=6.
```

La variable macro !dim doit être égale au double du nombre de rapports à calculer.

L'utilisateur doit modifier le nom du sous-répertoire où les fichiers sont sauvegardés dans la variable macro !let !folder.

L'exécution des programmes en se servant des variables de l'exemple produit des données de sortie semblables à celles-ci :

Estimations, variances et c.v. des totaux				1
OBS	DESCRIPTION	TOTAL	VARIANCE	c.v.
1	Pop. tot.	222967.15	29161285.48	2.42
2	Pop. 65+	185139.44	26768521.00	2.79
3	Femmes 80+	103702.22	19183726.64	4.22
Estimations, variances et c.v. des rapports				2
OBS	DESCRIPTION	RAPPORT	VARIANCE	c.v.
1	Liens étroit avec les membres du personnel 65+	0.40613	0.00044143	5.17
2	Anglais 65+	0.53973	0.00037237	3.58
3	Français 65+	0.17786	0.00018277	7.60
4	Anglais/français 65+	0.10223	0.00015822	12.30

## **12. Utilisation du fichier**

### **12.1 Convention appliquée pour nommer les variables**

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

#### **12.1.1 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 / composante
- Position 4 : Année/cycle durant lequel apparaît la variable
- Position 5 : Type de variable (c.-à-d. questionnaire, codé, dérivé, etc.)
- Positions 6 et 8 : Numéro/nom de la variable tiré du questionnaire

Par exemple, la signification du nom de variable DHI6GAGE:

- DH : dans la section du questionnaire sur la démographie et la composition du ménage;
- I : questions qui font partie de la composante des établissements de santé;
- 6 : figure dans le cycle de 1996-1997;
- G : variable groupée; et
- AGE : nom de la variable.

### 12.1.2 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 sociodémographiques
ED	Niveau de scolarité	SM	Usage du tabac
FI	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		

### 12.1.3 Position 3 : Type d'enquête / composante

- 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 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)

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

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

**12.1.5 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 transversale 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'interview (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.

### **12.1.6 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 DHI6GAGE en 1996-1997. 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.

## **12.2 Accès aux fichiers maîtres**

Les fichiers de microdonnées à grande diffusion doivent répondre aux normes rigoureuses de sécurité et de confidentialité prévues par la *Loi sur la statistique* en vue de protéger les personnes qui participent à l'enquête contre la divulgation de renseignements personnels. Pour s'assurer du respect de ces normes, chaque fichier de microdonnées est soumis à un processus officiel d'examen visant à confirmer que les données qu'il contient ne permettent d'identifier aucun répondant. Les valeurs rares des variables susceptibles de permettre d'identifier une personne sont supprimées du fichier ou regroupées en catégories plus générales de façon à réduire au minimum la divulgation de renseignements personnels. Cependant, ces variables sont souvent essentielles à l'analyse complète et détaillée des données d'enquête. Puisqu'une quantité considérable de ressources est consacrée à la collecte de ces données, il est important de prendre des mesures permettant d'exploiter pleinement le potentiel analytique des fichiers de microdonnées, donc de rentabiliser l'investissement statistique.

Les employés du service personnalisé à la clientèle de Statistique Canada peuvent aussi produire des totalisations spéciales. Ce service permet à l'utilisateur incapable de se servir des logiciels de totalisation d'avoir accès aux fichiers maîtres pour produire ses propres calculs personnalisés. Les résultats sont filtrés pour s'assurer qu'ils répondent aux normes de fiabilité et de non-divulgation des renseignements personnels avant d'être diffusés. Ce service n'est pas gratuit.