

PLAN DE SONDAGE DE LA COMPOSANTE RÉGIONALE DE L'ENQUÊTE SUR LA SANTÉ DANS LES COLLECTIVITÉS CANADIENNES

Yves Béland¹, Johane Dufour, Suzanne Lessard, Marianna Morano,
Orane Saint-Denis et Sylvain Thivierge

RÉSUMÉ

Dans le but de remédier aux principales lacunes statistiques en ce qui a trait aux déterminants de la santé, à l'état de santé et à l'utilisation du système de santé de la population canadienne à l'échelle des régions sociosanitaires, Statistique Canada a élaboré une nouvelle enquête, l'Enquête sur la santé dans les collectivités canadiennes (ESCC). L'ESCC est une enquête constituée de deux composantes : une composante régionale la première année, auprès d'un échantillon de plus de 130 000 répondants et une composante provinciale la deuxième année auprès d'un échantillon de 30 000 répondants. La composante régionale, dont la collecte a débuté en septembre 2000, a comme but premier de produire des estimations transversales pour 136 régions sociosanitaires au Canada. La composante provinciale, dont la collecte des données débutera en janvier 2002, a comme objectif principal de produire des estimations transversales portant sur différents aspects reliés à la santé mentale et au bien-être des Canadiens. Cet article décrit le plan de sondage de la composante régionale.

MOTS CLÉS : Enquête régionale; santé; base duale; sur-échantillonnage.

ABSTRACT

With the objective of addressing priority health determinant, health status and health system utilization data gaps at the health region level, Statistics Canada has designed a new survey, the Canadian Community Health Survey (CCHS). The CCHS comprises two survey components: a regional component, which data collection started in September 2000, using a sample of more than 130,000 respondents and a provincial component, scheduled to begin collection in 2002 with a sample of 30,000 respondents. The primary goal of the regional component is to produce level estimates for 136 health regions whereas the provincial component's main objective is to produce level estimates on several aspects related to the mental health and well-being of Canadians. This paper describes the sample design of the regional component.

KEY WORDS: Regional survey; Health; Dual frame; Over-sampling.

1. INTRODUCTION

La demande d'estimations transversales infra-provinciales relatives à la santé de la population s'amplifie de plus en plus dans le contexte actuel de l'élaboration des politiques et programmes de la santé. L'actuelle enquête nationale sur la santé, l'Enquête nationale sur la santé de la population (ENSP), constitue une excellente source d'information sur la santé de la population. Toutefois, l'échantillon de l'ENSP n'est pas suffisamment vaste pour isoler les différences entre les régions infraprovinciales du point de vue de l'état de santé. Par conséquent, l'Enquête sur la santé dans les collectivités canadiennes (ESCC) vise d'abord et avant tout à remédier aux principales

lacunes statistiques qui existent à l'échelle de la région sociosanitaire en ce qui a trait aux déterminants de la santé, à l'état de santé et à l'utilisation du système de santé.

En 1998, le Conseil consultatif sur l'infrastructure de la santé, l'Institut canadien d'information sur la santé (ICIS) et Statistique Canada ont réuni plus de 500 personnes pour mettre en place une stratégie permettant d'obtenir une meilleure information sur le système de santé au Canada. Les consultations auprès d'experts ont mené à la création de *l'Initiative du Carnet de route de l'information sur la santé* (ICIS, 1999) qui présente une vision canadienne commune pour moderniser et normaliser le système

¹ Yves Béland, 16H immeuble R.H. Coats, Parc Tunney, Ottawa, Ontario, K1A 0T6, yves.beland@statcan.ca

d'information sur la santé dans les diverses régions du pays.

Au printemps 1999, Statistique Canada a mené des consultations à travers le pays auprès de plus de 225 intervenants clés du domaine de la santé afin de mieux cerner les besoins en matière de données. Ces consultations ont également permis d'établir certaines exigences qui ont dicté le développement de l'ESCC.

Cet article décrit le plan d'échantillonnage mis en place pour la composante régionale de l'ESCC. La section 2 présente les différentes options qui ont été envisagées à Statistique Canada pour réaliser l'ESCC étant donné plusieurs contraintes imposées par l'infrastructure de collecte de l'Agence. Suit ensuite, à la section 3, une discussion en profondeur sur le développement méthodologique de l'enquête de la composante régionale. On termine avec une description des plans futurs de l'ESCC.

2. STRATÉGIE D'ENQUÊTE DU CYCLE BIENNAL

Afin de répondre efficacement aux besoins des utilisateurs tels qu'exprimés lors des consultations et en tenant compte de l'infrastructure existante de collecte de l'Agence, un cycle biennal de collecte de données se veut un choix logique. En effet, la nécessité de produire i) des estimations pour un vaste éventail de sujets pour 136 régions sociosanitaires et ii) des estimations pour un sujet thématique à l'échelle provinciale exige un échantillon d'une taille telle que recueillir l'information sur une plus courte période est opérationnellement difficile à gérer.

Parmi les différents scénarios envisagés, deux ont fait l'objet d'une évaluation sérieuse. L'option A requiert le développement de deux enquêtes distinctes et séparées : une enquête à l'échelle de la région sociosanitaire la première année auprès d'un échantillon d'un peu plus de 130 000 répondants et une enquête provinciale la deuxième année portant sur un échantillon de 30 000 répondants. L'enquête régionale aurait un contenu très diversifié qui serait adapté aux besoins régionaux alors que l'enquête provinciale traiterait en profondeur d'une thématique particulière.

L'option B représente une version annualisée de l'ESCC où une enquête principale à l'échelle de la région sociosanitaire, et adaptée aux besoins régionaux, est menée sur une période de deux ans auprès d'un échantillon de 160 000 répondants. À cette enquête, un supplément provincial (avec thématique particulière) serait administré auprès d'un sous-

échantillon de 30 000 répondants.

Les lignes directrices mises en place à Statistique Canada concernant la qualité des données (Statistique Canada, 1998a) ont dicté le choix de la stratégie d'enquête pour l'ESCC. Des critères de qualité tels que la pertinence, l'exactitude, l'intelligibilité, la cohérence mais surtout l'actualité des données ont été les facteurs déterminants dans le choix de l'option A. Le cycle biennal de l'ESCC est donc formé de deux enquêtes : une enquête régionale la première année et une provinciale la deuxième.

3. PLAN DE SONDAGE DE LA COMPOSANTE RÉGIONALE

L'objectif premier de cette enquête vise d'abord et avant tout à fournir des estimations transversales fiables à l'échelle de la région sociosanitaire en ce qui a trait aux déterminants de la santé, à l'état de la santé et à l'utilisation du système de santé.

3.1. Formation des régions sociosanitaires

Les RSS sont des régions géographiques que les provinces utilisent à des fins administratives dans le domaine de la santé. Il existe 133 RSS réparties dans les 10 provinces auxquelles s'ajoutent une RSS pour chacun des trois territoires. (Comme l'enquête dans les trois territoires utilise un plan de sondage différent, le présent article ne fait référence qu'au plan utilisé dans les 10 provinces dans le but d'alléger l'écriture.)

Pour des raisons d'ordre pratique, Statistique Canada, avec l'accord des provinces, a révisé quelque peu les limites de certaines RSS selon la géographie du Recensement de la population de 1996 pour ainsi permettre la production de projections démographiques pour différents groupes âge-sexe.

3.2. Population cible

L'ESCC vise les personnes vivant dans des logements privés qui sont âgées de 12 ans ou plus. Sont exclus du champ de l'enquête les personnes qui vivent sur les réserves indiennes et sur les terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces armées canadiennes et les résidents de certaines régions éloignées. L'ESCC couvre environ 97% des 12 ans ou plus de la population canadienne.

3.3. Taille et répartition de l'échantillon

Afin de fournir des estimations fiables aux 133 RSS et en tenant compte du budget alloué à l'enquête, un échantillon net de base de 130 750 répondants est visé.

Bien que l'objectif de produire des estimations fiables à l'échelle de la RSS soit primordial, la qualité des estimations à l'échelle provinciale pour certaines caractéristiques clés demeure tout de même un but à atteindre. Divers scénarios ont été envisagés pour répartir cet échantillon entre les provinces et les RSS. Comme la taille et le nombre des RSS varient considérablement d'une province à l'autre, il est difficile d'établir un équilibre entre les besoins régionaux et provinciaux. Une approche axée principalement vers les RSS pénalise les provinces fortement peuplées ayant un nombre relativement petit de RSS tandis qu'une approche axée vers les provinces procure l'effet inverse. La stratégie qui a été adoptée est une démarche en trois étapes qui permet d'accorder une importance relativement égale aux RSS et aux provinces.

À la première étape, on répartit entre les provinces à peu près la moitié de l'échantillon total de 130 750 répondants en fonction du nombre de RSS qu'elles renferment : 500 unités d'échantillonnage sont attribuées à chaque RSS. Une exception toutefois: les RSS où la fraction de sondage résultante est trop élevée; auquel cas le taux d'échantillonnage est majoré à 1 ménage sur 20. Un total de 65 830 unités d'échantillonnage ont ainsi été réparties suite à cette première étape.

La deuxième étape consiste à répartir le reste de l'échantillon total (64 920 unités) entre les provinces proportionnellement à la taille de population avec l'exception d'attribuer 1 000 unités à la province de l'Île-du-Prince-Édouard. Le tableau 1 fournit une répartition détaillée par province.

Tableau 1 – Tailles provinciales

Prov.	Taille ('000)	# de RSS	Étape 1 500/RSS	Étape 2 Prop. X	Échant. Total
T-N	551	6	2 780	1 230	4 010
Î-P-É	135	2	1 000	1 000	2 000
N-É	909	6	3 000	2 040	5 040
N-B	738	7	3 500	1 650	5 150
Qué.	7 139	16	8 000	16 280	24 280
Ont.	10 714	37	18 500	23 760	42 260
Man.	1 114	11	5 500	2 500	8 000
Sask.	990	11	5 400	2 320	7 720
Alb.	2 697	17	8 150	6 050	14 200
C-B	3 725	20	10 000	8 090	18 090
Canada	29 000	133	65 830	64 920	130 750

Tableau 2 – Taille par catégorie de RSS

Catégorie de RSS	Taille de la population	# de RSS	Taille moyenne
Petite	Moins de 75 000	41	530
Moyenne	75 000 – 240 000	60	900
Grande	240 000 – 640 000	25	1 500
Très grande	640 000 et plus	7	2 500

La troisième et dernière étape de la démarche consiste à répartir l'échantillon provincial entre les RSS de la province proportionnellement à la racine de la taille de la population des régions, tout en respectant un minimum de 500 unités par RSS. Deux exceptions toutefois : 1 000 unités ont été attribuées à chaque RSS de l'Î-P-É et le minimum de 500 unités a été majoré à 800 pour la province de Québec. Cette démarche en trois étapes permet d'avoir pour chaque RSS un échantillon suffisant tout en ne perturbant que très légèrement la répartition provinciale des tailles d'échantillon. Le tableau 2 donne un aperçu sommaire de la répartition des RSS selon leur population, ainsi que de la taille moyenne de l'échantillon prévue par catégorie de RSS.

La taille finale visée est de 130 750 répondants et ce, malgré l'opportunité qu'ont eue les RSS et les autorités provinciales d'augmenter les tailles d'échantillon en achetant des unités supplémentaires; les tailles allouées ont été jugées satisfaisantes. Ces tailles ont été gonflées avant d'effectuer la collecte pour tenir compte des logements vacants et de la non-réponse.

3.4. Choix de la base de sondage

Il existe plusieurs bases de sondage à Statistique Canada qui sont utilisées par les différentes enquêtes auprès des ménages et pouvant être souhaitables pour une enquête sur la santé. D'une part, la base aéroilaire mise en place pour l'Enquête sur la population active (EPA) est certes une base de choix pour les enquêtes auprès des ménages. Avec cette base, il est possible de sélectionner soit un échantillon de logements nouveaux, soit un échantillon de logements échus de l'EPA. (Il est bon de noter que pour l'EPA, les logements sélectionnés demeurent dans l'échantillon pour une durée de six mois consécutifs et sont par la suite remplacés; ces logements deviennent alors disponibles pour d'autres enquêtes.) D'autre part, le registre des adresses (RA) de Statistique Canada, la base de composition aléatoire (CA) de numéros de téléphone et les fichiers administratifs provinciaux de

la santé sont également des bases potentielles.

Le choix d'une base de sondage pour tirer un échantillon dépend de plusieurs facteurs mais la base doit d'abord et avant tout correspondre le plus possible à la population cible de l'enquête. De plus, la création, l'utilisation, la mise à jour et la vérification de la base de sondage doivent respecter les contraintes opérationnelles et budgétaires de l'enquête. Pour cette enquête, il a été décidé d'utiliser deux bases de sondage chevauchantes: la base aréolaire mise en place pour l'EPA et la base CA de numéros de téléphone. Dans le cadre de l'ESCC, la base aréolaire est utilisée comme base principale alors que la base CA agit comme complément à la base principale dans certaines RSS.

Outre le fait que la population cible de l'ESCC soit la même que celle de l'EPA, les avantages reliés au choix de la base aréolaire mise en place pour l'EPA pour tirer l'échantillon de l'ESCC sont indéniables. L'infrastructure déjà en place pour effectuer les mises à jour en fonction des nouveaux logements, des logements démolis et des unités hors du champ d'observation de même que tout le processus d'évaluation de la couverture de la base sont des atouts certains. De plus, étant donné que plusieurs autres enquêtes auprès des ménages de Statistique Canada utilisent également cette base aréolaire, le contrôle des chevauchements d'échantillon entre les enquêtes est facilité. Pour plusieurs raisons, mais principalement pour une question de fardeau de réponse, il a été décidé de sélectionner un échantillon de logements nouveaux et non un échantillon de logements échus de l'EPA.

Les limites reliées à l'utilisation de la base CA sont bien connues: sous-couverture des ménages sans téléphone (~2%) ou avec téléphone cellulaire seulement (estimé entre 1% et 2% selon les secteurs), taux de réponse généralement plus faibles et la nécessité d'effectuer plusieurs appels avant de contacter un ménage valide. L'utilisation de deux bases de sondage chevauchantes est toutefois nécessaire pour les trois raisons suivantes: i) minimiser les coûts de collecte des données dans les régions où la disponibilité des intervieweurs est restreinte, ii) pallier l'incapacité de la base aréolaire de fournir l'échantillon nécessaire pour certaines RSS et iii) disposer en permanence d'une structure flexible de collecte de données par téléphone.

3.5. Stratégie d'échantillonnage

L'échantillon total visé par l'ESCC est de 130 750 répondants; la majorité (115 000 répondants) provient

de la base aréolaire où des interviews personnelles sont menées alors que le reste provient de la base CA où, bien sûr, des interviews téléphoniques sont réalisées.

3.5.1. Échantillonnage des ménages

Le plan de base de l'EPA est un échantillon stratifié à plusieurs degrés dans lequel le logement est l'unité d'échantillonnage finale (Statistique Canada, 1998b). Pour les secteurs géographiques sélectionnés au premier degré, une liste des logements est effectuée et mise à jour sur le terrain. Un échantillon de logements est par la suite sélectionné au deuxième degré à partir de chaque liste. Les ménages demeurant dans les logements sélectionnés constituent l'échantillon de ménages. Afin d'atteindre la taille visée de 97 000 ménages, quelques 123 000 logements doivent être sélectionnés à partir de la base aréolaire afin de tenir compte des logements vacants et de la non-réponse.

La stratégie d'échantillonnage des ménages par composition aléatoire (CA) de numéros de téléphone utilise la méthode d'élimination des banques inactives (MÉBI). Une banque (indicatif régional à trois chiffres plus les cinq premiers chiffres d'un numéro de téléphone qui en compte sept) est considérée comme active aux fins de l'échantillonnage si elle comprend au moins un numéro de téléphone. L'information est obtenue à partir des fichiers de facturation fournis par les compagnies de téléphone ainsi que d'autres fichiers administratifs publiques. Les banques actives sont regroupées pour former des strates CA. À l'intérieur d'une strate CA, une banque est choisie aléatoirement et un numéro entre 00 et 99 est généré de façon aléatoire pour former un numéro de téléphone complet à 10 chiffres (3+7). Ce processus est répété jusqu'à ce que le nombre de numéros de téléphone souhaités à l'intérieur de la strate CA soit atteint. Souvent, le numéro généré correspond à un numéro de téléphone qui est hors du champ de l'enquête ou à un numéro qui n'est pas en service. Afin d'atteindre les objectifs visés quant à la taille d'échantillon nécessaire, de nombreux numéros de téléphone additionnels doivent être générés. Dans le cadre de l'ESCC, les taux de succès varient de 15% à 61% parmi les RSS où un échantillon de ménages de la base CA est nécessaire. À l'échelle globale, un échantillon d'un peu plus de 51 000 numéros de téléphone est donc nécessaire (en tenant également compte des ménages non-répondants).

3.5.2. Échantillonnage des personnes

L'ESCC cible les personnes vivant dans des logements privés occupés qui sont âgées de 12 ans ou plus. Il est important de mentionner qu'une des recommandations

qui est ressortie du processus de consultation était la nécessité d'obtenir des informations très détaillées sur la santé des jeunes (12-19 ans) et des personnes âgées (65 ans ou plus). Afin de tenir compte de cette demande, il a été décidé d'augmenter la représentativité de ces sous-groupes d'intérêt.

Interviewer plus d'une personne dans un même ménage permet une économie des coûts de collecte; une grande partie de ces coûts étant en effet attribuable aux démarches pour rejoindre le ménage, interviewer plus d'une personne ne coûte guère plus. Par contre, les fortes similitudes observées chez les membres d'un même ménage peuvent occasionner un certain effet de grappe non souhaitable pour quelques caractéristiques importantes de l'enquête. De plus, le fardeau de réponse du ménage se trouve accru.

D'un autre côté, sélectionner une seule personne par ménage représente un accroissement significatif des coûts de collecte car un plus grand nombre de ménages doit être échantillonné. Et encore, comme les chances de faire partie de l'échantillon sont inversement proportionnelles au nombre de personnes à l'intérieur du ménage, certaines catégories de personnes (selon l'âge) sont soit sous, soit surreprésentées. Pour des raisons de coût, d'efficacité du plan, du fardeau de réponse ainsi que des contraintes opérationnelles, la règle de sélection des personnes mise en place pour cette enquête se veut un compromis.

Pour tous les ménages provenant de la base CA, une seule personne est choisie parmi tous les membres du ménage. Pour les ménages provenant de la base aréolaire, une seule personne est sélectionnée dans 79 000 ménages alors que deux personnes sont choisies dans 18 000 ménages. La règle de sélection des personnes pour les ménages de la base aréolaire est définie en fonction de la composition du ménage et elle a pour but d'augmenter la représentativité des 12 à 19 ans dans l'échantillon. Le tableau 3 décrit la règle de sélection.

Tableau 3 – Règle de sélection des personnes selon la composition du ménage – base aréolaire

Nombre de 12-19	Nombre de personnes âgées de 20+					
	0	1	2	3	4	5+
0	-	A	A	A	A	B
1	A	A	C	C	C	B
2	A	C	C	C	C	C
3+	A	C	C	C	C	C

A: sélection aléatoire d'une personne de 12+

B: sélection aléatoire de deux personnes 12+

C: sélection aléatoire d'une personne dans le groupe 12-19 ET d'une personne de 20+

Tableau 4 – Distribution anticipée de l'ESCC

Groupe d'âge	Recens. 1996	* ESCC (1 pers.)	* ESCC (1 et 2 pers.)
Col. 1	Col. 2	Col. 3	Col. 4
12-19	13.2	8.3	14.9
20-29	16.4	14.3	13.1
30-44	30.8	29.1	28.1
45-64	25.8	27.9	26.3
65+	13.8	20.2	17.6

- Moyenne des distributions sur 100 simulations

Le tableau 4 présente les distributions par groupe d'âge observées du Recensement de la population de 1996 (colonne 2), d'un échantillon simulé de l'ESCC où une seule personne par ménage est sélectionnée (colonne 3) ainsi que la distribution anticipée de l'échantillon de l'ESCC en sélectionnant deux personnes (selon la règle décrite au tableau 3) dans 18 000 ménages et une personne de façon aléatoire dans tous les autres ménages sélectionnés. Les résultats montrent que les deux sous-groupes d'intérêt (12 à 19, et 65 et plus) sont maintenant surreprésentés dans l'échantillon sans toutefois pénaliser outre mesure les autres groupes d'âge.

3.6. Conception du questionnaire

Une des recommandations importantes qui est ressortie du processus de consultation était de recueillir des données sur des préoccupations uniques aux RSS. Afin d'atteindre cet objectif, une stratégie innovatrice à l'ESCC a été mise en place.

Le développement du contenu final du questionnaire a été divisé en deux parties : un contenu commun d'une durée de 35 minutes et un contenu optionnel d'une durée de 10 minutes adapté aux besoins régionaux. Les provinces et les RSS ont eu à choisir parmi une liste de 28 modules, de durée variable, des sujets pertinents à leurs préoccupations régionales jusqu'à concurrence de 10 minutes. À la fin de ce processus, 27 versions différentes du questionnaire ont été identifiées. Pour une copie de toutes les questions du questionnaire de l'ESCC, le lecteur est référé à [http : www.statcan.ca/enquetes_sante](http://www.statcan.ca/enquetes_sante).

3.7. Pondération et estimation

Le plan de sondage adapté aux régions sociosanitaires combiné à la stratégie de collecte des données répartie sur 12 mois permettra la diffusion d'estimations préliminaires au courant de l'année. Les poids d'échantillonnage seront calculés après 3, 6, 9 et 12 mois de collecte. Étant donné l'utilisation de deux bases de sondage ayant chacune leur plan

d'échantillonnage, deux systèmes de pondération ayant chacun sa série d'ajustements particuliers (incluant un ajustement pour la non-réponse totale) seront mis en place. Les deux systèmes seront, par la suite, intégrés afin de ne pas doubler les estimations dans la population. Bien qu'une décision finale n'ait pas été prise, une technique d'intégration tenant compte des effets de plan et des tailles effectives d'échantillon de chaque plan sera vraisemblablement utilisée. Les poids intégrés seront finalement calibrés pour 10 groupes âge-sexe à l'intérieur de chaque région sociosanitaire.

4. PLANS FUTURS

La composante provinciale de l'ESCC, dont la collecte débutera en janvier 2002, aura comme objectif principal de produire des estimations transversales portant sur différents aspects reliés à la santé mentale et au bien-être des Canadiens à l'échelle provinciale. Cette enquête sera menée auprès d'un échantillon de 30 000 répondants répartis dans les dix provinces.

À la fin du présent cycle de l'ESCC et ce, même si l'octroi du budget final est encore à l'étude, il est prévu répéter le cycle biennal de l'ESCC en utilisant une approche similaire. Une composante régionale avec un vaste échantillon se tiendrait en 2003, et elle serait suivie d'une composante provinciale en 2004 ayant une thématique particulière portant sur la nutrition et les mesures physiques.

RÉFÉRENCES

Institut canadien de l'information sur la santé (1999). *Carnet de route de l'information sur la santé : Entreprendre le voyage*. (1-895581-32-X)

Statistique Canada (1998a). *Lignes directrices concernant la qualité*. Troisième Édition, octobre 1998, Statistique Canada. Cat. no. 12-539-X1F.

Statistique Canada (1998b). *Méthodologie de l'Enquête sur la population active du Canada*, Statistique Canada. Cat. no. 71-526-XPB.