

No 11-522-XIF au catalogue

**La série des symposiums internationaux  
de Statistique Canada - Recueil**

**Symposium 2006 : Enjeux  
méthodologiques reliés à la  
mesure de la santé des  
populations**



2006



Statistics  
Canada

Statistique  
Canada

Canada

## Plan d'échantillonnage de l'Enquête canadienne sur les mesures de la santé

Suzelle Giroux<sup>1</sup>

### Résumé

Dans la présente communication, le plan d'échantillonnage élaboré pour répondre aux objectifs et à la logistique de l'enquête est décrit. Parmi les défis à relever, notons la nécessité d'avoir les répondants sélectionnés à proximité des cliniques, la difficulté d'atteindre la taille d'échantillon désirée pour les jeunes ainsi que le sous-échantillonnage pour les mesures reliées à l'exposition aux agents environnementaux. Le plan d'échantillonnage présentera des solutions à ces défis entre autres, la formation de sites de collecte, l'utilisation de plusieurs bases de sondage ainsi qu'une stratégie de sélection des personnes.

MOTS CLÉS: enquête sur la santé, mesures directes, base aréolaire, site de collecte, échantillon à plusieurs degrés

### 1. Introduction

La toute nouvelle Enquête canadienne sur les mesures de la santé (ECMS) a comme objectifs d'estimer la prévalence et la distribution d'indicateurs de santé tels que l'embonpoint, l'obésité, l'activité physique, la nutrition, ainsi que l'exposition aux agents environnementaux polluants. À l'aide d'entrevues et de mesures directes, l'enquête aidera à estimer le nombre de Canadiens ayant certaines conditions chroniques et facteurs de risques. La dernière enquête similaire, l'Enquête santé Canada, a été effectuée en 1978 et 1979.

### 2. Particularités de l'enquête

#### 2.1 Collecte

La collecte devrait débuter en mars 2007 et s'effectuera sur une période de 2 ans. La collecte des données s'effectuera en deux étapes. Tout d'abord, des interviewers se présenteront au domicile des répondants et leur demanderont de répondre à des questions sur différents aspects de leur santé et ainsi que sur d'autres variables sociodémographiques. Ensuite, les répondants seront invités à prendre rendez-vous à la clinique de l'ECMS où des professionnels de la santé prendront des mesures physiques telles que la grandeur, le poids, la pression artérielle, l'évaluation de la condition physique, des échantillons de sang et d'urine pour analyse en laboratoire. Une fois les mesures terminées, un rapport de santé préliminaire leur sera remis et on leur demandera de porter un accéléromètre pendant une période d'une semaine et de le retourner dans une enveloppe préaffranchie. Les répondants seront remboursés pour les dépenses encourues pour le déplacement à la clinique.

#### 2.2 Logistique

La clinique mobile de l'ECMS consiste en deux grandes roulottes reliées entre elles par une passerelle. Cette clinique sera installée dans un site et sera opérationnelle pendant une période de 6 à 8 semaines, après quoi elle sera déplacée sur le prochain site. La capacité de la clinique pour chacune des périodes est évaluée à environ 350 répondants.

Pour s'assurer que le répondant puisse se déplacer à la clinique dans un délai et à l'intérieur d'une distance raisonnable, des sites de collecte ont été créés. La définition des sites est décrite dans la prochaine section.

---

<sup>1</sup>Suzelle Giroux, Statistique Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A-0T6 (sgiroux@statcan.ca)

### **3. Plan de sondage**

#### **3.1 Population cible et taille d'échantillon**

La population cible est définie par des personnes âgées de 6 à 79 ans au moment de l'enquête, vivant dans des logements privés. De cette population sont exclus les personnes demeurant sur les réserves indiennes et terres de la Couronne, les membres réguliers des forces canadiennes, les résidents d'institution ainsi que ceux des régions éloignées.

Les objectifs analytiques de l'enquête consistent à obtenir des estimations nationales fiables pour 10 groupes âge-sexe, c'est à dire, cinq groupes d'âge ( 6 à 11 ans, 12 à 19 ans, 20 à 39 ans, 40 à 59 ans et 60 à 79 ans) par sexe (homme et femme). Le nombre de répondants désirés à la clinique a été fixé à 5 000, réparti également dans les 10 groupes. Ce nombre permettra d'estimer une prévalence de 10% ou plus pour chacun des groupes, avec un coefficient de variation de 16,5%, basé sur un effet de plan de 1,5.

#### **3.2 Base de sondage**

Une base aréolaire a été utilisée pour créer les sites de collecte : celle de l'enquête sur la population active (EPA). Les grappes de l'EPA, petites unités géographiques contenant généralement environ 150 à 200 logements, ont été regroupées pour former des sites selon les critères suivants :

1. un minimum de 10 000 personnes dans le site;
2. les limites du site doivent respecter les frontières des provinces et des régions métropolitaines de recensement (RMR);
3. la distance maximum du centre du site à ses limites doit être de :
  - a. moins de 50 kilomètres, si le site est à l'intérieur d'une RMR
  - b. moins de 100 kilomètres, si le site est à l'extérieur d'une RMR.

En suivant ces critères, 257 sites de collecte ont ainsi été créés. Environ 3.7% de la population canadienne est exclue.

#### **3.3 Plan d'échantillonnage**

Un plan d'échantillonnage à plusieurs degrés est utilisé pour rejoindre la population cible. Tout d'abord, un échantillon de sites sera sélectionné. Ensuite, dans chacun des sites, des logements puis des personnes seront sélectionnés. La section suivante décrit la sélection des sites.

##### **3.3.1 Sélection des sites de collecte**

Le nombre de répondants désirés étant fixé à 5 000 et la capacité de la clinique étant estimée à 350 répondants par site, on obtient donc  $5\ 000 / 350 = 15$  sites de collecte qui devront être échantillonnés.

Étant donné ce nombre restreint d'unités primaires et afin d'assurer une bonne représentativité à travers le Canada, les 257 sites de collecte ont été stratifiés selon les cinq régions géographiques canadiennes. Les 15 sites de collecte ont été répartis proportionnellement selon la taille de la population cible à l'intérieur de chaque région. Le tableau 1 donne la répartition régionale des sites.

**Tableau 1 : Répartition de l'échantillon des sites de collecte**

Région	Population âgée de 6 à 79 ans, Recensement de 2001	Nombre de sites dans la population	Nombre de sites dans l'échantillon
Atlantique	2 061 425	36	1
Québec	6 560 375	50	4
Ontario	10 248 545	61	6
Prairies	4 538 970	77	2
Colombie-Britannique	3 540 000	33	2
<b>Total</b>	<b>26 949 315</b>	<b>257</b>	<b>15</b>

À l'intérieur de chacune des régions, les sites ont été ordonnés selon qu'ils appartiennent ou non à une région métropolitaine de recensement et selon la taille de la population du site. La sélection des sites a été effectuée en utilisant une méthode d'échantillonnage systématique proportionnelle à la taille de la population à l'intérieur de chacune des régions.

### 3.3.2 Sélection des logements et des personnes

Un des défis de cette enquête est de réussir à obtenir un nombre égal de répondants pour chaque groupe d'âge par sexe étant donné que la composition des ménages n'est pas toujours connue au moment de la sélection. Les répondants les plus difficiles à rejoindre sont les jeunes dans les groupes d'âge de 6 à 11 ans et de 12 à 19 ans.

Deux options sont envisagées pour la sélection des logements et des personnes pour l'ECMS.

#### 3.3.2.1 Option 1

Pour cette option, deux sources sont utilisées pour échantillonner les logements puis les personnes pour chacun des sites.

##### Sélection des logements

La première source est le cycle 3.1 de l'Enquête sur la santé dans les collectivités canadiennes (ESCC 3.1). Les logements situés dans les 15 sites de collecte sélectionnés et où des jeunes âgés de 6 à 11 ans ou de 12 à 19 ans sont présents dans les ménages sont choisis.

Cette source ne permettant pas d'obtenir suffisamment de répondants dans ces deux groupes d'âge, elle est complétée par un autre échantillon de logements provenant de la base de sondage l'EPA. L'infrastructure de l'EPA est utilisée pour choisir ces logements, c'est à dire, tout d'abord, des grappes sont choisies, ensuite, des logements sont choisis systématiquement à l'intérieur de ces grappes. Le nombre de logements choisis dans l'EPA est déterminé en fonction du nombre de logements provenant de l'ESCC 3.1 et le nombre requis de logements pour obtenir un nombre de répondants égal par groupe d'âge et sexe. Étant donné que la composition des ménages des logements provenant de l'EPA est inconnue a priori, on doit échantillonner plus de logements que nécessaire et rejeter des ménages au moment de la collecte pour les groupes d'âge où la population est plus fréquente par rapport aux autres groupes d'âge plus difficiles à cibler.

### Sélection des personnes

Lors de l'entrevue au domicile du répondant, on demande de fournir de l'information sur les membres du ménage (incluant l'âge et le sexe). Pour la sélection des personnes pour l'enquête, des vecteurs de sélection avec probabilités variables par groupe d'âge sont utilisés.

Dans le cas des logements sélectionnés à partir de l'ESCC 3.1, une seule personne sera sélectionnée par ménage, soit un jeune de 6 à 11 ans ou un jeune de 12 à 19 ans. Si aucun jeune n'est présent, aucune personne ne sera sélectionnée et le ménage sera rejeté.

Dans le cas de logements sélectionnés à partir des grappes de l'EPA, deux personnes par ménage seront sélectionnées si un jeune de 6 à 11 ans est présent, c'est à dire, un jeune de 6 à 11 ans et une autre personne parmi les autres groupes d'âge (12 à 79 ans). Si aucun jeune de 6 à 11 ans n'est présent, alors une seule personne sera sélectionnée parmi les personnes âgées de 12 à 79 ans. Il est à noter qu'une méthode de rejet aléatoire est nécessaire pour rejeter les ménages avec personnes âgées de 20 à 59 ans, puisque ces groupes d'âge sont trop fréquents par rapport aux autres groupes d'âge et il est nécessaire de rejeter ces ménages afin d'obtenir une taille d'échantillon égale par groupe d'âge-sexe.

### Avantages de cette option

Il est possible de prendre avantage de l'infrastructure élaborée pour l'EPA tel que l'énumération des logements dans certaines grappes et ainsi réduire les coûts. L'énumération récente des grappes permet également d'inclure toute nouvelle construction (nouveau logement), donc une meilleure couverture. Finalement, l'infrastructure de l'EPA permet d'éviter le chevauchement d'enquête, donc, réduction du fardeau des répondants.

### Désavantages de cette option

L'utilisation de deux sources pour la sélection de l'échantillon (ESCC 3.1 et EPA) de même que l'ajout d'un degré d'échantillonnage (grappe) avant la sélection des logements provenant de l'EPA diminue l'efficacité de l'échantillon. Il y a un fardeau supplémentaire pour les ménages et personnes sélectionnés à partir de l'ESCC 3.1. De plus, le nombre restreint de ménages avec jeunes provenant de l'ESCC 3.1 résulte en un échantillon plus grand provenant de l'EPA, entraînant un fardeau et un coût plus élevé pour les ménages contactés qui sont ensuite rejetés.

## **3.3.2.2 Option 2**

Pour cette option, on utilise le recensement comme base de sondage. La liste des adresses des logements identifiés lors du recensement de 2006 est obtenue, et ce, pour chacun des 15 sites. L'information sur l'âge et le sexe des personnes présentes dans ces logements en date du recensement peut être utilisée.

### Sélection des logements

Pour chacun des sites, les logements sont stratifiés selon l'âge des membres du ménage. L'âge est déterminé en utilisant la date de naissance des membres du ménage et la date du début de la collecte pour un site donné. La stratification prend en considération les groupes d'âge les plus difficiles à rejoindre et est définie comme suit :

- strate 1 : au moins une personne âgée de 6 à 11 ans est présente dans le logement; sinon,
- strate 2 : au moins une personne âgée de 12 à 19 ans est présente dans le logement; sinon,
- strate 3 : au moins une personne âgée de 60 à 79 ans est présente dans le logement; sinon,
- strate 4 : au moins une personne âgée de 20 à 39 ans est présente dans le logement; sinon,
- strate 5 : au moins une personne âgée de 40 à 59 ans est présente dans le logement; sinon,
- strate 6 : logements ne faisant pas partie de la population cible selon la composition du ménage ou logements vacants en date du recensement.

Dans chacune des strates, un échantillon aléatoire de logements sera ensuite sélectionné.

### Sélection des personnes

Tout comme dans l'option 1, lors de l'entrevue au domicile du répondant, on demande de fournir de l'information sur les membres actuels du ménage (incluant l'âge et le sexe). Il est donc possible que les personnes présentes au moment du recensement ne soient plus les mêmes au moment de la collecte.

Deux personnes par ménage seront sélectionnées si un jeune de 6 à 11 ans est présent, c'est à dire, un jeune de 6 à 11 ans et une autre personne parmi les autres groupes d'âge (12 à 79 ans). Si aucun jeune de 6 à 11 ans n'est présent, alors une seule personne sera sélectionnée parmi les personnes âgées de 12 à 79 ans. Des vecteurs de sélection avec probabilités variables par groupe d'âge sont utilisés et diffèrent selon la strate.

### Avantages de cette option

La composition des ménages à l'intérieur des logements en date du recensement est connue, permettant de stratifier et ainsi mieux cibler les groupes d'âge désirés. Le rejet des ménages ne sera pas nécessaire, du moins au début, étant donné que l'on est près de la date du recensement. L'utilisation d'une seule source pour échantillonner les personnes et l'élimination d'un degré d'échantillonnage (grappes) par rapport à l'option 1 permet d'avoir un plan de sondage plus efficace.

### Désavantages de cette option

La capacité des strates à rejoindre les groupes d'âge désirés diminuera à mesure qu'on s'éloigne de la date du recensement. On devra trouver une autre source pour ajouter les nouvelles constructions ou les logements manqués lors du recensement.

### Derniers développements

Au moment de rédiger cet article, l'option 2 a été retenue pour sélectionner les logements et les personnes. Les données du recensement ont pu être obtenues et analysées pour développer et appliquer le plan d'échantillonnage à temps pour le début de l'enquête.

## **3.4 Autres considérations**

### **3.4.1 Contaminants environnementaux**

L'analyse de certains contaminants environnementaux étant trop coûteuse pour être effectuée sur l'ensemble des 5 000 répondants, des sous-échantillons seront identifiés parmi l'échantillon original. Les personnes sélectionnées auront des analyses supplémentaires effectuées sur leur échantillon de sang ou d'urine. Les sous-échantillons sont indépendants, varient par groupe d'âge, et, dans la mesure du possible, ne doivent s'effectuer que sur une seule personne du ménage, dans le cas où un 6 à 11 ans est présent. Afin de sous-échantillonner ces personnes, des intervalles de sélection ont été définis par groupe d'âge et par contaminant. Un nombre aléatoire par ménage est généré pour chacun des contaminants. Si le nombre aléatoire est inclus dans l'intervalle de sélection défini pour le contaminant et le groupe d'âge correspondant au répondant du ménage, ce répondant est sous-échantillonné et une analyse sera effectuée pour mesurer la présence du contaminant.

Dans le cas où deux personnes par ménage sont sélectionnées pour l'enquête, une méthode de collocation est utilisée pour le sous-échantillonnage afin d'éviter que les deux personnes du ménage soient sous-échantillonnées pour la même mesure. Les 6 à 11 ans auront un intervalle de sélection défini à une extrémité de l'intervalle (0 à p) alors que les autres groupes d'âge auront un intervalle de sélection défini à l'autre extrémité (de 1-p à 1).

### **3.4.2 Rendez-vous le matin ou l'après-midi et soir**

Les personnes sélectionnées pour l'enquête seront assignées aléatoirement à un rendez-vous le matin ou bien un rendez-vous l'après-midi et soir. Certaines analyses demandent que le répondant soit à jeun pendant 12 heures et on demandera aux répondants du matin seulement de se conformer à ces instructions, puisqu'il serait difficile pour le répondant (fardeau) de s'y conformer si le rendez-vous est l'après-midi ou le soir. Cette répartition aléatoire a pour but d'éviter qu'il y ait un biais relié au choix du rendez-vous par le répondant.

### **3.4.3 Répétition de mesures physiques**

Pour s'assurer de la qualité de certaines mesures physiques prises à la clinique, certains répondants seront sélectionnés pour un des trois groupes de mesures qui seront répétées à la fin de la visite. Les mesures faisant partie de ce contrôle sont : les mesures anthropométriques (taille, taille en position assise, poids, tour de taille, plis cutanés), la force de préhension, ainsi que la flexion du tronc vers l'avant en position assise. Les mesures répétées seront comparées aux mesures originales pour évaluer les erreurs de mesure.

### **3.4.4 Ordre de collecte des sites**

La collecte des données sera effectuée un seul site à la fois, un après l'autre. La collecte des 15 sites s'effectuera donc sur une période de deux ans. L'ordre de collecte des sites considère à la fois la répartition temporelle, c'est à dire, que les sites sont répartis également entre la première et la deuxième année selon la région (sauf la région de l'Atlantique où il n'y a qu'un site), ainsi que la saisonnalité dans la mesure du possible, tout en tenant compte des contraintes opérationnelles.

## **4. Prochaines étapes**

Dans les prochaines semaines, c'est à dire à l'hiver 2007, un test technique des différentes composantes de l'enquête sera effectué, suivi d'une répétition générale de l'enquête avec des volontaires. Le début de la collecte pour le premier site est prévu pour mars 2007.

## **Remerciements**

L'auteur aimerait remercier Sarah Maude Dion, Nicolas Lavigne et Rebecca Morrison pour leur participation au développement du plan de sondage de l'ECMS.

## **Références**

- Giroux S. (2004), "Study on the number of sites for the Canadian Health Measures Survey", unpublished report, Statistics Canada
- Giroux S., Lavigne N. (2005), "Sampling strategy for the Canadian Health Measures Survey, Part 1: selection of collection site", unpublished report, Statistics Canada
- Laflamme G., Dochitoui (2005), "Creating sites for the Canadian Health Measures Survey", unpublished report, Statistics Canada
- Lavigne N. (2006), "Stratégie de sélection des personnes", unpublished report, Statistics Canada
- Statistics Canada (1998), "Methodology of the Labour Force Survey", Catalogue 71-256-XPB, Ottawa : Statistics Canada

Tambay J.L., Mohl C. (1995), "Improving sample representativity through the use of a rejective method", 1995 ASA proceedings of the Survey Methods Section