

ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION DE 1998-1999
COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ
TABLE DES MATIÈRES

1. INTRODUCTION	3
2. CONTEXTE.....	4
3. OBJECTIFS	5
4. CONTENU DE L'ENQUÊTE	6
4.1 CRITÈRES	6
4.2 RÉVISIONS DU CONTENU POUR L'ENQUÊTE DU CYCLE 3 (1998-1999).....	6
5. PLAN D'ÉCHANTILLONNAGE	10
5.1 PLAN D'ÉCHANTILLONNAGE DE 1994-1995.....	10
5.2 SÉLECTION DE L'ÉCHANTILLON DE 1994-1995.....	11
5.3 ÉCHANTILLON DE 1998-1999	11
6. COLLECTE DES DONNÉES	14
6.1 CONCEPTION DU QUESTIONNAIRE ET MÉTHODE DE COLLECTE DES DONNÉES.....	14
6.2 NON-RÉPONSE À L'ENSP	14
7. TRAITEMENT DES DONNÉES	15
7.1 SAISIE DES DONNÉES ET VÉRIFICATION.....	15
7.2 CODAGE	15
7.3 CRÉATION DE VARIABLES DÉRIVÉES.....	15
8. QUALITÉ DES DONNÉES.....	16
8.1 TAUX DE RÉPONSE LONGITUDINAUX	16
8.1.1 <i>Taux de réponse du cycle 1 (1994-1995)</i>	16
8.1.2 <i>Taux de réponse du cycle 2 (1996-1997)</i>	17
8.1.3 <i>Taux de réponse du cycle 3 (1998-1999)</i>	17
8.2 ERREURS RELATIVES À L'ENQUÊTE	18
8.2.1 <i>Erreurs d'échantillonnage</i>	18
8.2.2 <i>Erreurs non dues à l'échantillonnage</i>	19
8.3 IMPUTATION.....	20

9.	LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION.....	22
9.1	LIGNES DIRECTRICES POUR L'ARRONDISSEMENT.....	22
9.2	LIGNES DIRECTRICES POUR LA PONDÉRATION DE L'ÉCHANTILLON EN VUE DE LA TOTALISATION.....	23
	9.2.1 Définitions des catégories d'estimation : nominale par opposition à quantitative.....	23
	9.2.2 Totalisation d'estimations de type nominal.....	24
	9.2.3 Totalisation d'estimations quantitatives.....	25
9.3	LIGNES DIRECTRICES POUR L'ANALYSE STATISTIQUE.....	25
9.4	LIGNES DIRECTRICES POUR LA DIFFUSION.....	26
10.	PONDÉRATION.....	28
10.1	PROBABILITÉ DE SÉLECTION DES ÉTABLISSEMENTS EN 1994-1995.....	28
	10.1.1 Calcul des poids de base de 1994-1995 pour les établissements et corrections pour la non-réponse.....	29
10.2	PROBABILITÉ DE SÉLECTION DES RÉSIDENTS EN 1994-1995.....	30
	10.2.1 Poids de base de 1994-1995 pour les résidents et correction pour la non-réponse au niveau des résidents.....	30
	10.2.2 Ajustement pour la non-réponse en 1996-1997.....	31
	10.2.3 Ajustement pour la non-réponse en 1998-1999.....	31
	10.2.4 Ajustement pour la stratification à posteriori.....	31
11.	CALCUL DE LA VARIANCE.....	32
11.1	MÉTHODE DU BOOTSTRAP.....	32
11.2	ESTIMATION DE LA VARIANCE AVEC LE PROGRAMME BOOTVARF.SAS.....	33
12.	UTILISATION DU FICHIER.....	34
12.1	CONVENTION APPLIQUÉE POUR NOMMER LES VARIABLES.....	34
	12.1.1 Structure élémentaire des noms des variables.....	34
	12.1.2 Positions 1 et 2 : Nom de la variable / section du questionnaire.....	35
	12.1.3 Position 3 : Type d'enquête / composante.....	35
	12.1.4 Position 4 : Variable de l'année de référence / du cycle.....	36
	12.1.5 Position 5 : Type de variable.....	36
	12.1.6 Positions 6 à 8 : Nom de la variable.....	37
12.2	ACCÈS AUX FICHIERS MAÎTRES.....	37

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 troisième cycle de collecte des données a débuté en février 1999 par le suivi des résidents interrogés en 1994-1995 lors du premier cycle de l'enquête. Contrairement au cycle 2, aucun échantillon de remise à niveau transversal n'a été sélectionné pour le cycle 3. Également nouveau au cycle 3, la collecte de données des personnes faisant partie de l'échantillon des établissements qui sont retournées en ménages privés a été faite par la composante des établissements de santé et non la composante des ménages, comme cela avait été fait lors du cycle 2. Les questionnaires posés aux résidents des ménages n'étaient pas les mêmes que ceux posés aux résidents en établissements, mais plus de 90% des questions étaient identiques.

Ce manuel a été produit pour faciliter l'utilisation du fichier maître du cycle 3 (1998-1999) de la composante des établissements de santé de l'ENSP.

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

Renseignements sur le télé-accès aux fichiers maîtres :

1-613-951-1653

Colette Koeune, ENSP,

Division des statistiques sur la santé

Internet : nphs-ensp@statcan.ca

Télécopieur :

1-613-951-4198

Renseignements sur le contenu de l'enquête :

Mario Bédard, ENSP,

Division des statistiques sur la santé

Internet : Mario.Bedard@statcan.ca

Télécopieur :

1-613-951-8933

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 brosser un tableau complet de l'état de santé de la population ni de la multitude de facteurs qui ont un effet sur la santé.

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

On a ajouté à l'enquête 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 du cycle 3 (1998-1999)

Voici la liste des questions qui ont été modifiées, ajoutées ou supprimées :

- Les questions posées aux répondants dépendent de quel échantillon ils font maintenant partie. Si le répondant vit en établissement, il répond aux questions des sections B à P. Si le répondant provenait de l'échantillon I (établissements) mais est retourné vivre à la maison, il répond aux questions des sections CC à QQ.

COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE L'ENSP DE 1998-1999

- Le formulaire 1 — *Liste des établissements du bureau régional* et le formulaire 2 — *Renseignements pour la section de l'échantillon* n'ont pas été utilisés car il n'y a pas eu d'échantillon de remise à niveau transversal. La situation finale des codes et des contacts avec l'établissement est sur la page couverture du formulaire de *Contrôle pour l'établissement*.
- La *Liste des résidents à long terme* (formulaire 99) n'a pas été utilisée puisqu'elle résume les données sur l'admission des résidents admis avant et après la fin de la collecte du cycle 1 afin de sélectionner un échantillon de remise à niveau et qu'il n'y a pas eu d'échantillon de remise à niveau dans le cycle 3. Les intervieweurs ont reçu un tableau contenant de l'information supplémentaire sur le dépistage et les contacts de chaque résident en établissement et chaque résident des ménages.
- Le contenu du formulaire 3 — *État du répondant de l'enquête longitudinale* a été incorporé au questionnaire du répondant (section A) pour simplifier la collecte de l'information sur les répondants des composantes Établissements et Ménages de l'enquête longitudinale.
- Les confirmations du sexe, de la date de naissance et du numéro provincial d'assurance-santé ont été supprimées puisqu'elles ont été confirmées aux cycles 1 et 2.
- La taille et le poids ont été supprimés puisqu'ils étaient utilisés pour calculer l'indice de la masse corporelle, qui n'est plus calculé pour les personnes de plus de 65 ans.
- Le consentement à l'entrevue (donné par le plus proche parent) a été incorporé au questionnaire du répondant, sections B, N, et QQ. Là où le consentement écrit était demandé par l'établissement, une déclaration générale pour la divulgation d'information a été ajoutée au formulaire *Diffusion de l'information par l'établissement — consentement du répondant*. Ce formulaire a été changé pour incorporer la formulation du questionnaire principal et pour abandonner la référence à la collecte des données sur la taille et le poids.
- La date de l'entrevue a été supprimée par erreur. Une variable avec une date fixe a été créée pendant le traitement pour la remplacer.
- Dans le formulaire de *Contrôle pour l'établissement* (auparavant le formulaire 5), le numéro de formulaire a été supprimé, les champs d'adresse ajoutés, le numéro de l'intervieweur supprimé (cela dédoublait le numéro de la tâche) et le numéro séquentiel (pré-imprimé sur les formulaires) ajouté pour lier les formulaires de contrôle et les questionnaires.

COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE L'ENSP DE 1998-1999

- Le *Questionnaire pour les résidents* (formulaire 6) a été renommé *Questionnaire pour les répondants* (le numéro du formulaire a été supprimé) pour le cycle 3 puisqu'il a été utilisé à la fois pour les résidents et les gens en ménages.
- "Échantillon de type = C ou I" a été ajouté à l'étiquette du questionnaire afin de distinguer les échantillons Ménages (C) et Établissements (I).
- Les questions sur l'endroit et le numéro séquentiel ont été ajoutées sur la page couverture du *Questionnaire pour les répondants* pour aider à relier les questionnaires et les formulaires de contrôle.
- Le code de statut final 077 a été ajouté pour les gens de l'échantillon C (ménages) qui sont retournés en ménages. Les intervieweurs de la composante ménages vont retracer ces cas lors du Q5.
- La formulation de l'introduction a été changée pour correspondre à celle de l'enquête sur les ménages.
- La formulation de la question "est... toujours un résident à long terme de cet établissement?" a été changée pour "...vit-il/elle toujours à (lisez les renseignements sur l'étiquette)?" afin d'incorporer la formulation pour les répondants des ménages.
- La catégorie "maison privée" est devenue "ménage privé" pour insister encore plus sur le déplacement des institutions vers un ménage. "Un autre établissement de soins de santé" est devenu "Établissement de soins de santé" pour incorporer les répondants des ménages.
- *État du répondant de l'enquête longitudinale* (section A): des questions sur la ville et la province du décès ont été ajoutées pour faciliter la correspondance avec les fichiers de mortalité. Une question sur le numéro de téléphone de l'établissement a été ajoutée pour faciliter le retraçage et un saut impératif pour le type d'échantillon et l'endroit a été ajouté pour incorporer les gens de l'échantillon principal qui se sont déplacés et les gens de l'échantillon des établissements qui sont maintenant des résidents des ménages.
- La section *Information sur le répondant choisi* (sections C et CC) a été changée pour *Répondant choisi*, afin d'être consistant avec la terminologie des résidents des ménages. Les répondants vivant en ménages ne reçoivent pas les questions C4, C5, C6 et C7.
- Il n'y a eu aucun changement à la section *État de santé général* (sections D et DD), ni à la section *État de santé* (sections E et EE).

- Le cancer a été ré-ajouté en tant que problème de santé chronique (il avait été exclus au cycle 1) , à la section *Problèmes de santé chroniques* (sections F et FF).
- *Limitation des activités* (sections G et GG): la question RA-Q1 (GG1) pour les gens qui vivent en ménages prend la formulation du questionnaire de la composante ménages. La question RA-Q7 (GG8) utilise la formulation du questionnaire des établissements mais remplace "institution" par "ménage". La question RA-Q5 sur le deuxième problème de santé a été séparée en deux questions: G5/GG5 sont des questions Oui/Non, puis G6/GG6 est la réponse longue pour le deuxième problème.
- *Équilibre* (sections H et HH): à la question FL_Q4 (H4/HH4) on a ajouté "commotion cérébrale" au questionnaire français.
- Pas de changements à la section *Usage du tabac* (sections I et II).
- *Consommation d'alcool* (sections J et JJ): les questions AL-Q2 et AL-Q3 ont changé de place. Les instructions "passez à..." ont aussi changé.
- *Soutien social* (section K): quelques catégories "Aucun" ont été remplacées par l'instruction à l'intervieweur "Si aucun entrez 00". Une instruction "passez à..." dans la question SS_Q11 (K12) a été changée pour permettre à ceux qui quittent l'établissement pour des sorties moins d'une fois par mois de recevoir la question sur les types de sorties, SS_Q12 (K13). Deux nouvelles catégories ont été ajoutées: "avez-vous fait des promenades en auto" et "êtes-vous allé au restaurant pour dîner ou souper?". La question sur la flexibilité de l'horaire, présente au cycle 1, a été ré-ajoutée.
- *Renseignements socio-démographiques* (sections L et LL): "assurance-chômage" a été changé pour "assurance-emploi" et les questions sur le lieu de naissance l'ethnie, la race, la langue et l'éducation ont été supprimées.
- *Personnes à contacter* (sections M et MM): on a ajouté le code de la province à l'adresse du contact et "personnel de l'établissement" à la liste des personnes contacts.
- *Consommation de médicaments* (sections O et OO): le nom d'un membre du personnel a été supprimé et une instruction à l'intervieweur a été ajoutée: "Demandez à la personne de regarder sur la bouteille, le tube ou la boîte".
- *Utilisation des soins de santé* (sections P et PP): les personnes en ménages ne se font plus poser de questions sur les transferts en établissements de soins actifs.

5. Plan d'échantillonnage

L'échantillon longitudinal de la composante des établissements de santé de l'ENSP de 1998-1999 est composé des répondants longitudinaux choisis au cycle 1. Cet échantillon est composé de 2 287 personnes qui habitaient dans un établissement de santé en 1994-1995.

La population cible de la composante des établissements de santé de l'ENSP de 1994-1995 comprenait les résidents des établissements de soins de longue durée (plus de 6 mois) de toutes les provinces, sauf ceux des territoires, des réserves indiennes et des bases des Forces canadiennes. 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. Les établissements considérés étaient les établissements de longue durée où les soins de santé étaient les principaux types de soins offerts. Les établissements devaient avoir un minimum de 4 lits et les résidents ne devaient pas être autonomes. À partir de cette liste, on a sélectionné un échantillon d'établissements, puis, dans chaque établissement sélectionné, 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 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

Autres établissements de réadaptation	présentant des troubles mentaux ou des troubles de retard du développement et les hôpitaux psychiatriques; 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.

5.2 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.

5.3 Échantillon de 1998-1999

La taille de l'échantillon diminue progressivement due à l'érosion du panel depuis 1994. L'ENSP définit l'érosion comme étant la non-réponse, les refus et les personnes impossibles à retracer. Cette diminution de la taille de l'échantillon est très petite et ne devrait pas accroître la variance des estimations. Il faut noter que même si les personnes décédées ont été exclues de la collecte de 1998-1999, elles demeurent valides pour les analyses et sont

COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE L'ENSP DE 1998-1999

incluses dans le fichier principal. Les personnes décédées sont considérées comme des réponses complètes pour des analyses longitudinales. Le Tableau 1 montre la taille de l'échantillon longitudinal de la composante des établissements de santé de l'ENSP de 1998-1999.

Tableau 1		
Taille de l'échantillon longitudinal		
Échantillon longitudinal de 1994	Fichier longitudinal complet de 1998: <i>Réponse complète en 1994, 1996 and 1998</i>	Personnes décédées (sur 2178)
2287	2178	1250

Puisqu'aucune unité longitudinale n'a été sélectionnée au cycle 3, la population couverte par l'échantillon longitudinal représente la population de 1994-1995.

Les membres de l'échantillon longitudinal peuvent avoir changé de situation depuis 1994. Toutes les situations sont résumées au tableau 2.

Tableau 2	
Situation du membre du panel longitudinal	Mesure prise
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.
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).
Avait déménagé dans un ménage privé	Les membres du panel longitudinal qui avaient déménagé dans un ménage privé ont été interviewés.
Décédé	Les membres décédés du panel longitudinal sont exclus de la collecte mais font partie du fichier longitudinal de la composante des établissements de santé.

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 troisième 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.

6. Collecte des données

6.1 Conception du questionnaire et collecte des données

Les questions de la composante des établissements de santé de l'ENSP sont conçues pour être posées lors d'une entrevue sur place par la méthode papier et crayon. L'entrevue téléphonique est permise dans les cas où une personne répondant par procuration ne peut être interrogée sur place. Pour chaque établissement, l'administrateur ou une personne-ressource a déterminé pour lesquels des résidents sélectionnés les réponses devaient être recueillies par procuration. Cette décision était basée sur l'état de santé des résidents sélectionnés. La personne répondant par procuration pouvait être un parent, un membre du personnel ou un bénévole de l'établissement. En tout, 57 % des entrevues ont eu lieu par procuration (dans 71% 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 et les contacts avec les professionnels de la santé.

La collecte des données a eu lieu de février à mai 1999. Les entrevues 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 intervieweur et l'administrateur ou une personne-ressource de l'établissement. Durant cette visite de liaison, l'intervieweur a fait remplir un court questionnaire sur les politiques de l'établissement. La liste des résidents pour lesquels les réponses devaient être recueillies par 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.

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

6.2 Non-réponse à l'ENSP

Les intervieweurs ont reçu pour instructions de faire tous les efforts raisonnables en vue d'interviewer les résidents sélectionnés. 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, ceci ayant comme résultat que tous les établissements ont participé à l'enquête.

7. Traitement des données

7.1 Saisie des données et vérification

Les réponses au questionnaire ont été saisies manuellement dans les bureaux régionaux avec la méthode DC2. Les programmes de saisie de données des questionnaires sont conçus pour éviter la saisie de valeurs aberrantes.

Le formulaire de *Contrôle pour les établissements* a été saisi manuellement. Cette information a été vérifiée à 100% dû au petit nombre de données.

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

Après la saisie des données, on s'est assuré que l'enchaînement des questions était correct et on a procédé à des contrôles de cohérence sur certaines zones des enregistrements. Sauf dans le cas de l'Indice de l'état de santé (IES), aucune imputation n'a été effectuée (voir la section 8.3).

7.2 Codage

Les problèmes de santé causant une limitation des activités ont été codés conformément à la Classification internationale des maladies, 9^e 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 canadien (ATC), développé par Santé Canada.

7.3 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.

8. Qualité des données

8.1 Taux de réponse longitudinaux

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

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

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. Le calcul du taux de réponse des établissements est le suivant :

$$\begin{aligned} & \frac{\text{Nombre d'établissements sélectionnés qui ont accepté de participer}}{\text{Nombre total d'établissements où logeaient des personnes faisant partie du panel}} \times 100 \\ &= \frac{214}{224} \times 100 \\ &= 95,5 \% \end{aligned}$$

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

$$\begin{aligned} & \frac{\text{Nombre de résidents qui ont participé complètement ou partiellement à l'interview}}{\text{Nombre total de résidents sélectionnés dans les établissements participants}} \times 100 \\ &= \frac{2287}{2444} \times 100 \\ &= 93,6 \% \end{aligned}$$

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

8.1.2 Taux de réponse du cycle 2 (1996-1997)

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 :

$$\begin{aligned} & \frac{\text{Nombre d'établissements sélectionnés qui ont accepté de participer}}{\text{Nombre total d'établissements où logeaient des personnes faisant partie du panel}} \times 100 \\ &= \frac{314}{314} \times 100 \\ &= 100 \% \end{aligned}$$

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

$$\begin{aligned} & \frac{\text{Nombre de résidents qui ont participé complètement ou partiellement à l'interview}}{\text{Nombre total de résidents sélectionnés dans les établissements participants}} \times 100 \\ &= \frac{2193}{2287} \times 100 \\ &= 95,9 \% \end{aligned}$$

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

8.1.3 Taux de réponse du cycle 3 (1998-1999)

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

COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE L'ENSP DE 1998-1999

ont accepté que celle-ci soit menée auprès de leurs résidents. Le calcul du taux de réponse des établissements est le suivant :

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

$$= \frac{352}{352} \times 100$$

$$= 100 \%$$

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

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

$$= \frac{2251}{2287} \times 100$$

$$= 98,4 \%$$

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.

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

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. Il y a non-réponse partielle lorsque le répondant refuse de répondre à une question ou qu'il ne peut se souvenir des renseignements demandés. Il y a non-réponse totale lorsque l'intervieweur a été incapable de communiquer avec la personne chargée de répondre par procuration ou que le répondant sélectionné a refusé de participer à l'enquête. Dans le cas de la composante des établissements de santé, la non-réponse tant partielle que totale est faible. On traite les cas de non-réponse totale en corrigeant le poids des résidents qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu. La méthode employée pour calculer ces poids est décrite à la section 10.

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 HSI8DHSI 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 (avec la possibilité d'avoir des valeurs négatives), est calculée d'après les réponses à une série de questions sur l'état de santé¹. 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 environ 5 % des répondants. Il a été décidé de procéder à l'imputation des valeurs manquantes pour le calcul de l'IES pour le cycle de 1998-1999 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

¹ Pour plus de renseignements sur le calcul de l'IES, voir l'annexe D : variables dérivées.

manquantes, afin de pouvoir calculer l'IES global pour les personnes concernées. Il convient de souligner que la méthode utilisée était la même que celle appliquée en 1994 et 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 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 quelle 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 données de l'enquête. Ces lignes directrices devraient leur permettre de reproduire les nombres publiés par Statistique Canada et de produire aussi des nombres non encore publiés conformes aux lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement

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

- a) Les estimations qui figurent dans un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux de marginaux des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près selon la méthode d'arrondissement classique.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et/ou dénominateurs), puis arrondis à une décimale par la méthode d'arrondissement classique. Si l'on veut arrondir une estimation à un seul chiffre décimal par cette méthode et que le 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

diffusés de toute autre façon différent 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 tous les répondants. Même pour produire des estimations simples, y compris des tableaux statistiques standards, l'utilisateur doit employer le poids d'échantillonnage approprié. Sinon, les estimations calculées à partir du présent fichier ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles qui pourraient être produites par Statistique Canada.

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

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

Deux principaux genres d'estimations ponctuelles des caractéristiques de la population peuvent être produites d'après le fichier de données 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

SMI8_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

SMI8_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 données en additionnant les poids de tous les enregistrements contenant la ou les caractéristiques voulues. Pour calculer les proportions et les rapports de la forme \hat{Y} / \hat{X} , on doit :

- a) additionner les poids des enregistrements contenant la caractéristique voulue pour le numérateur (\hat{Y});

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

9.2.3 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de données en multipliant la valeur de la variable voulue par le poids de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements concernés. Par exemple, pour obtenir une estimation du nombre *total* de cigarettes que fument chaque jour les personnes qui fument tous les jours, on multiplie la valeur déclarée à la question SMI8_3 par le poids de l'enregistrement (WTI8LF), puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMI8_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 SMI8_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

Fiabilité de l'estimation	c.v. (%)	Lignes directrices
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	<p>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 :</p> <p>« 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. »</p>

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

Contrairement à la pondération transversale, la pondération longitudinale tient compte de la probabilité de sélection de l'unité d'analyse au moment de la sélection de l'échantillon. Les poids attribués aux unités répondantes de la composante des établissements de santé de l'ENSP se fondent sur la probabilité d'avoir sélectionné cette unité au moment de la sélection de l'échantillon en 1994.

La pondération du cycle 3 est donc en grande partie basée sur la pondération du cycle 1. 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. Des ajustements sont aussi apportés pour tenir compte de la non-réponse au cycle 3. Quelques ajustements sont aussi apportés, notamment pour tenir compte de la non-réponse aux cycle 2 et 3.

La section qui suit donne une brève description de la méthode de pondération de 1994 qui reste applicable pour la pondération longitudinale du cycle 3, ainsi qu'une description des ajustements spécifiques au cycle 3.

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 5.3 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 Poids de base de 1994-1995 pour les établissements et correction pour la non-réponse

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

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

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

10.2 Probabilité de sélection des résidents en 1994-1995

Notation :

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

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

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

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

10.2.1 Poids de base de 1994-1995 pour les résidents et correction pour la non-réponse au niveau des résidents

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

La correction supplémentaire est apportée au poids de base des personnes pour tenir compte de la non-réponse des résidents:

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

La multiplication du poids final de l'établissement par le poids de base du résident ajusté pour la non-réponse donne le **poids du résident au cycle 1**.

10.2.2 Ajustement pour non-réponse en 1996-1997

La non-réponse est aussi possible au cycle 2. Les non-répondants de ce cycle ne font pas partie du fichier complet du cycle 3. Leurs poids doivent donc être redistribués aux répondants du cycle 2. Le poids du cycle 1 des répondants au cycle 2 est multiplié par:

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

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

On obtient ainsi le **poids du résident au cycle 2**.

10.2.3 Ajustement pour non-réponse en 1998-1999

Un ajustement est aussi apporté pour tenir compte de la non-réponse au cycle 3. Le poids des répondants (poids du résident au cycle 2) est multiplié par:

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

Cette correction est apportée séparément pour chaque catégorie possible de réponse longitudinale (la variable LONGPAT, ie le genre d'établissement pour chacun des cycles).

10.2.4 Ajustement pour la stratification à posteriori

Puisque le nombre total de personnes au Canada vivant en établissement de santé est inconnu (selon la définition d'établissement de l'ENSP), il est impossible d'effectuer une post-stratification en se basant sur ces totaux. Cependant, une post-stratification est faite en utilisant les poids totaux obtenus au cycle 1, pour chaque région, type d'établissement et catégorie âge-sexe. La post-stratification est faite en deux étapes. Premièrement au niveau de cinq régions, puis pour chaque type d'établissement et catégorie âge-sexe. Le poids de base du résident ajusté pour la non-réponse et pour la stratification à posteriori donne le **poids final du résident (WTI8LF)**.

11. Calcul de la variance

La méthode utilisée pour calculer la variance des estimations au cycle 3 diffère de celle employée au cycle 2. Au cycle 3, la méthode du bootstrap est utilisée. Il s'agit de la même méthode employée pour la composante principale de l'ENSP. Cette méthode est expliquée dans cette section.

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

11.1 Méthode du bootstrap

Les plans d'échantillonnage pour les enquêtes sur la santé sont complexes. Comme le calcul de la variance pour de tels plans ne peut être fait à l'aide de simples formules, on a recours à une méthode de rééchantillonnage pour calculer des estimations de la variance.

La méthode du bootstrap consiste à sous-échantillonner l'échantillon initial. À l'intérieur de chaque strate, on choisit un échantillon aléatoire simple (ÉAS), avec remise, de $n-1$ grappes parmi les n grappes de la strate. Un certain nombre B de nouveaux échantillons (ou répétitions) sont ainsi créés. La même estimation est ensuite calculée pour chacun des B échantillons, ce qui permet d'obtenir B estimations différentes. L'estimation de la variance correspond à la simple variance entre les B estimations. Afin d'obtenir chacune des B estimations, un poids spécifique à chaque échantillon est nécessaire. Pour chaque échantillon ÉAS, le facteur de pondération est donc recalculé pour chaque enregistrement dans la strate. Ces B poids, les poids bootstrap, ont été produits et sont disponibles avec les données.

En résumé, la méthode du bootstrap consiste à:

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

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

11.2 Estimation de la variance avec le programme BOOTVARF.SAS

Le programme qui permet d'estimer la variance est le programme BOOTVARF.SAS. Ce programme est fourni avec le fichier de données. Il permet de calculer des estimations de variance pour des totaux, des rapports, des différences entre des rapports, des paramètres de régression linéaire et logistique et des modèles linéaires généralisés.

L'utilisateur doit s'assurer que les références au nom des fichiers sont cohérents lors de l'utilisation du programme. Pour plus de détails sur l'utilisation de BOOTVARF.SAS, consulter le guide d'utilisation fourni avec le programme.

12. Utilisation du fichier

12.1 Convention appliquée pour nommer les variables

En 1996-1997, on a adopté une convention pour nommer les variables de l'ENSP qui permet aux utilisateurs des données de faire facilement référence à des données similaires provenant d'autres périodes de collecte et aux 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

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'entrevue (p. ex., indicateur de travail).
G	Variable groupée	Variables recueillies, codées, supprimées ou dérivées agrégées en un groupe (p. ex., groupes d'âge).
L	Variable longitudinale dérivée	Variable calculée d'après des variables provenant d'au moins deux cycles d'enquête.

12.1.6 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

Il existe trois moyens d'accéder aux fichiers maîtres de l'enquête. Le premier moyen est d'utiliser le service de télé-access. On fournit à l'utilisateur un fichier maître « fictif » et le cliché d'enregistrement correspondant. Grâce à ces outils, il ou elle peut mettre au point son propre ensemble de programmes analytiques en se servant du fichier fictif pour confirmer que les routines fonctionnent convenablement. Il ne reste plus qu'à envoyer le programme pour les totalisations personnalisées par courriel à nphs-ensp@statcan.ca. Le programme de totalisation est transmis au réseau interne protégé de Statistique Canada et appliqué au fichier maître de données de la composante des établissements de santé de l'ENSP approprié. Les données de sortie sont filtrées pour s'assurer qu'elles sont conformes aux normes de confidentialité et de fiabilité puis sont transmises au client. Ce service est gratuit.

Une deuxième méthode offerte à tous les clients consiste à demander au personnel spécialisé des Services personnalisés à la clientèle de la Division de la statistique de la santé de produire des totalisations personnalisées (courriel hd-ds@statcan.ca). Ce service permet aux utilisateurs d'obtenir des totalisations à partir du fichier maître sans soumettre de programme informatique. Comme dans le cas du télé-access, les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Contrairement au télé-access, ce service est offert contre remboursement des frais.

Finalement, un programme de recherche permet aux chercheurs de soumettre à Statistique Canada un projet de recherche utilisant les données des fichiers maîtres de l'ENSP.

COMPOSANTE DES ÉTABLISSEMENTS DE SANTÉ DE L'ENSP DE 1998-1999

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