



Composante du produit n° 82-003-S  
Rapports sur la santé

## Article

# Enquête canadienne sur les mesures de la santé : aperçu de la stratégie d'échantillonnage

par Suzelle Giroux

5 décembre 2007



Canada

# Enquête canadienne sur les mesures de la santé : aperçu de la stratégie d'échantillonnage

Suzelle Giroux

## Résumé

Le présent article donne un aperçu de la stratégie d'échantillonnage élaborée pour l'Enquête canadienne sur les mesures de la santé (ECMS). Cette enquête est conçue en vue de recueillir des renseignements clés sur la santé grâce à une entrevue sur place assistée par ordinateur et un examen de santé physique. Elle a débuté en mars 2007 et se poursuivra jusqu'en 2009. Environ 5 000 personnes de 6 à 79 constituant un échantillon représentatif de la population nationale seront interviewées à leur domicile, puis invitées à passer une visite dans une clinique mobile où des professionnels de la santé mesureront plusieurs aspects de leur santé physique. L'ECMS pose plusieurs défis, dont la nécessité de sélectionner des participants qui vivent à une distance raisonnable des cliniques, la difficulté d'obtenir la taille d'échantillon souhaitée pour les jeunes et le sous-échantillonnage pour les mesures de l'exposition aux contaminants présents dans l'environnement. La stratégie d'échantillonnage décrite dans le présent article permettra de relever ces défis.

## Mots-clés

base aéroloire, emplacements de collecte, études transversales, mesures directes, enquêtes sur la santé, échantillonnage à plusieurs degrés, méthodes de rejet, sous-échantillonnage, vecteur de probabilités de sélection.

## Auteur

Suzelle Giroux (613-951-1481; Suzelle.Giroux@statcan.ca) travaille à la Division des méthodes d'enquête auprès des ménages, à Statistique Canada, Ottawa (Ontario) K1A 0T6.

Au moyen d'entrevues à domicile et d'exams de santé physique, l'Enquête canadienne sur les mesures de la santé (ECMS) permettra de recueillir des renseignements sur l'état de santé général et sur le mode de vie de la population canadienne. Les résultats de cette enquête à participation volontaire permettront d'estimer la prévalence de maladies et de facteurs de risque particuliers.

Le questionnaire de l'ECMS est conçu pour recueillir des renseignements sur les antécédents médicaux, le régime alimentaire, les habitudes d'usage du tabac, la consommation d'alcool, l'état de santé courant et le comportement sexuel, ainsi que sur les caractéristiques démographiques et socioéconomiques. En outre, des renseignements seront recueillis sous forme de mesures physiques directes, comme la taille, le poids et la pression artérielle, le prélèvement d'échantillons de sang et d'urine et l'évaluation de la forme physique. Les résultats serviront de données de référence nationales sur la prévalence de problèmes de santé qui suscitent des inquiétudes, comme l'obésité, l'hypertension,

la maladie cardiovasculaire, ainsi que l'exposition aux maladies infectieuses et aux contaminants présents dans l'environnement. L'ECMS est l'une des rares enquêtes conçue par Statistique Canada en vue de recueillir des mesures directes.

Le présent aperçu décrit la stratégie d'échantillonnage suivie afin de satisfaire aux exigences de collecte de données et d'estimation de l'ECMS. D'autres articles de la publication décrivent le contexte de l'enquête<sup>1</sup>, les questions éthiques, juridiques et sociales soulevées par l'ECMS<sup>2</sup>, les défis logistiques et opérationnels<sup>3</sup>, ainsi que les résultats du prétest<sup>4</sup>.

### **Plan de sondage**

La collecte des données de l'ECMS durera deux ans et comporte deux étapes. En premier lieu, un intervieweur se rend au domicile des participant à l'enquête pour recueillir leurs réponses au questionnaire sur le ménage. En deuxième lieu, on demande aux personnes interrogées de prendre rendez-vous avec une clinique mobile de l'ECMS installée à proximité, afin que des professionnels de la santé puissent faire certaines mesures de santé physique. On leur demande aussi de porter un moniteur d'activité physique (accéléromètre) pendant une semaine pour mesurer leur niveau d'activité.

La clinique mobile de l'ECMS est installée à chaque emplacement pendant une période de six à sept semaines. Lors de la détermination des emplacements de collecte, il convient de tenir compte de la localisation des logements des participants à l'enquête afin de s'assurer qu'ils se situent à une distance raisonnable de la clinique<sup>3</sup>.

L'ECMS a pour champ d'observation les personnes de 6 à 79 ans qui vivent dans des logements privés. Les habitants des réserves indiennes ou des terres de la Couronne, les personnes vivant en établissement, les membres à temps plein des Forces canadiennes et les habitants de certaines régions éloignées sont exclus.

### **Taille de l'échantillon**

Afin de pouvoir produire des estimations nationales de référence pour divers indicateurs de la santé,

l'ECMS doit être réalisée auprès de 5 000 personnes réparties également selon le groupe d'âge (de 6 à 11 ans, de 12 à 19 ans, de 20 à 39 ans, de 40 à 59 ans et de 60 à 79 ans) et le sexe, ce qui donne 10 groupes en tout. Avant que débutent les opérations sur le terrain, on augmente ce nombre afin de tenir compte des logements hors du champ d'observation et des non-répondants. Cette taille d'échantillon devrait produire, pour des problèmes de santé dont la prévalence est égale ou supérieure à 10 %, des estimations de niveau national selon le sexe pour chacun des 5 groupes d'âge dont le coefficient de variation est de 16,5 %.

Des sous-échantillons de personnes participant à l'enquête sont également nécessaires pour les analyses d'échantillons de sang et d'urine afin de mesurer l'exposition aux contaminants présents dans l'environnement.

### **Base de sondage – Création, répartition de l'échantillon et choix des emplacements de collecte**

Comme les participants à l'ECMS doivent se présenter à une clinique, ils devraient pouvoir s'y rendre en un temps raisonnable. La base de sondage de l'Enquête sur la population active (EPA) a été utilisée pour concevoir les emplacements de collecte et contrôler leur taille afin de répondre à ces exigences.

À partir des grappes de la base de sondage de l'EPA<sup>5</sup>, on a créé 257 emplacements de collecte<sup>6</sup>. Les grappes sont de petites unités géographiques qui contiennent environ 200 logements. Un emplacement de collecte est une région géographique comptant au moins 10 000 habitants et est définie telle que la distance que doivent parcourir les participants à l'enquête n'excède pas 100 kilomètres (km) (50 km dans les régions urbaines et 100 km dans les régions rurales). Les régions ne répondant pas à ces critères ont été exclues. Les emplacements couvrent 96,3 % de la population canadienne<sup>7</sup>.

Bien que des estimations nationales seulement soient nécessaires, les emplacements de collecte ont été stratifiés en cinq régions pour s'assurer que la répartition de l'échantillon soit représentative de la

Tableau 1  
Sélection des emplacements de collecte de l'Enquête canadienne sur les mesures de la santé, par région

Région	Population cible estimée, de 6 à 79 ans, Recensement de 2001	Emplacements dans la région	Emplacements répartis	Échantillon
<b>Total</b>	<b>26 949 315</b>	<b>257</b>	<b>15</b>	<b>4 995</b>
Atlantique (Terre-Neuve-et-Labrador, Île-du-Prince-Édouard, Nouvelle-Écosse, Nouveau-Brunswick)	2 061 425	36	1	333
Québec	6 560 375	50	4	1 332
Ontario	10 248 545	61	6	1 998
Prairies (Alberta, Saskatchewan, Manitoba, y compris Yellowknife)	4 538 970	77	2	666
Colombie-Britannique (y compris Whitehorse)	3 540 000	33	2	666

population nationale. On s'est fondé pour cela sur les limites régionales types de Statistique Canada, à savoir la Colombie-Britannique, les Prairies (Alberta, Manitoba et Saskatchewan), l'Ontario, le Québec et les provinces atlantiques (Terre-Neuve-et-Labrador, Île-du-Prince-Édouard, Nouvelle-Écosse et Nouveau-Brunswick).

Un grand nombre d'emplacements de collecte comptant un petit nombre de participants est recommandé pour optimiser la précision des estimations<sup>8</sup>. Toutefois, compte tenu des contraintes logistiques et budgétaires inhérentes à l'utilisation de cliniques mobiles, le nombre d'emplacements de collecte a été limité à 15. À chaque emplacement, des données seront recueillies auprès d'environ 350 personnes, ce qui donnera un échantillon total d'environ 5 000 personnes. Les 15 emplacements de collecte ont été répartis entre les régions proportionnellement à la taille de la population (tableau 1).

Dans chaque région, les emplacements de collecte ont été triés en fonction de leur appartenance à une région métropolitaine de recensement (RMR), puis en fonction de la taille de la population avant la sélection. Une RMR est une région constituée d'une ou de plusieurs municipalités contiguës entourant une grande région urbaine (appelée noyau urbain). Le noyau urbain doit compter au moins 100 000 habitants pour former une RMR. Les emplacements de collecte ont ensuite été échantillonnés systématiquement avec une probabilité de sélection proportionnelle à la taille de leur population. Cette méthode de sélection, conjuguée au tri des emplacements selon l'appartenance ou non à une RMR et selon la taille

de la population, a permis de s'assurer que les emplacements sélectionnés soient répartis entre les RMR et les non-RMR, ainsi qu'entre les régions très peuplées et peu peuplées.

Bien qu'il n'y en aura pas dans toutes les provinces et tous les territoires, les emplacements de collecte des données de l'ECMS ont été choisis de manière qu'ils soient représentatifs de la population canadienne, d'est en ouest, compte tenu des grandes et faibles densités de population (voir *Emplacements de collecte de l'Enquête canadienne sur les mesures de la santé*).

### Emplacements de collecte de l'Enquête canadienne sur les mesures de la santé

#### Atlantique

- Région de Moncton, Nouveau-Brunswick

#### Québec

- Région de la ville de Québec
- Région de Montréal (deux emplacements)
- Région du sud de la Mauricie

#### Ontario

- Région d'Oshawa
- Région de Toronto (deux emplacements)
- Région de St. Catharine's-Niagara
- Région de Kitchener-Waterloo
- Région du comté de Northumberland

#### Prairies

- Région d'Edmonton, Alberta
- Région de Red Deer, Alberta

#### Colombie-Britannique

- Région de Vancouver
- Région de Williams Lake et Quesnel

## Échantillonnage des logements

Plusieurs options ont été étudiées afin de décider du meilleur moyen d'obtenir le nombre requis de participants par groupe d'âge. L'une d'elles consistait à utiliser les logements pour lesquels la composition du ménage était connue d'après l'Enquête sur la santé dans les collectivités canadiennes de 2005 (cycle 3.1 de l'ESCC) et d'y ajouter un nouvel échantillon de logements à composition du ménage inconnue tiré de la base de sondage de l'Enquête sur la population active<sup>9</sup>. Cette approche permettait d'atteindre le nombre cible de participants par groupe d'âge, mais l'adoption d'une méthode de rejet aléatoire<sup>10</sup> des ménages était nécessaire afin d'éviter d'interviewer un trop grand nombre de personnes appartenant au groupe d'âge représentant une part relativement importante de la population (de 20 à 59 ans) avant d'atteindre le nombre cible de participants pour les groupes d'âge plus avancés.

Une autre option consistait à prendre le Recensement de 2006 comme base de sondage. La composition des ménages occupant les logements en mai 2006 était connue et pouvait être utilisée pour établir un plan de sondage satisfaisant plus efficacement les exigences d'échantillonnage dans chaque groupe d'âge. Cette option est celle qui a été choisie.

Dans chaque emplacement de collecte, les logements dont on connaît la composition du ménage au moment du Recensement de 2006 sont stratifiés selon l'âge des occupants au moment de l'enquête. Cinq strates selon le groupe d'âge, qui correspondent aux cinq groupes d'âge de l'ECMS (de 6 à 11 ans, de 12 à 19 ans, de 20 à 39 ans, de 40 à 59 ans et de 60 à 79 ans), ont été créées. L'âge a été déterminé en se basant sur la date du début de la collecte des données à chaque emplacement.

- Strate 1 : logements où au moins un enfant de 6 à 11 ans est présent, sinon,
- Strate 2 : logements où au moins un jeune de 12 à 19 ans est présent, sinon,
- Strate 3 : logements où au moins une personne de 60 à 79 ans est présente, sinon,
- Strate 4 : logements où au moins une personne de 20 à 39 ans est présente, sinon,

- Strate 5 : logements où au moins une personne de 40 à 59 ans est présente, sinon,
- Strate contenant les logements non compris dans les strates susmentionnées, comme les logements inoccupés au moment du Recensement ou ceux habités par des personnes en dehors de la tranche d'âge visée par l'ECMS d'après la composition du ménage au moment du Recensement.

Chaque strate a une forte probabilité de contenir des logements habités par des personnes appartenant aux groupes d'âge souhaités. À chaque emplacement, un échantillon aléatoire simple de logements sera sélectionné dans chaque strate. La taille d'échantillon est répartie dans chaque strate de sorte que, en combinaison avec la stratégie d'échantillonnage des participants à l'enquête, un nombre égal de répondants par groupe d'âge puisse être obtenu. En principe, aucun rejet aléatoire de logements ne devrait avoir lieu au début de l'enquête. En revanche, vers la fin de la période de collecte de deux ans, la stratification basée sur le Recensement de 2006 risque d'être une stratégie moins efficace en vue d'atteindre certains groupes d'âge et le rejet aléatoire des ménages constitués principalement de personnes appartenant au groupe de 20 à 59 ans pourrait être nécessaire.

On prend contact avec chaque logement sélectionné afin de dresser la liste courante des membres du ménage que l'on utilise ensuite pour sélectionner les participants à l'enquête.

Comme la collecte des données de l'ECMS se poursuivra jusqu'en 2009, d'autres sources pourraient être utilisées pour compléter les adresses fournies au moment du Recensement de 2006 afin de tenir compte des nouveaux logements ou des logements manqués lors du recensement et de réduire le sous-dénombrement. Des sources complémentaires possibles sont le Registre des adresses et la base de sondage de l'Enquête sur la population active<sup>5</sup>.

## Échantillonnage des participants à l'enquête

Dans chaque strate, des probabilités de sélection différentes selon le groupe d'âge sont appliquées

afin d'être certains d'atteindre les cibles d'échantillonnage. On prend contact avec les logements sélectionnés afin d'obtenir la liste courante des membres des ménages. Dans chaque logement, une ou deux personnes sont sélectionnées selon la composition du ménage. Étant donné que les enfants doivent être accompagnés à la clinique, deux personnes sont sélectionnées dans les ménages comptant des enfants de 6 à 11 ans, à savoir un enfant choisi au hasard parmi les membres âgés de 6 à 11 ans et une deuxième personne choisie parmi celles âgées de 12 à 79 ans. Si aucun enfant de 6 à 11 ans ne vit dans le ménage, une seule personne est sélectionnée parmi les membres de 12 à 79 ans.

Le vecteur des poids pour la sélection des personnes de 12 à 79 ans a été construit de façon à éviter des poids d'échantillonnage individuels élevés. Puisque le poids pour certains groupes d'âge est deux fois plus grand que pour d'autres, il est possible que le poids d'échantillonnage d'une personne sélectionnée soit très élevé si les membres du ménage sont nombreux dans un logement. Par conséquent, lorsqu'un nombre minimum spécifié de personnes de 12 à 79 ans vivent dans un ménage, le poids appliqué à chacune est ramené à 1. Le cas échéant, tous les membres du ménage ont une même chance d'être sélectionnés.

Diverses études et simulations ont permis d'établir un juste équilibre entre les paramètres requis pour chacune des mesures mises en place. En principe, le nombre cible de participants ( $n=1\ 000$ ) par groupe d'âge sera atteint.

### **Sous-échantillonnage**

Des sous-échantillons de participants à l'enquête sont également sélectionnés pour le prélèvement d'échantillons de sang et d'urine afin de mesurer l'exposition aux contaminants présents dans l'environnement<sup>3</sup>. On ne se sert pas de l'échantillon complet de la population pour ces analyses à cause de leur coût élevé.

Les sous-échantillons sont sélectionnés de manière indépendante, c'est-à-dire sans tenir compte des sujets qui ont été sélectionnés pour le dépistage des autres groupes de substances. Donc, un

répondant pourra être sélectionné pour mesurer l'exposition à un, deux ou tous les contaminants.

Toutefois, deux personnes appartenant au même ménage ne devraient pas être sélectionnées pour le dépistage de la même substance. On pense en effet que les substances chimiques environnementales affectent les membres d'un même ménage de façon comparable. Afin d'éviter d'obtenir des renseignements plus ou moins identiques, il faut donc veiller à ne pas sélectionner deux membres d'un même ménage pour l'analyse de la même substance. Deux personnes pourraient être sélectionnées dans un ménage lorsque des sous-échantillons sont nécessaires pour les enfants de 6 à 11 ans ainsi que pour d'autres groupes d'âge. Une méthode d'échantillonnage « colloqué » (*collocated sampling*) a été utilisée afin de réduire au minimum ou de prévenir cette situation, en utilisant des intervalles de sélection différents pour le groupe des 6 à 11 ans et pour les autres groupes d'âge.

### **Autres facteurs considérés**

Chaque logement échantillonné est désigné aléatoirement pour l'attribution d'un rendez-vous à la clinique le matin ou l'après-midi au membre du ménage interrogé. Un rendez-vous le matin oblige la personne sélectionnée à jeûner toute la nuit, tandis que des restrictions alimentaires de moins longue durée sont imposées à celles dont le rendez-vous est fixé l'après-midi. Cette répartition aléatoire réduit le risque de biais qui pourrait exister s'il était permis aux participants à l'enquête de choisir le moment de leur rendez-vous. Les femmes enceintes, les diabétiques et d'autres personnes dans des situations particulières n'ont pas à jeûner, même si leur rendez-vous a lieu le matin.

Dans le cadre du programme de contrôle et d'assurance de la qualité, des logements sont sélectionnés aléatoirement afin de surveiller la qualité de certaines mesures physiques faites sur les répondants à la clinique. Pour les cas sélectionnés, les mesures physiques visées sont répétées sur la même personne et comparées aux premières faites.

Durant la période de collecte de deux ans, Statistique Canada veillera à ce que soient

échantillonnés les nombres requis de personnes pour chaque groupe âge-sexe. Par conséquent, des rajustements seront apportés au besoin aux paramètres du fichier de l'échantillon et au nombre de logements afin d'atteindre le nombre prévu de participants à l'enquête et de réduire le biais éventuel de non-réponse.

La collecte des données aux 15 emplacements se fera séquentiellement au cours des deux années. Les emplacements ont été classés de façon à tenir compte de la saisonnalité selon la région et de l'effet temporel, sous réserve des contraintes opérationnelles et logistiques. Par effet temporel, on entend que le nombre d'emplacements par région est réparti uniformément entre la première et la deuxième année (à l'exception de la région atlantique où il n'y aura qu'un seul emplacement).

### Mot de la fin

Le plan d'échantillonnage de l'Enquête canadienne sur les mesures de la santé (ECMS) est unique. Pour

la première fois depuis des décennies, des données sur le ménage ainsi que des mesures physiques seront recueillies auprès d'un échantillon représentatif au niveau national. L'ECMS s'inscrit dans la lancée des enquêtes sur la santé antérieures et permettra de combler les lacunes statistiques afin de répondre aux nouveaux besoins.

### Remerciements

L'auteure remercie Sarah Maude Dion, Nicolas Lavigne et Rebecca Morrison, de la Division des méthodes d'enquête auprès des ménages, pour leur participation à l'élaboration du plan d'échantillonnage de l'Enquête canadienne sur les mesures de la santé. Elle remercie également Johane Dufour et Yves Béland, de la Division des méthodes d'enquête auprès des ménages, ainsi que Sarah Connor Gorber, de la Division des mesures physiques de la santé, de leurs commentaires au sujet du présent article.

### Références

- 1 Tremblay MS, Wolfson M. Canadian Health Measures Survey: Background and rationale. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): XX-XX.
- 2 Day B, Langlois R, Tremblay M, et al. Canadian Health Measures Survey: Ethical, legal and social issues. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): XX-XX.
- 3 Bryan SN, St-Denis M, Wojtas D. Canadian Health Measures Survey: Operations and logistics of survey collection. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): XX-XX.
- 4 Tremblay MS, Langlois R, Bryan SN, et al. Canadian Health Measures Survey pre-test: Design, methods, results. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): XX-XX.
- 5 Statistique Canada. *Méthodologie de l'enquête sur la population active du Canada*, (n° 71-256-XPB au catalogue) Ottawa: Ministère de l'industrie, 1998.
- 6 Laflamme G, Dochitoui C. *Creating Sites for the Canadian Health Measures Survey*, April 2005, internal paper.
- 7 Giroux S, Lavigne N. *Sampling Strategy for the Canadian Health Measures Survey, Part I: Selection of Collection Sites*, HSMD, March 2005, Internal paper.
- 8 Giroux S. *Study on the Number of Sites for the Canadian Health Measures Survey*, March 26, 2004, internal paper.
- 9 Lavigne N. *Stratégie de sélection de personnes*, DMEM, January 2006, rapport non-publié.
- 10 Tambay J-L, Mohl C. Improving sample representativity through the use of a rejective method. *Proceedings of the Annual Meeting of the American Statistical Association: Survey Research Methods Section, August 1995*. Orlando, Florida: American Statistical Association, 1995.