

Enquête nationale sur la santé de la population

Volet Établissements de soins de santé

Cycle 1 (1994-1995)

Guide du fichier de microdonnées à grande diffusion

Statistique Canada

TABLE DES MATIÈRES

1.	Introduction.....	1
2.	Renseignements de base	2
3.	Objectifs	3
4.	Contenu de l'enquête.....	4
5.	Plan d'échantillonnage.....	5
	5.1 Plan de la base de sondage	5
	5.2 Stratification et répartition	5
	5.3 Prélèvement de l'échantillon	7
6.	Collecte des données	8
	6.1 Conception du questionnaire et méthode de collecte des données	8
	6.2 Essai	8
	6.3 Interviews.....	8
	6.4 Surveillance et contrôle.....	9
	6.5 Non-réponse à l'ENSP	9
7.	Traitement des données	10
	7.1 Saisie des données.....	10
	7.2 Vérification	10
	7.3 Codage	10
	7.4 Création de variables dérivées	11
	7.5 Pondération	11
	7.6 Suppression de renseignements confidentiels.....	11
8.	Qualité des données.....	12
	8.1 Taux de réponse	12
	8.2 Erreurs relatives à l'enquête	13
	8.3 Imputation	15
9.	Lignes directrices pour la totalisation, l'analyse et la diffusion.....	16
	9.1 Lignes directrices pour l'arrondissement.....	16
	9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	17
	9.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatives	18
	9.2.2 Totalisation d'estimations de type nominal	19
	9.2.3 Totalisation d'estimations quantitatives	19
	9.3 Lignes directrices pour l'analyse statistique	20
	9.4 Lignes directrices pour la diffusion.....	20
10.	Pondération	22
	10.1 Probabilités de sélection	22
	10.1.1 Probabilité de sélection d'un établissement.....	22
	10.1.2 Probabilité de sélection d'un résident à l'intérieur d'un établissement.....	23

10.2	Calcul des poids et corrections	23
10.2.1	Poids de base pour les établissements	24
10.2.2	Correction pour la non-réponse au niveau des établissements..	24
10.2.3	Poids de base pour les personnes	24
10.2.4	Correction pour la non-réponse au niveau des personnes.....	24
11.	Calcul de la variance	26
11.1	Exécution du programme de calcul de la variance	27

Annexe A: Questionnaire

Annexe B: Cliché d'enregistrement

Annexe C: Dictionnaire des données

Annexe D: Variables dérivées

1. Introduction

L'Enquête nationale sur la santé de la population (ENSP) a pour but de recueillir de l'information au sujet de la santé de la population canadienne. Afin de broser un tableau complet de la santé des Canadiens, une enquête spéciale a été élaborée pour les personnes qui résident dans des établissements de santé, hôpitaux, foyers de soins infirmiers et centres d'hébergement pour personnes handicapées. Cette enquête produira non seulement des données transversales, mais aussi des données longitudinales qui seront recueillies tous les deux ans auprès d'un échantillon permanent d'individus.

En 1994 et en 1995, des résidents des ménages de chaque province ont été interviewés dans le cadre du volet principal de l'ENSP. Les données de cette enquête -ménage ont été diffusées à l'automne de 1995. De plus, une enquête distincte a été menée au Yukon et dans les Territoires du Nord-Ouest; les résultats de cette enquête seront diffusés plus tard.

L'Enquête nationale sur la santé de la population, volet établissements de soins de santé a été menée par Statistique Canada de janvier à mars 1995. Ce guide a pour but de faciliter l'utilisation du fichier de microdonnées des résultats de l'enquête.

Pour toute question au sujet de l'ensemble de données ou de son utilisation, s'adresser à :

Unité d'accès aux données
(613) 951-1653
nphs-ensp@statcan.ca

2. Renseignements de base

À l'automne de 1991, le Conseil national d'information sur la santé (CNIS) a recommandé la réalisation d'une enquête permanente nationale sur la santé de la population. Cette recommandation découlait des pressions économiques et fiscales exercées sur le régime de soins de santé et du besoin correspondant d'information pour améliorer l'état de santé de la population canadienne. Les sources existantes de données sur la santé étaient insuffisantes pour pouvoir dresser un tableau complet de l'état de santé de la population et de la multitude de facteurs qui ont un effet sur la santé.

Dès le mois d'avril 1992, Statistique Canada a obtenu des fonds pour élaborer l'Enquête nationale sur la santé de la population. L'enquête a été conçue de manière à être souple et à produire des données valables, fiables et actuelles. Il fallait aussi que l'enquête puisse être adaptée aux besoins, aux politiques et aux intérêts changeants.

Une enquête spéciale a été menée auprès des résidents des établissements de soins de santé parce que cette population est rarement incluse dans les enquêtes nationales et que ses caractéristiques de santé sont vraisemblablement différentes 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 en vue de la réalisation d'études analytiques qui aideront à comprendre les facteurs déterminants de la santé;
- recueillir des données sur les corrélations entre la santé et les facteurs économiques, sociaux, démographiques et environnementaux;
- éclaircir la relation entre l'état de santé et l'utilisation des soins de santé;
- fournir des renseignements sur un échantillon permanent d'individus suivis dans le temps, afin d'illustrer la dynamique de la santé et de la maladie et de déterminer les facteurs qui ont un effet sur le placement en établissement;
- donner aux provinces, aux territoires et à d'autres clients une capacité d'enquête sur la santé pour leur permettre de compléter le contenu ou l'échantillon;
- permettre de faire des liens entre les données de l'enquête et des données administratives recueillies couramment, 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

Le contenu du volet établissements de soins de santé de l'ENSP a été choisi d'après les critères suivants :

- 1) l'enquête doit recueillir des renseignements sur l'état de santé des Canadiens qui résident dans un établissement de santé;
- 2) dans la mesure du possible, les données recueillies doivent pouvoir être comparées à celles de la population des ménages;
- 3) l'enquête doit éclaircir les conditions liées au placement en établissement;
- 4) les renseignements fournis doivent permettre d'étudier, dans le temps, la transition entre un ménage et un établissement et vice versa;
- 5) l'enquête doit produire des données au niveau national.

Les répondants ont été choisis au hasard parmi des établissements de soins de santé sélectionnés. Le questionnaire portait sur l'état de santé, les facteurs de risque, le soutien social, les contacts avec des professionnels de la santé, ainsi que la situation démographique et socio-économique. Par exemple, l'état de santé a été mesuré à l'aide de questions sur la perception personnelle de la santé, la capacité fonctionnelle, les états chroniques et les limitations d'activité. Les questions sur les facteurs de risque à l'égard du comportement portaient sur l'usage du tabac et la consommation d'alcool. Le degré de soutien social a été mesuré au moyen de questions sur la fréquence des contacts avec des amis et des parents à l'intérieur et à l'extérieur de l'établissement. Les questions sur les caractéristiques démographiques et socio-économiques portaient sur l'âge, le sexe, la scolarité, l'origine ethnique et le revenu personnel. Le questionnaire de l'enquête figure à l'annexe A.

5. Plan d'échantillonnage

La population cible du volet établissements de soins de santé de l'ENSP était formée de tous les résidents de longue durée des établissements de santé de toutes les provinces, à l'exclusion des territoires. Une liste des établissements ayant des résidents de longue durée a été dressée et stratifiée selon la région géographique, le type et la taille des établissements. Un échantillon d'établissements a été sélectionné et un échantillon subséquent de résidents a été prélevé à l'intérieur de ces établissements.

5.1 Plan de la base de sondage

La base de sondage a été établie à partir des listes des établissements de soins de santé et des hôpitaux que tient la Division des statistiques sur la santé (DSS) de Statistique Canada. Ces listes ont été vérifiées et mises à jour par les ministères provinciaux de la santé pour s'assurer de leur exactitude. Ces établissements ont été classés selon le principal type de soins qu'ils fournissent; et, seuls les établissements qui assurent des soins de longue durée ont été retenus¹. Parmi la liste des établissements de soins, on a retenu ceux qui offrent des soins de longue durée aux personnes âgées, aux enfants ayant des troubles affectifs et aux personnes ayant des troubles de développement, une déficience physique ou des troubles psychiques. La liste des hôpitaux comprenait des hôpitaux généraux ayant des unités de soins de longue durée, des établissements de soins prolongés et des centres de réadaptation, ainsi que des hôpitaux spécialisés ayant des unités de soins de longue durée, comme les hôpitaux pédiatriques et les hôpitaux psychiatriques. Le nombre de lits de soins de longue durée de chaque établissement était connu.

La population de référence a été limitée aux établissements ayant au moins quatre lits et offrant des soins de longue durée aux personnes qui ont des problèmes de santé. Les établissements ayant moins de quatre lits n'étaient pas inclus dans les listes des établissements de soins et des hôpitaux que tient la DSS. Les établissements de soins et les hôpitaux situés dans les réserves indiennes, les bases des Forces canadiennes et les centres correctionnels ont été supprimés des listes pour des raisons d'ordre opérationnel.

5.2 Stratification et répartition

La taille de l'échantillon total a été fixée à 2 600 résidents. En supposant un taux de réponse de 85 %, un échantillon de cette taille est suffisant pour produire des estimations

¹ Les établissements offrant uniquement des soins de courte durée, comme les centres de traitement de la toxicomanie, ont été exclus car les résidents des établissements de soins de courte durée sont compris dans le champ du volet des ménages de l'ENSP.

nationales avec un coefficient de variation (c.v.) de 10 % pour les variables présentes dans au moins 10 % de la population.

La liste des établissements de santé a été stratifiée au départ selon la région géographique (strate géographique) et, ensuite, selon le type d'établissement (strate de caractéristique) et le nombre de lits (strate de taille).

Les strates géographiques étaient composées de cinq régions (les provinces de l'Atlantique, le Québec, l'Ontario, les provinces des Prairies et la Colombie-Britannique). Trois strates de caractéristique ont été créées dans chaque strate géographique :

Établissements pour personnes âgées	les établissements de soins pour personnes âgées et les hôpitaux de soins prolongés.
Établissements de soins cognitifs	y compris les établissements de soins pour les enfants ayant des troubles affectifs, les personnes ayant des troubles psychiques et les personnes ayant des troubles de développement, ainsi que les hôpitaux psychiatriques.
Autres établissements de réadaptation	y compris les hôpitaux de réadaptation, les hôpitaux pédiatriques et les autres hôpitaux spécialisés, les hôpitaux généraux ayant des unités de soins de longue durée, ainsi que les établissements de soins pour les personnes ayant une déficience physique.

À l'intérieur de chacune des strates géographiques / caractéristique, les établissements ont été divisés en strates de taille en groupant ceux ayant un nombre de lits comparable. Le nombre de strates de taille dépendait du nombre total de lits dans la strate géographique / caractéristique. Une fois le nombre de strates de taille connu, on a établi les limites des différentes strates de taille selon 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 les cas où l'effectif de base de l'échantillon d'une strate de taille était inférieur à 30 résidents, l'effectif de cet échantillon a été élargi à 30 résidents.

Tableau 1			
Taille de l'échantillon selon la strate de caractéristique			
Personnes âgées	Soins cognitifs	Autres établissements de réadaptation	Total
1 931	411	282	2 624

5.3 Prélèvement de l'échantillon

Le nombre d'établissements sélectionnés dans une strate de taille dépendait de la part de l'échantillon attribuée à la strate et de la taille des établissements compris dans la strate. Dans les strates comportant de grands établissements, un plus grand échantillon de résidents a été sélectionné à l'intérieur de chaque établissement. Cela a eu pour effet de réduire le nombre d'établissements visités. Après avoir déterminé le nombre d'établissements à sélectionner dans chaque strate de taille, un échantillon systématique d'établissements a été prélevé dans la liste de la strate selon la méthode d'échantillonnage avec probabilité proportionnelle à la taille (PPT). La taille était fonction du nombre de lits de soins de longue durée. Il était possible de trouver dans la liste un bureau central regroupant plusieurs petits établissements. En pareil cas, une liste de tous les établissements relevant de ce bureau central a été obtenue et deux d'entre eux ont été sélectionnés - le plus grand (d'après le nombre de lits) et un autre qui a été sélectionné au hasard selon la méthode d'échantillonnage PPT. Une liste des résidents de longue durée de chaque établissement sélectionné a été dressée et un échantillon systématique de résidents a été prélevé à partir de cette liste juste avant les interviews.

6. Collecte des données

6.1 Conception du questionnaire et méthode de collecte des données

Les questions du volet établissements de soins de santé de l'ENSP ont été conçues pour être posées au moyen d'une interview sur place selon la méthode papier et crayon. Une interview téléphonique était permise dans les cas où il était impossible d'interviewer une personne interposée sur place.

C'est l'administrateur ou un représentant désigné de l'établissement qui a déterminé quels répondants sélectionnés il fallait interviewer par personne interposée à cause de maladie ou d'incapacité du répondant. La personne interposée pouvait être un parent, un membre du personnel ou un bénévole de l'établissement. Une proportion de 57 % des interviews ont été faites par personne interposée (dans 40 % des cas, la personne interposée était un parent du résident). Un membre du personnel de l'établissement a fourni des renseignements sur les médicaments pris par les résidents sélectionnés et sur leurs contacts avec des professionnels de la santé.

6.2 Essai

Un essai a eu lieu à l'automne de 1994 avant la réalisation de l'enquête principale sur le terrain. Cet essai a été effectué par des intervieweurs expérimentés de Statistique Canada. Le principal objectif de l'essai était d'observer les réactions des répondants à l'enquête, d'estimer la durée de l'interview et d'évaluer la possibilité d'obtenir un taux de réponse élevé.

6.3 Interviews

La collecte des données a eu lieu du 30 janvier au 31 mars 1995. Les interviews ont été menées par des intervieweurs de Statistique Canada.

Avant la collecte des données, les intervieweurs principaux ont téléphoné à tous les établissements pour organiser une rencontre entre un intervieweur et l'administrateur de l'établissement. Lors de la première visite à l'établissement, l'intervieweur a fait remplir un court questionnaire sur les politiques de l'établissement et a procédé à la sélection de l'échantillon. C'est durant cette visite qu'on a déterminé quels résidents il fallait interviewer par personne interposée et qu'on a obtenu le nom et le numéro de téléphone d'un proche parent de ces résidents; ensuite, on a téléphoné aux proches parents et on leur 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.

Les interviews ont été effectuées sur place dans la mesure au possible. L'interview complète a duré en moyenne 40 minutes dans le cas d'une interview sans personne interposée et 30 minutes dans le cas d'une interview par personne interposée.

6.4 Surveillance et contrôle

Tous les intervieweurs travaillent sous la surveillance d'une équipe d'intervieweurs principaux. Les intervieweurs principaux sont chargés de veiller à ce que les intervieweurs connaissent bien les concepts et les méthodes de l'enquête. Ils contrôlent périodiquement le travail des intervieweurs et revoient les documents qu'ils remplissent. Les intervieweurs principaux travaillent à leur tour sous la surveillance des gestionnaires de programme postés à chacun des bureaux régionaux de Statistique Canada.

6.5 Non-réponse à l'ENSP

Les intervieweurs ont reçu pour instructions de faire tous les efforts raisonnables pour interviewer les résidents sélectionnés. Les cas de refus de la part des établissements ont fait l'objet d'un suivi effectué par les intervieweurs principaux, les 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

Le programme de saisie des données du questionnaire a été conçu pour empêcher la saisie de données 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 revu le questionnaire pour s'assurer d'avoir suivi correctement l'enchaînement des questions. Les bureaux régionaux ont fait une vérification ultérieure des questionnaires pour s'assurer qu'ils étaient complets et lisibles et que les réponses étaient cohérentes. Il était ainsi 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 la saisie des données, tous les enregistrements ont fait l'objet d'un contrôle descendant pour vérifier l'enchaînement des questions. Sauf dans le cas de l'Indice général de santé (IGS), aucune imputation n'a été effectuée (voir la section 8.3).

7.3 Codage

Les renseignements contenus dans les réponses écrites ont été codés en de nouvelles catégories distinctes. Les problèmes de santé causant une limitation des activités ont été codés selon la Classification internationale des maladies, version 9 (CIM-9), ou selon le Système de codage supplémentaire pour les incapacités ostéo-articulaires et musculaires établi aux fins de l'Enquête sur la santé et l'incapacité au Canada. Les médicaments ont été codés selon le Système de classification anatomique, thérapeutique et chimique (ATC) pour les médicaments pour usage humain.

7.4 Création de variables dérivées

Pour faciliter l'analyse des données, un certain nombre de variables du fichier de microdonnées ont été calculées en utilisant les réponses au questionnaire du volet établissements de soins de santé de l'ENSP. Le nom des variables dérivées est habituellement précédé des lettres DV et suivi de caractères se rapportant au numéro de la question ou au sujet. Les détails du calcul de ces variables figurent à l'annexe D.

7.5 Pondération

L'estimation pour un échantillon probabiliste (comme celui de l'ENSP) repose sur le principe selon lequel chaque personne faisant partie de l'échantillon représente, à part elle-même, plusieurs autres personnes qui n'en font pas partie. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 membres de la population. Dans la terminologie de ce texte, on peut dire que chaque personne a un coefficient de pondération (ou poids) de 50.

La phase de pondération sert à calculer le poids pour chaque personne. Ce poids est inscrit dans le fichier de microdonnées et il faut l'utiliser 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 SMOK-Q62 à la section 9.2.1), il faut choisir les enregistrements se rapportant aux personnes comprises dans l'échantillon qui possédant cette caractéristique. Il faut ensuite faire la somme des poids inscrits sur ces enregistrements.

La méthode employée pour calculer ces poids est décrite en détail au chapitre 10.

7.6 Suppression de renseignements confidentiels

Il y a des différences importantes entre le fichier de microdonnées à grande diffusion et le fichier principal conservé par Statistique Canada. Ces différences sont dues à des mesures prises pour protéger l'anonymat des répondants à l'enquête. Les utilisateurs qui veulent obtenir des renseignements exclus du fichier de microdonnées à grande diffusion peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, à la condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 de ce guide.

8. Qualité des données

8.1 Taux de réponse

Le volet établissements de soins de santé de l'ENSP comporte deux taux de réponse distincts. Le taux de réponse des établissements représente le pourcentage d'établissements admissibles qui ont accepté de permettre que l'enquête soit menée auprès de leurs résidents. Les résidents ne pouvaient pas être interviewés sans l'autorisation de l'établissement. Le taux de réponse des établissements se calcule comme suit :

$$\frac{\text{Nombre d'établissements sélectionnés ayant accepté de participer}}{\text{Nombre total d'établissements sélectionnés admissibles}} \times 100$$

$$= \frac{214}{224} \times 100$$

$$= 95.5\%$$

Le taux de réponse individuel représente le pourcentage de résidents sélectionnés des établissements participants qui ont été interviewés. Le taux de réponse individuel se calcule comme suit :

$$\frac{\text{Nombre de résidents ayant été interviewés complètement ou partiellement}}{\text{Nombre total de résidents sélectionnés dans les établissements participants}} \times 100$$

$$= \frac{2287}{2444} \times 100$$

$$= 93.6\%$$

Note : la multiplication des deux taux de réponse ne donne pas un résultat valable, car le nombre de résidents sélectionnés dans chaque établissement n'est pas le même.

8.2 Erreurs relatives à l'enquête

L'enquête permet de produire des estimations à partir des données portant sur un échantillon et recueillies auprès de ses membres. On aurait pu obtenir des estimations quelque peu différentes si on avait procédé à un recensement complet en utilisant le même questionnaire et en faisant appel aux mêmes intervieweurs, surveillants, méthodes de traitement, etc. que pour cette enquête. L'écart entre les estimations découlant de l'échantillon et celles tirées d'un dénombrement complet effectué dans des conditions semblables est appelé erreur d'échantillonnage des estimations.

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 inscrites dans le questionnaire et 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 auront peu d'effet sur les estimations tirées des données de l'enquête. Toutefois, les erreurs qui se produisent systématiquement causeront des biais dans les estimations de l'enquête. Beaucoup de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Par exemple, on a fait appel à des intervieweurs hautement qualifiés, on leur a donné une formation poussée sur les méthodes d'enquête et le questionnaire, et on a observé leur travail afin de s'assurer qu'ils comprenaient bien les instructions.

L'effet de la non-réponse à l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage. L'ampleur de la non-réponse varie de la non-réponse partielle (le fait de ne pas répondre à une ou à plusieurs questions) à la non-réponse complète. La non-réponse complète et la non-réponse partielle au volet établissements de soins de santé de l'ENSP ont été minimales. Il y a eu non-réponse partielle lorsque le répondant a refusé de répondre à une question ou qu'il n'a pas pu se souvenir d'un renseignement demandé. Il y a eu non-réponse complète lorsque l'intervieweur a été incapable de contacter une personne interposée ou que l'établissement ou le répondant a refusé de participer à l'enquête. Pour traiter la non-réponse complète, le poids des résidents qui ont répondu à l'enquête a été corrigé afin de tenir compte de ceux qui n'ont pas répondu.

Les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage. Conformément aux règles de bonne pratique en matière de statistique, les chercheurs doivent fournir aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. Cette partie du guide décrit les mesures de l'erreur

d'échantillonnage que Statistique Canada emploie couramment. Il est fortement conseillé aux utilisateurs qui produisent des estimations à partir du fichier de microdonnées d'employer eux aussi ces mesures.

La mesure de l'importance éventuelle des erreurs d'échantillonnage est fondée sur l'erreur-type des estimations tirées des résultats de l'enquête. En raison de la grande 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. Cette mesure s'exprime en pourcentage de l'estimation.

Par exemple, supposons qu'à partir des résultats de l'enquête, on estime que 12 % des résidents des établissements de santé fument des cigarettes tous les jours. L'erreur-type de cette estimation est établie à .0096. Le coefficient de variation de l'estimation se calcule alors comme suit :

$$\left(\frac{.0096}{.12} \right) \times 100\% = 8.28\%$$

La section 11 du guide contient plus de détails sur le calcul des coefficients de variation pour l'enquête.

8.3 Imputation

Les résultats de l'une des variables du volet établissements de soins de santé de l'ENSP ont été calculés par imputation. La variable IMHST indique le résultat obtenu par le résident à l'Indice général de santé (IGS). Il s'agit d'une mesure de l'état de santé général qui sert à évaluer la vision, l'ouïe, l'élocution, la mobilité, la dextérité, les sentiments, la mémoire, la pensée et les douleurs. Les résultats globaux à l'IGS, allant de 0 à 1, ont été calculés à partir des réponses à une série de questions sur l'état de santé². Si une réponse n'était pas donnée à l'un ou à plusieurs des éléments de l'IGS, il était impossible de calculer un résultat complet, ce qui s'est produit dans le cas de 20,5 % des répondants. Une forme d'imputation par la méthode « hot deck » a été employée pour attribuer des valeurs imputées aux éléments manquants, afin de pouvoir calculer un résultat global à l'IGS pour ces répondants.

Les enregistrements comportant une réponse à tous les éléments de l'IGS ont servi de donneurs. Douze groupes de donneurs ont été créés. À l'intérieur de chacune des trois states de caractéristique (établissements pour personnes âgées, établissements de soins cognitifs, autres établissements de réadaptation) les donneurs ont été classés en quatre catégories d'âge et de sexe (hommes/femmes; 0 à 64 ans/65 ans et plus). Les receveurs ont été classés de la même manière. Une valeur partielle de l'IGS (IGSP) a été produite pour chaque receveur à partir des éléments comportant une réponse. À l'intérieur du groupe correspondant de donneurs, une valeur IGSP a été produite pour chaque donneur au moyen des mêmes éléments. La personne du groupe de donneurs dont la valeur IGSP se rapprochait le plus de la valeur IGSP du receveur est devenue le donneur par imputation de cette personne. Les éléments manquants de l'enregistrement du receveur ont été imputés à partir de l'enregistrement du donneur et une valeur IGS imputée a été calculée pour le receveur. Dans le cas où plusieurs donneurs possibles avaient la même valeur IGSP, l'un d'eux a été choisi au hasard pour être le donneur. Les enregistrements comportant au moins un élément imputé sont désignés par IMPFLAG=1.

²Pour plus de renseignements sur le calcul de l'IGS, voir Annexe D: les variables dérivées.

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

Cette section décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent de quelque façon que ce soit des données tirées du fichier de microdonnées de l'enquête. Les chiffres obtenus à l'aide de ces lignes directrices devraient correspondre aux estimations produites par Statistique Canada.

9.1 Lignes directrices pour l'arrondissement

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

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

publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

- f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées de quelque façon que ce soit par les utilisateurs. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour le volet établissements de soins de santé de l'ENSP n'était pas autopondéré. Les personnes faisant partie de l'échantillon n'ont pas toutes le même poids d'échantillonnage. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids d'échantillonnage.

Si les poids exacts ne sont pas employés, les estimations calculées à partir du fichier de microdonnées ne pourront être considérées comme représentatives de la population observée et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs doivent également noter qu'en raison de la manière dont ils traitent le champ du poids, certains logiciels ne permettent peut-être pas d'obtenir des estimations qui correspondent exactement à celles de Statistique Canada.

9.2.1 Définitions des genres d'estimations : de type nominal par opposition à quantitatives

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

Estimations de type nominal :

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

Exemple de question de type nominal :

SMOK-Q62 Actuellement, fumez-vous 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 quotidiennement. Le numérateur est une estimation du nombre total de cigarettes que fument par jour les personnes qui fument quotidiennement. Le dénominateur est une estimation du nombre de personnes qui fument quotidiennement.

Exemple de question quantitative :

SMOK-Q64 Actuellement, combien de cigarettes fumez-vous chaque jour?

|_|_| Nombre de cigarettes

9.2.2 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre de personnes qui possédant une caractéristique donnée à partir du fichier de microdonnées en additionnant les poids de tous les enregistrements qui contiennent la ou les caractéristiques voulues. Les proportions et les rapports de la forme \hat{Y} / \hat{X} s'obtiennent :

- en additionnant les poids des enregistrements qui contiennent la caractéristique voulue pour le numérateur (\hat{Y}),
- en additionnant les poids des enregistrements qui contiennent la caractéristique voulue pour le dénominateur (\hat{X}),
- en divisant 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. Ensuite, on additionne cette quantité pour tous les enregistrements voulus. Par exemple, pour obtenir une estimation du nombre total de cigarettes que fument chaque jour les personnes qui fument quotidiennement, on multiplie la valeur déclarée à la question SMOK-Q64 par le poids de l'enregistrement, puis on additionne ce résultat pour tous les enregistrements dont la réponse à la question SMOK-Q62 est « tous les jours ».

Afin d'obtenir une moyenne pondérée exprimée sous la forme \hat{Y} / \hat{X} , le numérateur (\hat{Y}) est calculé 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 quotidiennement, il faut :

- estimer le nombre total de cigarettes que fument par jour les personnes qui fument quotidiennement, selon la méthode décrite ci-haut,
- estimer le nombre de personnes qui fument quotidiennement en additionnant les poids de tous les enregistrements dont la réponse à la question SMOK-Q62 est « tous les jours », puis
- diviser l'estimation obtenue en (a) par celle calculée en (b).

9.3 Lignes directrices pour l'analyse statistique

Le volet établissements de soins de santé de l'ENSP a un plan d'échantillonnage complexe qui comporte une stratification, de multiples degrés de sélection et des probabilités inégales de sélection des répondants. L'utilisation des données tirées d'enquêtes aussi complexes pose des difficultés aux analystes, car le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance à utiliser.

De nombreuses méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids. Cependant, le sens ou la définition de ces poids ne convient pas à une enquête par sondage. En général, les estimations obtenues au moyen de ces progiciels sont exactes, mais les variances calculées n'ont pratiquement aucune signification.

Dans le cas de nombreuses techniques d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de variance), il existe un moyen de rendre l'application des progiciels courants plus significative. Si l'on transforme les poids des enregistrements contenus dans le fichier de manière à ce que le poids moyen soit un (1), les résultats produits par les progiciels courants seront plus raisonnables. Même s'ils ne permettront pas de stratifier ou de répartir en grappes le plan d'échantillonnage, ils tiendront compte des probabilités inégales de sélection. On peut effectuer cette transformation en employant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux des unités d'échantillonnage (personnes) qui contribuent à l'estimateur en question.

9.4 Lignes directrices pour la diffusion

Avant de publier ou de diffuser une estimation quelconque tirée du fichier de microdonnées, il faut d'abord déterminer le nombre de résidents échantillonnés qui ont fourni les données ayant servi au calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée, quel que soit le c.v. de cette estimation. Pour les estimations pondérées basées sur des échantillons dont la taille est égale ou supérieure à 30, il faut calculer le c.v. de l'estimation arrondie et suivre les lignes directrices ci-dessous.

Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	C.v. (en %)	Lignes directrices
1. Sans réserve	0.0 à 16.5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est requise.
2. Sous réserve	16.6 à 33.3	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs suivants quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être désignées par la lettre E (ou d'une autre manière semblable).
3. Diffusion interdite	33.4 ou plus	Les estimations ne peuvent en aucun cas être diffusées sous quelque forme que ce soit. Dans les tableaux statistiques, ces estimations doivent être supprimées et remplacées par des tirets (--).

Le coefficient de variation (c.v.) est l'erreur-type (égale à la racine carrée de la variance de l'estimation), multipliée par 100 et divisée par l'estimation. Pour plus de détails sur le calcul de la variance, voir la section 11.

10. Pondération

Les poids attribués aux unités répondantes du volet établissements de soins de santé de l'ENSP étaient fondés sur les probabilités de sélection des résidents et sur les corrections pour la non-réponse au niveau de la strate ou de l'établissement.

Notation :

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

10.1 Probabilités de sélection

10.1.1 Probabilité de sélection d'un établissement

Les établissements ont été sélectionnés à partir de la base de sondage, la probabilité de sélection étant proportionnelle au nombre de lits. Par conséquent, dans la plupart des cas, la probabilité de sélection d'un établissement i était la suivante :

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

Dans les cas où un bureau central a été sélectionné (pour plus de détails, voir la section 5.3), la probabilité de sélection était la suivante :

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

où $P_{h,i,j}$ est la probabilité qu'un établissement j relevant du bureau central 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 bureau central i , à l'exclusion du plus grand.

10.1.2 Probabilité de sélection d'un résident à l'intérieur d'un établissement

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

$$\frac{r_{h,i}}{L_{h,i}} \text{ si } L_{h,i} \geq r_{h,i}$$

ou 1 si $L_{h,i} < r_{h,i}$

10.2 Calcul des poids et corrections

À ce stade, les poids de base peuvent être calculés tant pour les établissements que pour les résidents. Cependant, il peut y avoir non-réponse à ces deux niveaux. Des corrections doivent être apportées aux deux niveaux pour tenir compte des unités non répondantes.

10.2.1 Poids de base pour les établissements

Les poids des établissements correspondent au nombre d'établissements représentés par l'établissement échantillonné. Le **poids de base de l'établissement** est égal à l'inverse de la probabilité de sélection de l'établissement.

10.2.2 Correction pour la non-réponse au niveau des établissements

Dans les cas où des interviews n'ont pas eu lieu dans un établissement admissible sélectionné, une correction doit être apportée aux autres établissements faisant partie de la même strate de taille, afin de tenir compte de l'établissement non répondant. Cette correction prend la forme suivante :

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

Le **poids final de l'établissement** s'obtient en multipliant cette correction par le poids de base de l'établissement.

10.2.3 Poids de base pour les personnes

Le **poids de base de la personne** s'obtient en multipliant le poids final de l'établissement par l'inverse de la probabilité de sélection d'un résident à l'intérieur de l'établissement.

10.2.4 Correction pour la non-réponse au niveau des personnes

Certains résidents sélectionnés n'ont pas répondu au questionnaire. Une correction supplémentaire doit être apportée aux poids pour tenir compte des non-répondants. Cette correction est semblable à celle pour la non-réponse au niveau des établissements.

$$\frac{\text{nombre de résidents sélectionnés répondants et non répondants}}{\text{nombre de résidents sélectionnés répondants}}$$

Cette correction se fait au niveau de l'établissement.

Le **poids final de la personne** indiqué dans le fichier s'obtient en multipliant le poids de base de la personne par la correction pour la non-réponse au niveau de la personne. Étant

donné que l'enquête est axée sur les personnes et non sur les établissements, le poids final des établissements n'est pas indiqué dans le fichier.

11. Calcul de la variance

Une formule simple et bien connue a été employée pour calculer les variances et les coefficients de variation des estimations dans le volet établissements de soins de santé de l'ENSP. Cette formule suppose que les établissements sont sélectionnés avec probabilités inégales et avec remise. En réalité, les établissements ont été sélectionnés sans remise, c'est-à-dire qu'une fois sélectionné, un établissement ne pouvait pas être choisi de nouveau. Cela a été fait pour des raisons d'ordre opérationnel et non pour améliorer la variance; par conséquent, l'emploi d'une formule supposant un échantillonnage avec remise devrait avoir un effet négligeable.

Un programme de calcul de la variance, créé à l'aide du système SAS, accompagne le fichier de microdonnées. Ce programme peut être utilisé pour calculer les variances des moyennes et des totaux. Les formules employées pour calculer les variances d'un total Y ou d'un rapport $R=Y/X$ sont les suivantes :

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

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

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

N est le nombre de strates

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

\hat{X}_h est l'estimation de la strate h pour la variable X du dénominateur du rapport basée sur tous les répondants dans la strate h

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

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

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

11.1 Exécution du programme de calcul de la variance

Le programme SAS qui accompagne le fichier de microdonnées permet à l'utilisateur de calculer des moyennes et des totaux avec facilité. Voici les étapes à suivre pour exécuter le programme :

ÉTAPE 1 : À l'étape des données, sous « ÉTAPE 1 », l'utilisateur indique les variables pour lesquelles il veut obtenir des totaux et/ou des rapports. 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; les variables 0/1 sont désignées tot1-tot3. Dans l'exemple de rapport, les variables sont désignées num1-num4 et denom1-denom4, où num1/denom1 indique un rapport nécessitant une estimation de la variance. Il ne faut pas changer la désignation des variables, c'est-à-dire tot1-tot n , num1-num m et denom1-denom m , car le programme utilise automatiquement ces noms. Dans l'instruction «keep» à la fin de cette étape, changez les variables tot n , num m et denom m de manière à indiquer le nombre de totaux ou de rapports à calculer.

Les estimations quantitatives se calculent de la même façon. La seule différence est que la variable 0/1 est remplacée par une variable de quantité, où la quantité représente la valeur de la caractéristique que possède 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 de la même façon que n'importe quel autre programme SAS.

Note : Le programme comporte des sous-routines différentes, selon qu'il s'agit de calculer des totaux ou des rapports. Si l'utilisateur veut calculer seulement des totaux ou seulement 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 voulue entre /* et */. Par exemple, si l'utilisateur veut calculer seulement des totaux, il doit alors indiquer :

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

Seul la sous-routine de calcul de la variance des totaux s'exécutera.

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.

Si le programme est exécuté en utilisant les variables de l'exemple, voici à quoi ressemblera le résultat :

Estimations, variances et CV des totaux					1
OBS	DESCRIPTION	TOTAL	VARIANCE	CV	
1	Pop. tot.	227841.79	22325494.69	2.07	
2	Pop 65+	186496.38	22765493.02	2.56	
3	Femmes 80+	99343.07	15047102.4	3.9	
Estimations, variances et CV des taux					2
OBS	DESCRIPTION	TAUX	VARIANCE	CV	
1	Liens étroits avec membres du personnel 65+	0.39964	0.00039088	4.95	
2	Anglais 65+	0.56058	0.00031905	3.19	
3	Français 65+	0.15752	0.00012785	7.18	
4	Anglais/Français 65+	0.10677	0.00009991	9.36	