

Enquête sur la santé dans les collectivités canadiennes, Cycle 2.1
Iiyiyiu Aschii, 2003

Méthodes de l'enquête

AUTEUR

Denis Hamel
Unité Connaissance-surveillance, direction Recherche, formation et développement
Institut national de santé publique du Québec

ÉDITION ET COORDINATION

Gilles Légaré, Marie-Anne Kimpton et Mélanie Anctil
Unité Connaissance-surveillance, direction Recherche, formation et développement
Institut national de santé publique du Québec

ANALYSES STATISTIQUES

Denis Hamel
Unité Connaissance-surveillance, direction Recherche, formation et développement
Institut national de santé publique du Québec

AVEC LA COLLABORATION DE

Jill Torrie, Yv Bonnier-Viger, Marcellin Gangbè, Elena Kuzmina et Pierre Lejeune
Direction de santé publique
Conseil cri de la santé et des services sociaux de la Baie James

LECTEUR

Louis Rochette
Institut national de santé publique du Québec

MISE EN PAGE

Line Mailloux
Unité Connaissance-surveillance, direction Recherche, formation et développement
Institut national de santé publique du Québec

PUBLICATION

Conseil cri de la santé et des services sociaux de la Baie James &
Institut national de santé publique du Québec

***Ce document est disponible en version intégrale
sur le site Web du CCSSBJ au www.creepublichealth.org
et sur le site Web de l'INSPQ au www.inspq.qc.ca***

***Reproduction autorisée à des fins non commerciales
à la condition d'en mentionner la source.***

Photo : Jimmy Sam
Conception graphique : Katya Petrov

Document déposé à Santécom (<http://www.santecom.qc.ca>)
Dépôt légal – 2^e trimestre 2008
Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada
ISBN : 978-2-550-52898-2 (PDF)
© Conseil cri de la santé et des services sociaux de la Baie James (2008)

TABLE DES MATIÈRES

AVANT-PROPOS	4
INTRODUCTION	4
1. CONTEXTE DE L'ÉTUDE	5
2. PLAN D'ÉCHANTILLONNAGE	6
Population cible.....	6
Base de sondage	6
Taille et répartition de l'échantillon	6
Sélection de l'échantillon de logements.....	6
Échantillonnage des personnes interviewées.....	6
3. COLLECTE DES DONNÉES	7
4. PONDÉRATION	7
Ajustement 0 – Poids de sélection.....	8
Ajustement 1 – Retrait des unités hors champ	8
Ajustement 2 – Non-réponse ménage.....	8
Ajustement 3 – Poids de sélection personne	8
Ajustement 4 – Non-réponse personne.....	8
Ajustement 5 – Poststratification.....	8
5. QUALITÉ DES DONNÉES	9
Taux de réponse	9
Taux de réponse à l'échelle du ménage.....	9
Taux de réponse à l'échelle de la personne	9
Taux de réponse combiné.....	9
Erreurs dans les enquêtes	10
6. ACCÈS AUX DONNÉES ET CALCUL DE LA VARIANCE	10
Accès aux données	10
Mesures de précision des estimations.....	11
7. DIFFUSION DES RÉSULTATS	11
Données présentées dans les fascicules thématiques.....	11
Lignes directrices pour la diffusion.....	11
Comparaisons à l'intérieur de la région d'Iiyiyiu Aschii	11
Comparaisons entre la région d'Iiyiyiu Aschii et le reste du Québec.....	12
<i>Structure d'âge</i>	12
<i>Mode de collecte</i>	12
<i>Moment de la collecte</i>	12
Évolution de la situation depuis 1991 : Comparaisons entre l'Enquête Santé Québec (1991) et l'ESCC-Cri (2003).....	12
<i>Population couverte</i>	13
<i>Structure d'âge</i>	13
<i>Questionnaires</i>	13
RÉFÉRENCES	13
ANNEXE	14
Estimations à partir de l'ESCC 2003 Cycle 2.1 Technique d'estimation sur de petits domaines (<i>Small Area Estimation</i>)	14

AVANT-PROPOS

Ce fascicule présente les résultats d'une enquête de santé menée en 2003 auprès des ménages de la région d'Iiyiyiu Aschii¹. Une enquête similaire avait été réalisée par Santé Québec dans la région en 1991 (Santé Québec, 1994). Après plus de dix ans, il devenait indispensable pour la Direction de santé publique du Conseil cri de la santé et des services sociaux de la Baie James (CCSSSBJ) de disposer d'un nouveau portrait de l'état de santé de sa population. L'enquête de 2003 avait donc pour objectif de fournir des renseignements à jour sur les principaux problèmes de santé et leurs déterminants afin de mieux planifier, administrer et évaluer les divers programmes sociaux et sanitaires de la région.

Depuis 2001, la *Loi sur la santé publique* attribue aux directions de santé publique du Québec le mandat de mener des enquêtes périodiques pour évaluer l'état de santé de leur population. Pour remplir ce mandat, les régions sociosanitaires de la province – à l'exception d'Iiyiyiu Aschii et du Nunavik – participent depuis 2000-2001 à l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC) menée par Statistique Canada.

En 2003, la Direction de santé publique d'Iiyiyiu Aschii a décidé de se greffer à cette vaste démarche déjà en cours dans l'ensemble du Canada, et de conduire une enquête de type ESCC sur son territoire (Statistique Canada, 2003). Étant donné qu'elle fait partie du réseau du ministère de la Santé et des Services sociaux du Québec (MSSS), la Direction de santé publique du CCSSSBJ a pu bénéficier de l'expertise de l'Institut national de santé publique du Québec (INSPQ) pour la coordination de l'analyse des résultats. La rédaction des fascicules a été confiée à différents professionnels du réseau québécois de la santé et de la Direction de santé publique d'Iiyiyiu Aschii, ainsi qu'à des intervenants du milieu universitaire. Les analyses dévoilées présentent à la fois les résultats de divers aspects de la santé des résidents d'Iiyiyiu Aschii, mais offrent aussi des comparaisons avec les données de 1991 pour la région, et les données de 2003 pour le reste du Québec (Santé Québec, 1994; Statistique Canada, 2003). Les analyses s'adressent à tous les intervenants (professionnels, administrateurs, planificateurs et chercheurs) intéressés par l'état de santé des résidents d'Iiyiyiu Aschii.

Dix fascicules ont ainsi été produits dans le cadre de cette enquête :

- Caractéristiques démographiques et sociales de la population habitant Iiyiyiu Aschii
- Habitudes alimentaires, activité physique et poids corporel
- L'usage de la cigarette
- Habitudes de vie en matière de consommation d'alcool, de drogues et participation aux jeux de hasard et d'argent
- Pratiques préventives et changements pour améliorer sa santé
- État de santé, espérance de vie et limitation des activités
- Blessures et sécurité dans les transports
- Santé mentale
- Utilisation et appréciation des services de santé
- Méthodes de l'enquête.

Enfin, un dernier fascicule, les *Faits saillants de l'enquête*, présente brièvement l'ensemble des résultats de cette enquête sanitaire.

La réalisation de cette enquête a été rendue possible grâce à la participation de nombreux collaborateurs tout au long de son déroulement. Mentionnons notamment la contribution de Mme Jill Elaine Torrie, directrice des services spécialisés, et de M. Yv Bonnier-Viger, directeur de la santé publique du Conseil cri, lors des étapes de planification et de réalisation sur le terrain. Nous désirons aussi remercier la population crie dont on doit souligner la participation exceptionnelle.

INTRODUCTION

Ce fascicule présente un survol des aspects méthodologiques de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), volet cri. Ce survol est nécessaire à une bonne compréhension de la provenance des résultats qui sont dévoilés dans les différents fascicules thématiques. Nous exposons dans un premier temps le contexte de l'étude qui a mené, entre autres, au choix de l'enquête canadienne. Le plan d'échantillonnage et tous les autres aspects méthodologiques, tels que la collecte de données, la pondération, la qualité des données et la diffusion des résultats, sont expliqués par la suite.

¹ L'appellation crie *Iiyiyiu Aschii* est utilisée tout au long du texte et désigne la région sociosanitaire des Terres-Cries-de-la-Baie-James.

1. CONTEXTE DE L'ÉTUDE

La dernière enquête d'envergure sur la santé des Cris de la Baie James avait été conduite en 1991 par Santé Québec auprès d'un échantillon de 354 ménages (Santé Québec, 1994). Cette enquête s'inspirait de celle de 1987 menée au sud du Québec, et comportait des volets supplémentaires tels les mesures physiques et un rappel alimentaire de 24 heures. D'autres éléments spécifiques aux communautés crées avaient aussi été ajoutés. Cependant, les données de cette enquête n'ont pas toutes fait l'objet d'analyse spécifique.

Après plus de dix ans, il devenait indispensable pour le Conseil cri de la santé et des services sociaux de la Baie James (CCSSSBJ) de disposer d'un nouveau portrait fiable de l'état de santé de la population qu'il dessert. Cette seconde vaste enquête devait fournir des renseignements à jour sur les principaux problèmes de santé et leurs déterminants afin de mieux planifier, administrer et évaluer les divers programmes sociaux et sanitaires dans la région.

Au printemps 2003, deux principaux scénarios d'enquête ont été envisagés :

- reconduire simplement l'enquête proposée par Santé Québec en 1991;
- se greffer à l'Enquête sur la santé dans les collectivités canadiennes (ESCC) déjà en cours en 2003 dans l'ensemble du Canada.

La deuxième proposition fut retenue pour des raisons évidentes : 1) possibilité de procéder rapidement à la collecte de données puisque l'ESCC Cycle 2.1 était déjà en cours; 2) possibilité de comparaison avec les données recueillies auprès des autres Québécois; 3) coûts relativement faibles compte tenu des ressources déjà consenties par Statistique Canada (questionnaire validé, équipes d'interviewers, etc.); 4) possibilité de reconduire l'enquête dans le temps. En privilégiant cette enquête, il devenait par contre impossible d'apporter des modifications au questionnaire préétabli par Statistique Canada afin d'adapter les questions au contexte vécu par les Cris.

Le questionnaire planifié par Statistique Canada pour l'ESCC reprend en grande partie le contenu des enquêtes effectuées par Santé Québec en 1991 avec des thématiques assez comparables. Cela devait entre autres permettre d'effectuer certaines comparaisons dans le temps. Ce questionnaire comprend deux grandes parties : un contenu commun et un contenu optionnel. Le contenu commun comprend les questions posées à tous les répondants de l'enquête peu importe leur région

sociosanitaire de résidence. Le contenu optionnel comporte quant à lui divers modules de questions dont le choix relève des provinces. Des négociations ont eu lieu au printemps 2003 entre les représentants du CCSSSBJ et de Statistique Canada afin de sélectionner les modules les plus susceptibles de répondre aux objectifs de l'enquête et aux besoins de comparaison avec les données du reste du Québec. Le tableau 1 présente, à titre indicatif, les modules du contenu optionnel retenus pour la région d'Iiyiyiu Aschii².

Tableau 1
Présentation des modules optionnels de questions choisis par le Conseil cri et le reste du Québec

Modules	Description	Iiyiyiu Aschii	Québec
CIH	Changements pour améliorer la santé	X	
CMH	Utilisation des services - santé mentale	X	X
CPG	Jeu excessif	X	
DEN	Visites chez le dentiste	X	X
DIQ	Détresse et santé mentale (Québec)	X	X
DRV	Sécurité et conduite	X	
FDC	Choix alimentaires	X	
FIN	Insécurité alimentaire	X	X
HCS	Satisfaction du système des soins de santé	X	X
HUI	Indice de l'état de santé	X	X
IDG	Drogues illicites	X	
INC	Revenu	X	X
MAS	Maîtrise de soi	X	
NDE	Dépendance à la nicotine	X	X
NUS	Supplément relatif au personnel infirmier	X	
PCU	Examen général	X	X
PIC	Problèmes dans la communauté	X	
REP	Blessures	X	
SCA	Outils pour cesser de fumer	X	X
SCH	Usage du tabac - Étapes du changement	X	X
SPV	Valeurs spirituelles	X	
STR	Stress	X	
SWA	Satisfaction de l'accessibilité	X	X

Source : ESCC 2.1, 2003.

² On peut consulter le questionnaire complet à l'annexe A du *Guide du fichier de microdonnées à grande diffusion de l'ESCC 2.1* (Statistique Canada, 2005a).

2. PLAN D'ÉCHANTILLONNAGE

Le plan d'échantillonnage a d'abord été conçu en collaboration entre l'Institut national de santé publique du Québec (INSPQ) et le CCSSSBJ puis présenté aux méthodologistes d'enquête de Statistique Canada. Ces derniers ont revu ce plan et l'ont adapté aux exigences propres à l'ESCC. Les grandes lignes de l'échantillonnage ont été préservées dont la taille de l'échantillon nécessaire à la production de données fiables.

Population cible

Le volet de l'ESCC Cycle 2.1 auprès des résidents de la région sociosanitaire d'Iiyiyiu Aschii vise la population de 12 ans et plus vivant à domicile. Sont inclus, tous les blancs et les autochtones vivant dans les neuf communautés de cette région. En 1991, seuls les Cris faisaient partie de la population cible. Sont exclus du champ de l'enquête, les personnes vivant en institution et les membres à temps plein des Forces canadiennes.

Base de sondage

Le CCSSSBJ a fourni des listes de logements pour chacune des neuf communautés crie de la région. Ces listes ont été créées à partir de trois sources : les conseils de bande, la commission scolaire et le CCSSSBJ. Un nettoyage des listes a été effectué afin d'enlever les doublons et les logements hors du champ de l'enquête.

Taille et répartition de l'échantillon

Afin de produire des estimations fiables pour cette région sociosanitaire, un échantillon de 1 000 personnes était nécessaire. Avant de procéder à la répartition de l'échantillon, les neuf communautés ont d'abord été regroupées en quatre strates selon la taille de la population des communautés. Ensuite, l'échantillon de ces 1 000 personnes a été réparti entre chacune des strates proportionnellement à la racine carrée de la population estimée de ces strates³. Les estimations de la population par strate proviennent des données du Recensement de la population de 2001. Puis, dans les strates 3 et 4, la taille d'échantillon a été répartie à nouveau parmi les communautés proportionnellement à la racine carrée de la population estimée des communautés. Le tableau 2 présente la répartition de l'échantillon visé et réel. Il est à noter que les tailles d'échantillon initiales ont été gonflées avant la collecte

des données pour tenir compte de la non-réponse anticipée (~ 15 %) et des logements vacants (~ 5 à 20 % selon la communauté). L'échantillon final compte 920 personnes et la répartition de cet échantillon entre les neuf communautés respecte assez bien les proportions établies au départ.

Tableau 2

Répartition de l'échantillon visé et réel en fonction des strates du plan de sondage et des communautés

Strates	Communautés	Taille visée (personnes)	Taille obtenue (personnes)
1	Chisasibi	264	227
2	Mistissini	231	234
3	Waswanipi	90	75
	Waskaganish	106	96
	Wemindji	89	84
	Total	285	255
4	Whapmagoostui	60	53
	Eastmain	55	58
	Nemiscau	53	43
	Oujé-Bougoumou	52	50
Total	220	204	
Total Iiyiyiu Aschii		1 000	920

Source : ESCC 2.1 – Iiyiyiu Aschii, 2003.

Sélection de l'échantillon de logements

Dans chaque communauté, la sélection des logements a été faite à l'aide d'un échantillonnage systématique à partir de la liste des logements. Cette liste avait préalablement été triée selon l'adresse avant d'effectuer la sélection. La fraction de sondage à l'échelle du logement variait de 17 à 26 % selon la communauté avec une moyenne autour de 22 %.

Échantillonnage des personnes interviewées

Compte tenu de la taille d'échantillon visée, du nombre total de ménages dans la région d'Iiyiyiu Aschii et des contraintes opérationnelles, il a été convenu de sélectionner deux personnes par ménage lorsque c'était possible. La sélection des personnes dans un ménage a été conçue de façon à ce que tous les groupes d'âge soient bien représentés dans l'échantillon. La règle de sélection était basée sur la composition du ménage. Le tableau 3 décrit la règle de sélection des personnes dans un ménage.

³ En procédant ainsi, on maximise la précision des estimations produites à l'intérieur de chacune des quatre strates et pour l'ensemble des neuf communautés crie.

Tableau 3
Stratégie de sélection des personnes interviewées selon la composition du ménage

		Nombre de personnes âgées de 20 ans ou plus						
		0	1	2	3	4	5	6+
Nombre de personnes âgées de 12 à 19 ans	0		A	A	B	B	B	B
	1	A	A	C	C	C	C	B
	2	A	C	C	C	C	C	B
	3	B	C	C	C	C	C	B
	4	B	C	C	C	C	C	B
	5	B	C	C	C	C	C	B
	6+	B	B	B	B	B	B	B

A : Sélection de tous les membres du ménage (ménages à 1 ou 2 personnes).
B : Sélection aléatoire de deux personnes de 12 et plus dans le ménage.
C : Sélection aléatoire d'une personne âgée entre 12 et 19 ans, et sélection aléatoire d'une personne âgée de 20 ans ou plus.

Le tableau 4 présente la distribution anticipée de l'échantillon par groupe d'âge en utilisant la stratégie de sélection décrite au tableau 3.

Tableau 4
Distribution anticipée de l'échantillon par groupe d'âge

Groupe d'âge	Distribution du Recensement 2001 (%)	Distribution prévue de l'échantillon (%)
12-19 ans	21,7	21,5
20-29 ans	26,5	26,8
30-44 ans	28,7	29,9
45-64 ans	16,8	15,3
65 ans et +	6,3	6,6

3. COLLECTE DES DONNÉES

Lorsque la décision a été prise au printemps 2003 de se joindre à l'ESCC, la collecte du cycle 2.1 de l'enquête était alors en cours depuis le début de l'année dans l'ensemble du Canada. Statistique Canada ne voyait aucun inconvénient à ce que l'ESCC-Cri fasse partie de ce cycle, mais il fallait faire vite pour répondre à certaines contraintes opérationnelles. Parmi celles-ci figuraient la formation d'interviewers et une période de collecte relativement courte pour rejoindre une population vivant au rythme des cycles saisonniers. Le choix de la période de collecte s'est arrêté sur les mois d'août et de septembre afin de maximiser la participation des répondants car ces deux mois précèdent la période de chasse de l'automne chez les Cris.

Dans la très grande majorité des cas (85 %), le questionnaire a fait l'objet d'entrevue en face-à-face assistée par ordinateur (IPAO). Lorsqu'il n'était pas possible d'entrer en contact avec la personne sélectionnée au moment de la collecte de données sur le terrain, à la suite d'une absence prolongée par exemple, un rendez-vous téléphonique était alors fixé et l'entrevue se faisait par téléphone, toujours assistée par ordinateur (ITAO), au cours du dernier trimestre de l'année 2003. L'entrevue se déroulait dans la langue choisie par le répondant. Plus de 87 % des répondants ont préféré l'anglais, un peu moins de 7 % le français et environ 5 % le cri.

Pour connaître les autres aspects de la collecte des données communes à toutes les collectivités canadiennes, consultez la section 6 du *Guide du fichier de microdonnées à grande diffusion de l'ESCC 2.1* (Statistique Canada, 2005a).

4. PONDÉRATION⁴

Pour que les estimations produites à partir de données d'enquête soient représentatives de la population à l'étude, et non pas uniquement de l'échantillon comme tel, l'utilisateur doit incorporer les facteurs de pondération, appelés ici *poids d'enquête*, dans ses calculs. Un poids d'enquête est attribué à chaque personne comprise dans l'échantillon final, c'est-à-dire l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de personnes représentées par le répondant dans l'ensemble de la population. Le tableau 5 présente un sommaire des différents ajustements faisant partie de la stratégie de pondération dans l'ordre où ils ont été appliqués.

Tableau 5
Liste des ajustements dans la pondération

0	Poids de sélection ménage
1	Retrait des unités hors champ
2	Non-réponse ménage
3	Poids de sélection personne
4	Non-réponse personne
5	Poststratification

⁴ Cette section a été préparée par Statistique Canada et fait partie intégrante d'un document de travail remis au Conseil cri.

Ajustement 0 – Poids de sélection

Étant donné que l'échantillon initial est sélectionné à l'aide d'un échantillonnage systématique de logements à l'intérieur de chacune des neuf communautés, le poids de sélection (le poids initial) correspond à l'inverse de la probabilité de sélection d'un logement. Cette probabilité de sélection varie d'une communauté à l'autre. Le poids de sélection est alors nommé le poids 0.

Ajustement 1 – Retrait des unités hors champ

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est identifiée, lors de la collecte, comme étant hors du champ de l'enquête. Des logements détruits ou en construction, des logements vacants, saisonniers ou secondaires, de même que des établissements, sont tous des exemples de cas hors champ pour l'ESCC. Ces logements sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Ces derniers conservent donc le même poids qu'à l'étape précédente que l'on appelle maintenant poids 1.

Ajustement 2 – Non-réponse ménage

Lors de la collecte, une certaine proportion des ménages interviewés a inévitablement résulté en une non-réponse. Ceci survient habituellement lorsque le ménage refuse de participer à l'enquête, fournit des données inutilisables, ou encore, ne peut être rejoint pour réaliser l'entrevue. Le poids des ménages non-répondants est redistribué aux répondants selon des classes de réponse. Cependant, la seule information auxiliaire disponible pour les ménages non-répondants est la communauté, par conséquent, les classes de non-réponse correspondent simplement aux neuf communautés. À l'intérieur de chacune de ces communautés, un facteur d'ajustement est donc calculé :

$$\frac{\sum \text{poids 1 pour tous les ménages}}{\sum \text{poids 1 pour tous les ménages répondants}}$$

Le poids 1 des ménages répondants est ensuite multiplié par ce facteur d'ajustement pour produire le poids 2. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

Ajustement 3 – Poids de sélection personne

Puisque l'unité d'échantillonnage finale pour l'ESCC est la personne, le poids-ménage calculé jusqu'ici doit être converti en poids-personne. Celui-ci est obtenu en multipliant le poids 2 par l'inverse de la probabilité de sélection de la personne choisie dans le ménage pour ainsi obtenir le poids 3. Il est bon de noter que deux

personnes par ménage peuvent être sélectionnées pour l'enquête et que la probabilité de sélection des personnes dans le ménage dépend de la composition du ménage (voir la section Échantillonnage des personnes interviewées pour les détails sur l'algorithme de sélection des personnes).

Ajustement 4 – Non-réponse personne

Dans le cadre du volet de l'ESCC auprès des Cris d'Iiyiyiu Aschii, une entrevue peut être vue comme un processus en deux étapes. Dans un premier temps, l'intervieweur obtient la liste complète des personnes vivant dans le ménage, puis procède par la suite à l'entrevue de la ou des personnes sélectionnées dans le ménage. Dans certains cas, les intervieweurs ne réussissent qu'à compléter la première étape, soit parce qu'ils ne peuvent entrer en contact avec la ou les personnes sélectionnées, ou encore parce que la ou les personnes sélectionnées refusent d'être interviewées. De tels cas sont définis comme étant des non-réponses à l'échelle de la personne, et un facteur d'ajustement doit être appliqué aux poids des personnes répondantes pour compenser cette non-réponse. Tout comme pour la non-réponse à l'échelle du ménage, l'ajustement est appliqué à l'intérieur de classes définies à partir des caractéristiques disponibles pour les répondants et non-répondants. Toutes les caractéristiques recueillies lors du listage des membres du ménage étaient en fait disponibles pour créer ces classes. L'algorithme CHAID (*Chi-Square Automatic Interaction Detector*), disponible dans Knowledge Seeker⁵, permet d'identifier les caractéristiques qui divisent le mieux l'échantillon en groupes selon leurs propensions à répondre. Les caractéristiques suivantes ont été utilisées pour former les classes d'ajustement : le nombre de personnes sélectionnées dans le ménage, le sexe, le groupe d'âge et la communauté. Un facteur d'ajustement est calculé à l'intérieur de chaque classe de la façon suivante :

$$\frac{\sum \text{poids 3 pour toutes les personnes sélectionnées}}{\sum \text{poids 3 pour toutes les personnes sélectionnées répondantes}}$$

Le poids 3 des personnes répondantes est donc multiplié par ce facteur d'ajustement pour produire le poids 4. Les personnes non-répondantes sont éliminées de la pondération à partir de ce point.

Ajustement 5 – Poststratification

La dernière étape nécessaire afin d'obtenir le poids final est la poststratification. Celle-ci est appliquée afin

⁵ ANGOSS Software (1995). Knowledge Seeker IV for Windows - User's Guide. ANGOSS Software International Limited.

d'assurer que la somme des poids finaux corresponde aux estimations de population de la région d'Iiyiyiu Aschii pour chacun des 10 groupes d'âge-sexe d'intérêt, c'est-à-dire les groupes d'âge 12-19, 20-29, 30-44, 45-64 et 65 et plus, pour chacun des deux sexes. Les estimations de population utilisées pour l'année de référence 2003 sont basées sur les comptes du Recensement, de même que sur les estimations des comptes de naissance, décès, immigration et émigration. Le poids 4 est donc ajusté afin d'obtenir le poids final 5, à l'aide du facteur d'ajustement défini comme suit :

$$\frac{\text{Estimation de population pour le groupe âge-sexe du répondant}}{\sum \text{ poids 4 pour le groupe âge-sexe du répondant}}$$

Le poids 5 correspond au poids final de l'enquête auprès des résidents de la région sociosanitaire d'Iiyiyiu Aschii que l'on retrouve dans le fichier maître portant le nom de variable WTSC_M.

5. QUALITÉ DES DONNÉES

Taux de réponse⁶

Au total, et après avoir retiré les unités hors du champ de l'enquête, 646 ménages ont été sélectionnés pour participer au volet de l'ESCC auprès des résidents de la région sociosanitaire d'Iiyiyiu Aschii. De ce nombre, 581 ont accepté de participer à l'enquête, ce qui résulte en un taux de réponse à l'échelle du ménage de 89,9 %. Parmi ces ménages répondants, 1 074 personnes ont été sélectionnées pour participer à l'enquête parmi lesquelles 920 ont accepté, ce qui résulte en un taux de réponse à l'échelle de la personne de 85,7 %. À l'échelle de la région d'Iiyiyiu Aschii, un taux de réponse combiné de 77,9 % a donc été observé. Il est important de mentionner que le taux de réponse combiné n'est pas obtenu en multipliant les taux de réponse à l'échelle du ménage et de la personne car il y a eu une sélection de deux personnes dans plusieurs ménages. Le tableau 6 donne les taux de réponse combinés par communauté ainsi que l'information pertinente au calcul de ceux-ci.

Nous décrivons dans ce qui suit de quelle façon les différentes composantes de l'équation doivent être manipulées afin de calculer correctement les taux de réponse combinés.

HHS = nombre de ménages faisant partie du champ de l'enquête

HHR = nombre de ménages répondants

PS1 = nombre de personnes sélectionnées pour participer à l'enquête parmi les ménages où une seule personne a été sélectionnée

PS2 = nombre de personnes sélectionnées pour participer à l'enquête parmi les ménages où deux personnes ont été sélectionnées

PR1 = nombre de personnes répondantes parmi les ménages où une seule personne a été sélectionnée

PR2 = nombre de personnes répondantes parmi les ménages où deux personnes ont été sélectionnées

Taux de réponse à l'échelle du ménage

$$\text{HHRR} = \text{HHR} / \text{HHS}$$

Taux de réponse à l'échelle de la personne

$$\text{PPR} = \text{PR1} + \text{PR2}$$

$$\text{PPS} = \text{PS1} + \text{PS2}$$

$$\text{PPRR} = \text{PPR} / \text{PPS}$$

Taux de réponse combiné

$$\text{CBRR} = [\text{PR1} + (\text{PR2} / 2)] / \text{HHS}$$

Voici maintenant, étape par étape, un exemple de calcul du taux de réponse combiné utilisant l'information fournie au tableau 6.

$$\text{HHRR} = \text{HHR} / \text{HHS} = 581 / 646 = 89,9 \%$$

$$\text{PPR} = \text{PR1} + \text{PR2} = 87 + 833 = 920$$

$$\text{PPS} = \text{PS1} + \text{PS2} = 88 + 986 = 1\,074$$

$$\text{PPRR} = \text{PPR} / \text{PPS} = 920 / 1\,074 = 85,7 \%$$

$$\text{CBRR} = [\text{PR1} + (\text{PR2} / 2)] / \text{HHS} = [87 + (833 / 2)] / 646 = 77,9 \%$$

⁶ Cette sous-section a été préparée par Statistique Canada et fait partie intégrante d'un document de travail remis au Conseil cri.

Tableau 6

Présentation des taux de réponse et du nombre de personnes sélectionnées pour l'ESCC-Cri en fonction de la communauté d'appartenance des répondants

Communautés	Chisasibi	Mistissini	Waskaganish	Wemindji	Waswanipi	Eastmain	Nemiscau	Oujé-Bougoumou	Whapmagoostui	Total Iiyiyiu Aschii
Ménages										
Ménages cibles (Nbr)	175	158	64	56	62	31	30	35	35	646
Ménages répondants (Nbr)	153	144	57	54	50	31	29	30	33	581
Taux de réponse (%)	87,4	91,1	89,1	96,4	80,6	100	96,7	85,7	94,3	89,9
1 personne sélectionnée										
Personnes sélectionnées (Nbr)	23	21	10	9	6	1	6	5	7	88
Répondants (Nbr)	23	21	10	9	5	1	6	5	7	87
Taux de réponse (%)	100	100	100	100	83,3	100	100	100	100	98,9
2 personnes sélectionnées										
Personnes sélectionnées (Nbr)	260	246	94	90	88	60	46	50	52	986
Répondants (Nbr)	204	213	86	75	70	57	37	45	46	833
Taux de réponse (%)	78,5	86,6	91,5	83,3	79,5	95	80,4	90	88,5	84,5
Toutes les personnes										
Personnes sélectionnées (Nbr)	283	267	104	99	94	61	52	55	59	1 074
Répondants (Nbr)	227	234	96	84	75	58	43	50	53	920
Taux de réponse (%)	80,2	87,6	92,3	84,8	79,8	95,1	82,7	90,9	89,8	85,7
Taux de réponse combiné (%)	71,4	80,7	82,8	83	64,5	95,2	81,7	78,6	85,7	77,9

Source : ESCC 2.1 – Iiyiyiu Aschii, 2003.

Erreurs dans les enquêtes

Statistique Canada décrit généralement deux types d'erreurs lors de l'interprétation des données d'enquête : l'erreur due à l'échantillonnage et les erreurs non dues à l'échantillonnage. Pour ce document, on s'attardera principalement à l'erreur due à l'échantillonnage tout en renvoyant le lecteur au guide du fichier de microdonnées pour de plus amples informations sur les erreurs non dues à l'échantillonnage (Statistique Canada, 2005a).

Puisque la présente enquête permet de produire des estimations provenant d'un échantillon de personnes, celles-ci peuvent être quelque peu différentes de celles obtenues à partir d'un recensement complet de la population de la région d'Iiyiyiu Aschii. Les estimations ainsi produites comportent une erreur d'échantillonnage et si nous voulons les utiliser pour l'inférence, il faut chercher à quantifier l'ampleur de cette erreur. Une mesure communément utilisée, le coefficient de variation, est l'écart-type exprimé en fonction de l'estimation à laquelle il se rapporte. C'est cette mesure qui sera prise en compte pour dire si l'estimation possède une précision acceptable. Le

calcul des écarts-types des estimations sera expliqué plus en détail dans la prochaine section.

6. ACCÈS AUX DONNÉES ET CALCUL DE LA VARIANCE

Accès aux données

Le CCSSSBJ s'est orienté pour avoir plus d'autonomie dans l'accès aux données de santé recueillies sur son territoire. Une entente spéciale a ainsi été conclue avec Statistique Canada afin que le Conseil cri puisse avoir accès aux données intégrales fournies par les répondants de l'enquête. Statistique Canada doit cependant demander le consentement à chaque répondant. Pour cette enquête, 849 des 920 répondants initiaux, soit un peu plus de 92 %, ont donné leur accord pour que l'agence statistique puisse partager les données avec le CCSSSBJ. À la fin du traitement des données, un fichier de données, appelé fichier-partage, a alors été transmis et est identique au fichier de données maître de Statistique Canada avec 849 répondants.

L'entente stipule que seuls le CCSSSBJ, ses membres et ses employés directs, peuvent utiliser ces données. Ainsi, l'Institut national de santé publique du Québec, mandaté pour analyser les données, n'a aucunement accès à ces dernières. Après quelques discussions avec les représentants de Statistique Canada, l'agence a donné l'autorisation d'interroger le fichier-maître par l'intermédiaire de leur service de téléaccès. Ce service consiste à envoyer des programmes SAS aux représentants de Statistique Canada afin que ceux-ci les soumettent à leur fichier-maître. Après vérifications des règles de diffusion (voir, plus bas, la section Diffusion des résultats), les résultats sont retransmis aux utilisateurs sous forme de tableaux. En ayant recours à ce service, nous augmentons sensiblement le potentiel d'analyse puisque nous avons ainsi accès aux réponses de tous les répondants de l'enquête ESCC-Cri.

Mesures de précision des estimations

Comme le plan d'échantillonnage est relativement complexe, le calcul direct de la précision des estimations, en l'occurrence de la variance, est souvent ardu et n'est souvent pas possible. Pour pallier ce problème, il est commun de recourir à des techniques de rééchantillonnage tels le *Jackknife* et le *Bootstrap*. Pour l'ESCC, Statistique Canada a produit avec son fichier-maître un ensemble de 500 poids de bootstrap que l'on peut facilement utiliser avec le service de téléaccès. En gros, il s'agit de calculer 500 fois l'estimation désirée et la variance de ces 500 estimations deviendra alors la variance recherchée.

Statistique Canada met à la disposition des utilisateurs des syntaxes SAS ou SPSS (programme *Bootvarf_v30.sas*) pour obtenir les variances à l'aide de la technique du bootstrap (Statistique Canada, 2005b). Pour produire les estimations présentées dans les fascicules thématiques, nous avons utilisé ces programmes en les adaptant quelque peu.

7. DIFFUSION DES RÉSULTATS

Données présentées dans les fascicules thématiques

Les données présentées dans les divers tableaux sont des proportions ou des prévalences. Pour être inférées à la population de la région sociosanitaire d'Iiyiyiu Aschii, donc être représentatives de cette population, toutes les proportions obtenues de l'échantillon doivent être pondérées. La variable de poids considérée est celle décrite à la section Pondération. Les proportions de l'ESCC 2003 pour le Québec et pour le Canada sont également pondérées pour les mêmes raisons.

De plus, même si des comparaisons peuvent être faites entre deux populations ayant des structures d'âge fort différentes (par exemple, Iiyiyiu Aschii contre le reste

du Québec), les proportions présentées dans tous les fascicules thématiques n'ont fait l'objet d'aucun ajustement (estimations brutes). De cette manière, les estimations présentées reflètent la situation réelle observée chez les populations visées.

Lignes directrices pour la diffusion

Toutes les estimations présentées dans les fascicules thématiques respectent les règles de diffusion stipulées par Statistique Canada. Ainsi, aucune diffusion n'est permise pour les proportions dont le coefficient de variation dépasse 33,3 % ou dont le nombre de répondants dans l'échantillon associé au calcul du numérateur est inférieur à 10. Dans les tableaux de résultats, les lettres « NP » (non publiées) remplacent ces valeurs imprécises. Pour ce qui est des estimations présentant une forte variabilité, soit un coefficient de variation compris entre 16,6 % et 33,3 %, elles sont diffusées avec restriction. Un astérisque «*» accompagne ces estimations et indique une variabilité importante qu'il faut interpréter avec circonspection.

Par ailleurs, pour améliorer la précision de certaines estimations produites à l'échelle des sous-régions (communautés côtières vs continentales et strates du plan de sondage), des techniques d'estimation sur de petits domaines (*Small Area Estimation*) ont été considérées. L'estimateur composite empirique de Bayes⁷ a été ainsi retenu dans quelques cas et les résultats qui découlent de cet estimateur sont clairement identifiés dans les fascicules thématiques.

Comparaisons à l'intérieur de la région d'Iiyiyiu Aschii

Les analyses descriptives ne seraient pas complètes sans les comparaisons que procurent les croisements avec d'autres variables. Par exemple, il est fort intéressant de vérifier si les indicateurs de santé se comportent différemment d'un groupe d'âge à l'autre ou d'un sexe à l'autre. Pour ce faire, plusieurs tests de proportion ont été effectués à l'aide de la statistique suivante :

$$Z = \frac{p_i - p_j}{\sqrt{\text{Var}(p_i) + \text{Var}(p_j)}}$$

où p_i et p_j sont respectivement les proportions brutes estimées des modalités i et j et où $\text{Var}(p_i)$ et $\text{Var}(p_j)$ correspondent aux variances obtenues à l'aide des poids de bootstrap pour ces mêmes proportions. On suppose ici que la covariance entre les deux proportions à comparer est nulle.

⁷ Pour de plus amples renseignements sur cet estimateur, nous invitons le lecteur à consulter le document de travail *Estimations à partir de l'ESCC 2003 Cycle 2.1 - Technique d'estimation sur de petits domaines (Small Area Estimation) - Document de travail : Projet spécial pour la région Eeyou Istchee* présenté en annexe.

Le seuil retenu pour ces tests est 5 %. Puisque Z suit approximativement une loi normale de moyenne nulle et de variance 1, une différence est alors déclarée statistiquement significative si Z est inférieur à -1,96 ou supérieur à 1,96.

Outre l'âge et le sexe, les croisements avec la sous-région ou le regroupement selon la taille de la communauté sont fréquents dans les fascicules thématiques. Les communautés d'Iiyiyiu Aschii ont été divisées en deux sous-régions pour les besoins de comparaison. Les communautés côtières incluent Chisasibi, Wemindji, Eastmain, Waskaganish et Whapmagoostui, tandis que les communautés continentales regroupent Nemiscau, Mistissini, Oujé-Bougoumou et Waswanipi. Quatre regroupements de communautés ont aussi été définis en fonction de la taille de la population de chaque village au moment de l'enquête (ces regroupements correspondent aux strates du plan de sondage présentées au tableau 2) : 1. Chisasibi (plus de 3 000 habitants) ; 2. Mistissini (2 000 à 3 000 habitants) ; 3. Communautés de taille moyenne (1 000 à 2 000 habitants), c.-à-d. Waswanipi, Waskaganish et Wemindji ; 4. Communautés de petite taille (moins de 1 000 habitants), c.-à-d. Whapmagoostui, Eastmain, Nemiscau et Oujé-Bougoumou. Des croisements avec d'autres variables sont aussi possibles et varient selon la thématique des fascicules.

Comparaisons entre la région d'Iiyiyiu Aschii et le reste du Québec

L'un des principaux critères en faveur du choix de l'ESCC était la possibilité de comparer la région d'Iiyiyiu Aschii avec le reste du Québec pour les questions qui ont été posées aux répondants de ces deux populations. Les tests de comparaison s'effectuent à l'aide de la même statistique décrite précédemment mais en tenant compte des différences suivantes :

Structure d'âge

La population d'Iiyiyiu Aschii est beaucoup plus jeune que celle du reste du Québec. Comme plusieurs indicateurs de santé sont fortement associés à l'âge, il importe de tenir compte de cette disparité pour éviter de conclure à tort en une différence entre les deux populations alors que cette différence est uniquement due à l'âge. Pour contourner ce problème, il est courant d'ajuster les estimations des deux populations sur une même structure d'âge. La population de référence retenue est celle du Québec en 2001 et les groupes d'âge sont les 12-19, 20-29, 30-44 et 45 et plus (données de l'Institut de la statistique du Québec). Les tests pour ce type de comparaison ont donc été faits sur des proportions ajustées selon l'âge.

Mode de collecte

Tel que mentionné à la section Collecte de données, la très grande majorité des entrevues (85 %) réalisées dans la région d'Iiyiyiu Aschii ont été effectuées en face-à-face. Pour le reste du Québec, les entrevues téléphoniques ont été privilégiées (autour de 70 %) (St-Pierre et Béland, 2004). Côté et ses collègues (2005) ont étudié l'impact de ces deux modes de collecte sur les estimations et ont démontré des divergences importantes pour un grand nombre de variables qu'ils avaient analysées. Certaines différences significatives pourraient donc être dues à la différence entre les modes de collecte plutôt qu'à une réelle différence entre les proportions. Pour réduire ce problème, ces auteurs suggèrent d'abaisser à 1 % le seuil de signification habituel de 5 %.

Moment de la collecte

La période de collecte pour la région d'Iiyiyiu Aschii s'est échelonnée principalement sur deux mois (août et septembre), alors que pour les autres régions du Québec, les douze mois de l'année 2003 ont été couverts assez uniformément. Comme certaines questions de l'enquête font référence à une période antérieure à l'entrevue⁸, les réponses obtenues peuvent fortement dépendre du moment de l'entrevue. Pour les indicateurs qui découlent de ces questions, il est préférable de comparer les proportions de la région d'Iiyiyiu Aschii et du Québec sur la même période, soit pour les mois d'août et de septembre.

Pour conclure, les tests de comparaison entre la région d'Iiyiyiu Aschii et le reste du Québec pour l'ESCC 2003 sont effectués à un seuil de 1 % sur des proportions ajustées selon l'âge. Et pour les indicateurs qui peuvent être influencés par le moment de l'entrevue, les proportions comparées sont calculées uniquement à partir des entrevues faites aux mois d'août et de septembre.

Évolution de la situation depuis 1991 : Comparaisons entre l'Enquête Santé Québec (1991) et l'ESCC-Cri (2003)

L'analyse de l'évolution de la santé des résidents d'Iiyiyiu Aschii dans le temps est un objectif primordial. Dans un passé relativement récent, l'Enquête Santé Québec auprès des Cris de la Baie James de 1991 (ESQ 1991) était la seule véritable enquête d'envergure qui couvre plusieurs aspects de la santé et de ses déterminants. Cette enquête représente donc un point de repère indéniable. Cependant, la décision d'aller de l'avant avec l'ESCC en 2003 réduit le potentiel d'analyse pour les comparaisons car les thèmes couverts ne sont pas toujours les mêmes. De

⁸ Par exemple, pour l'activité physique on demandait au répondant s'il en avait pratiqué une au cours des trois derniers mois.

plus, on doit toujours se questionner sur la comparabilité de ces enquêtes puisque la formulation des questions et les choix de réponse peuvent différer. Chaque indicateur doit donc être traité séparément afin de vérifier la validité de la comparaison. Il serait toutefois ambitieux ici de faire le tour de tous les indicateurs retenus pour les besoins de comparaison dans le temps. Nous référons pour cette raison le lecteur aux divers fascicules thématiques dans lesquels ces indicateurs sont analysés. Par contre, il est important de mentionner les grandes lignes à suivre lorsque de telles comparaisons sont effectuées.

Population couverte

Les deux enquêtes ne couvrent pas tout à fait la même population. Tandis que l'ESCC-Cri 2003 s'adresse à toutes les personnes de 12 ans et plus (Cris et non Cris) vivant en ménage privé, l'Enquête Santé Québec de 1991 couvre uniquement les Cris âgés de 15 ans et plus vivant en ménage privé. Pour être comparables, les proportions provenant de l'ESCC-Cri 2003 ne doivent par conséquent être calculées que pour les Cris de 15 ans et plus. Il est possible de différencier les Cris des non-Cris dans l'ESCC grâce à une question portant sur l'appartenance ethnique.

Structure d'âge

En douze ans, la structure d'âge de la population crie a changé et, pour les mêmes raisons évoquées à la section Diffusion des résultats, il vaut mieux travailler avec les proportions ajustées selon l'âge pour que les comparaisons soient valides. La population de référence demeure le Québec en 2001, mais les groupes d'âge sont plutôt les suivants : 15-24, 25-44, 45-64, 65 et plus.

Questionnaires

Contrairement à l'ESCC 2003 qui ne compte qu'un seul questionnaire, l'ESQ 1991 comprend plusieurs instruments de collecte dont les deux principaux sont des questionnaires : le QRI, administré en face-à-face où une seule personne répond pour l'ensemble des membres de son ménage et le QAA, en principe auto-administré mais qui a principalement été réalisé en face-à-face pour chacune des personnes de 15 ans et plus du ménage échantillonné. Le mode de collecte en entrevue face-à-face est donc identique pour les deux enquêtes. Pour cette raison, on conserve le seuil habituel de 5 % pour les tests.

En résumé, les tests de comparaison entre l'enquête de 2003 (ESCC-Cri) et celle de 1991 (ESQ) sont effectués à un seuil de 5 % sur des proportions ajustées selon l'âge et calculées uniquement pour les personnes crie de 15 ans et plus.

RÉFÉRENCES

Côté L., Courtemanche R. & B. Caron (2005). *Comparabilité entre les cycles 1.1 et 2.1 de l'Enquête sur la santé dans les collectivités canadiennes : Impact du changement apporté à la répartition de l'échantillon selon la base de sondage*. Québec : Institut de la statistique du Québec.

Santé Québec. Daveluy, C., Lavallé, C., Clarkson, M., & Robinson, E. (dir.) (1994). *Et la santé des Cris, ça va? Rapport de l'Enquête Santé Québec auprès des Cris de la Baie James 1991*. Montréal : ministère de la Santé et des Services sociaux, gouvernement du Québec.

Statistique Canada. (2003). *Enquête sur la santé dans les collectivités canadiennes (ESCC), Cycle 2.1*. Ottawa : Division de la statistique de la santé. [En ligne]. http://www.statcan.ca/francais/concepts/health/cycle2_1/index_f.htm.

Statistique Canada (2005a). *Guide du fichier de microdonnées à grande diffusion – Enquête sur la santé dans les collectivités canadiennes (ESCC) Cycle 2.1 (2003)*, Statistique Canada.

Statistique Canada (2005b). *Estimation de la variance à l'aide des poids de bootstrap – Guide de l'utilisateur du programme BOOTVARF_V30.SAS (version 3.0)*. Statistique Canada.

St-Pierre, M. & Y. Béland (2004). *Mode effects in the Canadian Community Health Survey : a comparison of CAPI and CATI. 2004 Proceedings of the American Statistical Association Meeting, Survey research Methods Section*. Toronto: American Statistical Association.

ANNEXE

Estimations à partir de l'ESCC 2003 Cycle 2.1

Technique d'estimation sur de petits domaines (Small Area Estimation)

Document de travail : Projet spécial pour la région d'Iiyiyiu Aschii

Contexte

Pour la première fois, la région sociosanitaire d'Iiyiyiu Aschii, une région isolée du Nord du Québec, est incluse dans l'Enquête sur la santé dans les collectivités canadiennes (ESCC) au Cycle 2.1. En effet, le Conseil cri de la santé et des services sociaux de la Baie James (CCSSBJ) a acheté un échantillon supplémentaire qui permet de produire des estimations fiables de plusieurs variables ou indicateurs de la santé à l'échelle de cette région. Le choix des modules de questions suit de près celui du Québec, avec quelques exceptions pour tenir compte des problématiques particulières de cette région. Il est donc possible dans la plupart des cas de procéder à des comparaisons avec le Québec.

Comme le plan d'échantillonnage fait en sorte de couvrir chacune des neuf communautés d'Iiyiyiu Aschii, nous aimerions aller plus loin et produire des estimations à des échelles plus fines, comme les sous-régions (communautés côtières vs continentales), les regroupements selon la taille des communautés, voire même pour une communauté seulement. La technique d'estimation usuelle déduite à partir du plan d'échantillonnage, appelée ici estimateur direct, peut être utilisée dans certains cas mais présente souvent des problèmes d'instabilité dus à la faible taille d'échantillon sur ces petits domaines. Les valeurs des coefficients de variation ne respectent pas, dans la plupart des cas, les lignes directrices de Statistique Canada (voir le *Guide du fichier de microdonnées à grande diffusion de l'ESCC 2.1* (Statistique Canada, 2005a)). Cependant, il est possible d'améliorer la précision des estimations en proposant un estimateur composite qui est l'objet de ce texte.

Domaines d'estimation

Avant de décrire la technique d'estimation proprement dite, il est important de bien définir les domaines d'estimation qui nous préoccupent le plus ici. Ces domaines correspondent à des subdivisions de l'échantillon séparées selon des regroupements de villages et des catégories d'âge et de sexe. Plus précisément, nous désirons produire des estimations pour les combinaisons âge/sexe de même que pour les regroupements de régions suivants :

Sous-région :

- Communautés côtières : Chisasibi, Wemindji, Eastmain, Waskaganish, Whapmagoostui ;
- Communautés continentales : Nemiscau, Mistissini, Oujé-Bougoumou, Waswanipi.

Regroupements selon la taille de la communauté :

1. Chisasibi (plus de 3 000 habitants);
2. Mistissini (2 000 à 3 000 habitants);
3. Communautés de taille moyenne (1000 à 2000 habitants), c.-à-d. Waswanipi, Waskaganish et Wemindji;
4. Communautés de petite taille (moins de 1000 habitants), c.-à-d. Whapmagoostui, Eastmain, Nemiscau et Oujé-Bougoumou.

De même que pour les 8 combinaisons âge/sexe (12-19 ; 20-29 ; 30-44 ; 45 et plus).

Description de la technique retenue d'estimation sur de petits domaines : l'estimateur composite empirique de Bayes

La technique décrite ci-dessous est basée principalement sur les travaux de Chattopadhyay, Lahiri, Larsen et Reimnitz (1999)⁹. C'est également celle qui a été retenue par Statistique Canada pour produire des estimations de faibles prévalences pour quatre sous-régions de la province de l'Île-du-Prince-Édouard dans le cadre du Cycle 1.1 de l'ESCC. Cette technique permet d'améliorer tangiblement la précision des estimations tout en minimisant les biais.

Notation :

- i : indice identifiant la sous-région
- k : indice de la combinaison âge/sexe
- j : indice du répondant ($j = 1, \dots, n_{ik}$) où n_{ik} est la taille d'échantillon de la sous-région i pour la combinaison âge/sexe k

y_{ikj} : j ième observation (0 ou 1) de la k ième combinaison âge/sexe de la sous-région i

w_{ikj} : poids associé à l'individu de l'observation y_{ikj} .

⁹ Chattopadhyay, M., Lahiri, P., Larsen, M. et Reimnitz J. (1999). *Composite Estimation of Drug Prevalences for Sub-State Areas. Survey Methodology*, 25, no. 1, p. 81-86.

En utilisant la combinaison linéaire des estimateurs directs de la région et de la sous-région, on définit l'estimateur composite de la de la k ème combinaison âge/sexe de la sous-région i de la façon suivante :

$$\hat{\pi}_{ik}^{com} = \lambda_{ik} \hat{\pi}_{ik}^{dir} + (1 - \lambda_{ik}) \hat{\pi}_k^{dir}$$

où $\hat{\pi}_{ik}^{dir}$ représente l'estimateur direct de la k ème combinaison âge/sexe de la sous-région i et $\hat{\pi}_k^{dir}$ représente l'estimateur direct de la k ème combinaison âge/sexe de la région avec un coefficient λ_{ik} compris entre 0 et 1.

Le modèle suivant permet de proposer un estimateur empirique de Bayes de $\hat{\pi}_{ik}^{com}$ et ainsi de déterminer la valeur de la constante λ_{ik} .

Modèle

- Conditionnellement aux π_{ik} , les observations y_{ikj} sont non corrélées entre elles avec $E(y_{ikj} / \pi_{ik}) = \pi_{ik}$ et $\text{Var}(y_{ikj} / \pi_{ik}) = \pi_{ik}(1 - \pi_{ik})$
- Les π_{ik} sont également non corrélés avec comme espérance $E(\pi_{ik}) = \mu_k$ et variance $\text{Var}(\pi_{ik}) = d\mu_k^2$
- $\pi_{ik} \sim \text{Uniform}(\mu_k - \alpha\mu_k, \mu_k + \alpha\mu_k)$ où α prend une valeur comprise entre 0 et 1.
Alors, $d = \alpha^2/3$

En d'autres mots, le modèle suggère que les proportions réelles des sous-régions (pour chacune des combinaisons âge/sexe) suivent une loi uniforme dans l'intervalle $(\mu_k - \alpha\mu_k, \mu_k + \alpha\mu_k)$, donc centrées autour de la proportion réelle de la région (μ_k). La valeur α est déterminée à l'aide des données provenant des autres régions du Québec de l'ESCC Cycle 2.1. On en discutera plus en détail à la prochaine section.

L'estimateur linéaire empirique de Bayes de $\hat{\pi}_{ik}^{com}$, sous le modèle décrit précédemment et la fonction quadratique de perte associée, est le suivant :

$$\hat{\pi}_{ik}^{eb} = \hat{\lambda}_{ik} \hat{\pi}_{ik}^{dir} + (1 - \hat{\lambda}_{ik}) \hat{\pi}_k^{dir}$$

$$\text{où } \hat{\lambda}_{ik} = \frac{d(\hat{\pi}_k^{dir})^2}{\left[d(\hat{\pi}_k^{dir})^2 + c_{ik} \left(\hat{\pi}_k^{dir} - (d+1)\hat{\pi}_k^{dir} \right)^2 \right]}$$

$$\text{avec } c_{ik} = \frac{\sum_{j=1}^{n_{ik}} w_{ikj}^2}{\left(\sum_{j=1}^{n_{ik}} w_{ikj} \right)^2} \text{ et } d = \alpha^2/3$$

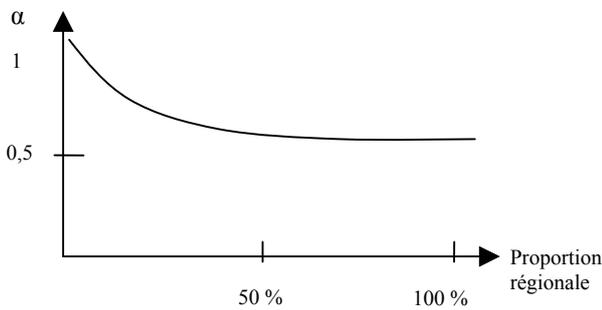
Choix de α

Dans la dernière formule décrite précédemment, tous les termes peuvent être facilement obtenus à l'exception de la constante α . Celle-ci représente l'étendue de la loi uniforme autour de la prévalence régionale pour une caractéristique donnée d'après le modèle. Nous pouvons aisément déduire que cette valeur dépendra directement de la prévalence régionale. En effet, pour une étendue similaire (ou variabilité), plus la prévalence sera importante, plus petite sera la valeur α . Pour vérifier cette affirmation et déterminer les valeurs α à utiliser pour des prévalences régionales données, nous pouvons utiliser les données de l'ESCC 2.1 pour estimer les variations régionales. Trois régions du Québec soit la région de la Capitale-Nationale, de l'Outaouais et des Laurentides ont acheté des échantillons supplémentaires afin de produire des estimations à des échelles infrarégionales : les territoires de CLSC. Nous pouvons donc évaluer la variabilité des estimations pour les domaines formés du CLSC et des huit combinaisons sexe/groupe d'âge et pour diverses caractéristiques autour des valeurs estimées pour la région correspondante. Les caractéristiques retenues sont :

- Perception de l'état de santé : excellente et très bonne
- Obésité : IMC > 30
- Fumeur
- Activité physique de loisir : inactif
- Scolarité : moins que le diplôme secondaire

Par exemple, nous avons calculé la variabilité parmi les proportions de fumeurs de chacun des huit CLSC de la région de la Capitale-Nationale autour de la proportion régionale de fumeurs chez les hommes de 12-19 ans. En répétant cet exercice pour les autres combinaisons sexe/groupe d'âge, les diverses caractéristiques retenues et les trois régions visées, on se retrouve alors avec plusieurs mesures de variabilité intrarégionale pour un même ordre de grandeur de proportion régionale. Nous pouvons alors déterminer une valeur α pour une proportion ou une prévalence régionale donnée en prenant le 95^{ème} percentile de ces mesures et en divisant par cette proportion. En mettant

les valeurs α ainsi obtenues en fonction de leurs proportions régionales associées, nous obtenons un graphique de la forme suivante :



À la lumière des résultats obtenus, nous suggérons d'estimer, à l'aide d'un modèle de régression quadratique, la valeur α pour une proportion donnée dont l'équation est décrite ci-bas. Mentionnons que les caractéristiques ayant une proportion similaire auront une même valeur α .

$$\alpha = 1,0689 - 0,03073\hat{\pi}_k^{dir} + 0,00028(\hat{\pi}_k^{dir})^2 ; R^2 = 0,6758$$

Puisque la valeur maximale de α est 1, toute valeur estimée par cette équation supérieure à 1 sera ramenée à 1.

Estimation de la variance pour un estimateur composite

Statistique Canada suggère d'utiliser les poids bootstrap pour estimer la variance des estimateurs directs de l'ESCC. Pour obtenir une estimation de la variance de l'estimateur composite, ces mêmes poids de bootstrap peuvent être utilisés sans problème avec une légère adaptation. En fait, on doit calculer séparément l'estimateur composite pour chaque échantillon bootstrap (c'est-à-dire 500 fois) où b représente l'échantillon bootstrap $b=1$ à 500.

$$\hat{\pi}_{ik(b)}^{eb} = \hat{\lambda}_{ik(b)} \hat{\pi}_{ik(b)}^{dir} + (1 - \hat{\lambda}_{ik(b)}) \hat{\pi}_{k(b)}^{dir}$$

La variabilité observée entre ces 500 estimations deviendra alors l'estimateur de la variance recherchée.

Règles de diffusion adaptées aux estimateurs composites

Pour toute production de prévalences ou de proportions obtenues à partir du fichier de l'ESCC 2003 par la méthode directe, Statistique Canada impose des règles de diffusion : le coefficient de variation (CV) de l'estimation doit être inférieur à 33 % et la taille de l'échantillon pour l'estimation doit comporter au moins 30 répondants (10 en utilisant les poids bootstrap). On essaiera dans la mesure du possible de produire et de présenter des estimations directes en respectant ces règles pour les petits domaines. Par contre, si et seulement si on ne peut répondre à ces exigences, on poursuivra avec la production d'estimations composites telles que décrites précédemment. Il faudra encore une fois que ces estimations respectent un minimum de précision, c'est-à-dire avoir un CV inférieur à 33 % et que la taille minimum pour chaque cellule soit de 10. Il se pourrait alors que, même si on privilégie une technique d'estimation sur de petits domaines, on se retrouve avec quelques cellules vides. Dans la plupart des cas, l'utilisation de la technique d'estimation composite fera en sorte que les CV diminueront tangiblement (avec une meilleure précision) et permettra ainsi de diffuser ces estimations qui autrement (avec les estimations directes) n'auraient pas été présentées. Il faudra bien identifier dans les tableaux les proportions ou prévalences obtenues grâce à la technique d'estimation composite et apposer au bas des tableaux une note appropriée.