

## **Bienvenue**

Le comité organisateur du 23<sup>e</sup> Symposium international sur les questions de méthodologie de Statistique Canada, «**Enjeux méthodologiques reliés à la mesure de la santé des populations**» vous souhaite la bienvenue. Nous espérons que vous trouverez le programme du Symposium intéressant, pertinent, stimulant, et utile, et que vous apprécierez rencontrer et discuter du sujet avec vos collègues pendant les trois prochains jours.

## **Le comité organisateur du Symposium 2006**

Milorad Kovacevic, Président

Colin Babyak

Marie P. Beaudet

Yves Béland

Chris Mohl

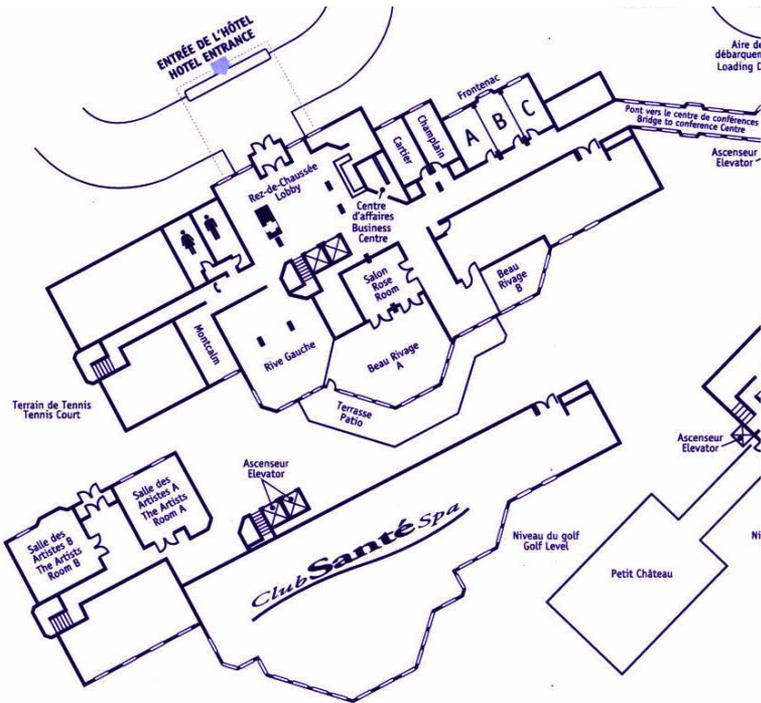
## Table des matières:

Cartes .....	4
Horaire .....	6
Programme .....	7
Résumés – Ateliers .....	20
Résumés – Sessions .....	22
Résumés – Affiches .....	59
Mot de la fin.....	71



# Château Cartier

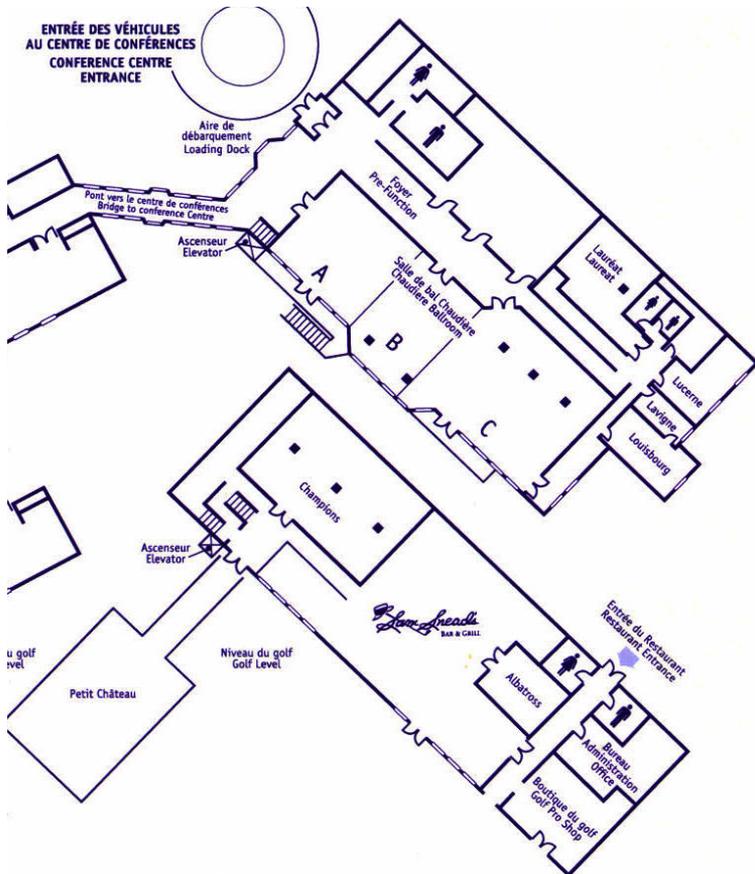
RELAIS · RESORT





# Château Cartier

RELAIS · RESORT



## HORAIRE

### Jour 1 – mercredi le 1er novembre

<b>8h à 9h</b>	Inscription aux ateliers, Rafrâichissements (Foyer de Chaudière ABC)	
<b>9h à 16h30</b> Ateliers concomitants	<b>Atelier 1</b> - Couplage d'enregistrements dans le cadre d'études sur la santé de la population — une vue d'ensemble (Chaudière B)	<b>Atelier 2</b> - Méthodes pour l'analyse des données d'enquêtes longitudinales sur la santé. Théorie et applications (Chaudière A)

### Jour 2 – jeudi le 2 novembre

<b>7h45 à 8h45</b>	Inscription, Rafrâichissements (Foyer de Chaudière ABC)	
<b>8h45 à 9h</b>	<b>Mot d'ouverture</b> (Chaudière ABC)	
<b>9h à 10h</b> Session plénière	<b>Session 1</b> : Discours Principal – Graham Kalton (Chaudière ABC)	
<b>10h à 10h30</b>	<b>Session d'affiches A</b> (Foyer de Chaudière ABC) Pause-café	
<b>10h30 à 12h</b> Sessions concomitantes	<b>Session 2</b> : Enjeux actuels relatifs à l'estimation pour petits domaines avec applications pour les enquêtes sur la santé (Chaudière AB)	<b>Session 3</b> : Analyse des données sur la santé – applications (Chaudière C)
<b>12h à 13h30</b>	<b>Dîner</b> (Champions / Rive Gauche / Beau Rivage A)	
<b>13h30 à 15h</b> Sessions concomitantes	<b>Session 4</b> : Questions relatives à l'analyse des données longitudinales sur la santé (Chaudière C)	<b>Session 5</b> : Combinaison de données provenant de sources différentes (Chaudière AB)
<b>15h à 15h30</b>	<b>Session d'affiches B</b> (Foyer de Chaudière ABC) Pause-café	
<b>15h30 à 17h</b> Sessions concomitantes	<b>Session 6</b> : Questions relatives à la pondération, à l'estimation et à la détection des observations aberrantes (Chaudière AB)	<b>Session 7</b> : Enquêtes sur la santé utilisant les mesures directes : problèmes et possibilités (Chaudière C)
<b>17h à 19h</b>	<b>Soirée rencontre</b> (Rive Gauche/Beau Rivage A)	

### Jour 3 – vendredi le 3 novembre

<b>8h à 9h</b>	Inscription, Rafrâichissements (Foyer de Chaudière ABC)	
<b>9h à 10h</b> Session plénière	<b>Session 8</b> : Discours du Prix Waksberg – Alastair Scott (Chaudière ABC)	
<b>10h à 10h30</b>	<b>Session d'affiches C</b> (Foyer de Chaudière ABC) Pause-café	
<b>10h30 à 12h</b> Sessions concomitantes	<b>Session 9</b> : Méthodes d'analyse des données sur la santé (Chaudière C)	<b>Session 10</b> : Effets du mode et collecte des données (Chaudière AB)
<b>12h à 13h30</b>	<b>Dîner</b> (Champions/Rive Gauche/Beau Rivage A)	
<b>13h30 à 15h</b> Sessions concomitantes	<b>Session 11</b> : Échantillonnage – théorie et applications (Chaudière AB)	<b>Session 12</b> : Mesure de la santé : une perspective internationale (Chaudière C)
<b>15h à 15h15</b>	Pause (Foyer de Chaudière ABC)	
<b>15h15 à 16h45</b> Sessions concomitantes	<b>Session 13</b> : Confidentialité et contrôle de la divulgation (Chaudière AB)	<b>Session 14</b> : Questions relatives à la conception et à la mise en œuvre des enquêtes sur la santé (Chaudière C)
<b>16h45 à 17h</b>	<b>Mot de clôture</b> (Chaudière C)	

# PROGRAMME

**Mercredi le 1<sup>er</sup> novembre 2006**

8h à 9h                      **Inscription aux ateliers–Foyer de Chaudière ABC**  
**Rafrâichissements - Foyer de Chaudière ABC**

---

9h à 16h30

**Ateliers concomitants**

10h à 10h30

**Rafrâichissements**

12h à 13h30

**Dîner** (Le dîner n'est pas inclus dans les  
frais d'inscription)

15h à 15h30

**Rafrâichissements**

- **(A,F) Couplage d'enregistrements dans le cadre d'études sur la santé de la population – une vue d'ensemble – Chaudière B**  
Karla Fox, ministère de la Défense nationale  
Patricia Whitridge, Élections Canada
- **(A) Méthodes pour l'analyse des données d'enquêtes longitudinales sur la santé. Théorie et applications – Chaudière A**  
Mary Thompson, Université de Waterloo, Canada

\* La lettre en majuscule devant le titre indique la langue de la  
présentation.

A=Anglais, F=Français

## Jeudi le 2 novembre 2006

7h45 à 8h45

**Inscription – Foyer de Chaudière ABC**  
**Rafraîchissements – Foyer de Chaudière ABC**

---

8h45 à 9h

**(A,F) Mot d'ouverture – Chaudière ABC**

François Maranda, Statistique Canada

---

9h à 10h

**Discours principal – Chaudière ABC**

- **(A) La mesure de la santé dans les enquêtes de population**  
Graham Kalton, Westat Inc., Washington, DC, Etats-Unis
- 

10h à 10h30

**Session d'affiches A– Foyer de Chaudière ABC**  
(Voir page 18 pour la liste des présentations)

---

**Pause-café – Foyer de Chaudière ABC**

---

10h30 à 12h

**Session 2 – Enjeux actuels relatifs à l'estimation pour petits domaines avec applications pour les enquêtes sur la santé – Chaudière AB**

Organisateur : Avi Singh, Statistique Canada

Président : Robert Fay, U.S. Census Bureau

- **(A) Modèles mixtes linéaires et non linéaires pour l'estimation pour petits domaines avec application à l'Enquête sur la santé dans les collectivités canadiennes**  
Avi Singh, Statistique Canada  
François Verret, Statistique Canada
- **(A) Traitement des observations influentes et des observations aberrantes dans l'estimation pour petits domaines**  
William R. Bell, U.S. Census Bureau  
Elizabeth T. Huang, U.S. Census Bureau

## Jeudi le 2 novembre 2006

- **(A) Inférence bayésienne prédictive robuste pour les quantiles de population finie d'un petit domaine**  
Balgobin Nandram, Worcester Polytechnic Institute, États-Unis  
Jai Won Choi, National Center for Health Statistics, CDC, États-Unis
- **(A) Estimation robuste de l'erreur quadratique moyenne (EQM) du meilleur prédicteur linéaire sans biais empirique (MPLNB) de la moyenne pour petits domaines**  
Partha Lahiri, University of Maryland, États-Unis

10h30 à 12h      **Session 3 – Analyse des données sur la santé – applications – Chaudière C**

Organisateur : Dave Dolson, Statistique Canada

- **(F) Les défis méthodologiques reliés à l'analyse des données nutritionnelles de l'Enquête sur la nutrition**  
François Verret, Statistique Canada
- **(F) Estimation des courbes de croissance de l'IMC chez l'enfant au Canada**  
Stéphane Tremblay, Statistique Canada  
Rebecca Morrison, Statistique Canada
- **(A) Écart entre les taux d'obésité chez les femmes américaines et canadiennes ainsi que chez les hommes américains et canadiens : résultats de l'Enquête conjointe Canada-États-Unis sur la santé**  
Jane F. Gentleman, National Center for Health Statistics, États-Unis  
Debra L. Blackwell, National Center for Health Statistics, États-Unis  
Michael E. Martinez, National Center for Health Statistics, États-Unis
- **(F) L'étude de scénarios d'intervention en nutrition : l'exemple des boissons gazeuses**  
Didier Garriguet, Statistique Canada

---

12h à 13h30      **Dîner – Champions / Rive Gauche / Beau Rivage A**

---

## Jeudi le 2 novembre 2006

13h30 à 15h

### **Session 4 – Questions relatives à l’analyse des données longitudinales sur la santé – Chaudière C**

Présidente : Georgia Roberts, Statistique Canada

- **(A) Erreur de mesure dans les données sur le cycle de vie**  
Grace Y. Yi, University of Waterloo, Canada
- **(A) Méthodes souples pour l’analyse des données longitudinales sur la santé de la population**  
Joel A. Dubin, University of Waterloo, Canada
- **(A) Analyse de données familiales longitudinales sur la santé provenant d’enquêtes à plan de sondage complexe : une vue d’ensemble**  
Brajendra Sutradhar, Memorial University of Newfoundland, Canada
- **(A) Estimation du nombre attribuable de décès et des erreurs-types d’après des cohortes échantillonnées selon des plans de sondage simples et complexes**  
Barry I. Graubard, National Cancer Institute, États-Unis  
Mitchell H. Gail, National Cancer Institute, États-Unis  
Katherine M. Flegal, National Center for Health Statistics, États-Unis  
David F. Williamson, Centers for Disease Control and Prevention, États-Unis

13h30 à 15h

### **Session 5 – Combinaison de données provenant de sources différentes – Chaudière AB**

Présidente : Jillian Oderkirk, Statistique Canada

- **(A) Études de qualité en vue d’évaluer le couplage de données aux données du Recensement de la population**  
Glenys Bishop, Analytical Services Branch, ABS, Australie

## Jeudi le 2 novembre 2006

- **(A) Appariement de données censitaires régionales avec des données d'enquête et des données administratives**  
Denis Gonthier, Statistique Canada  
Tina Hotton, Statistique Canada  
Cynthia Cook, Statistique Canada  
Russell Wilkins, Statistique Canada
- **(A) Une étude de cas du recours à l'estimation assistée par modèle pour intégrer des données d'enquête et des données administratives**  
Robert E. Fay, U.S. Census Bureau
- **(A) Couplage d'enregistrements exact et probabiliste**  
Claude Nadeau, Statistique Canada  
Marie P. Beaudet, Statistique Canada  
Jocelyne Marion, Statistique Canada  
Christel Le Petit, Statistique Canada

---

15h à 15h30      **Session d'affiches B – Foyer de Chaudière ABC**  
(Voir page 18 pour la liste des présentations)

---

**Pause-café – Foyer de Chaudière ABC**

---

15h30 à 17h      **Session 6 – Questions relatives à la pondération, à l'estimation et à la détection des observations aberrantes – Chaudière AB**

Président : Jack Gambino, Statistique Canada

- **(A) Modèles de sélection des variables pour l'élagage des coefficients de pondération**  
Michael Elliott, University of Michigan School of Public Health, États-Unis
- **(A) Combinaison de renseignements provenant de deux enquêtes pour améliorer l'analyse des données autodéclarées utilisées pour estimer les mesures de la santé**  
Nathaniel Schenker, National Center for Health Statistics, États-Unis  
Trivellore E. Raghunathan, National Center for Health Statistics, États-Unis  
Irina Bondarenko, National Center for Health Statistics, États-Unis

## Jeudi le 2 novembre 2006

- **(A) Évaluation des méthodes de détection et de traitement des observations aberrantes dans la « Survey of Occupational Injuries and Injuries » des États-Unis**  
John L. Eltinge, Bureau of Labour Statistics, États-Unis
- **(A) La combinaison de cycles de l'enquête sur la santé dans les collectivités canadiennes**  
Steven Thomas, Statistique Canada

15h30 à 17h      **Session 7 – Enquêtes sur la santé utilisant les mesures directes : problèmes et possibilités – Chaudière C**

Organisateur : Mark Tremblay, Statistique Canada  
Président : Michael Wolfson, Statistique Canada

- **(A) L'apprentissage des défis uniques que posent les enquêtes utilisant les mesures directes de la santé : l'expérience canadienne**  
Mark Tremblay, Statistique Canada
- **(A) La « National Health and Nutrition Examination Survey » des États-Unis : Ce que 40 années d'expérience nous ont appris**  
Cliff Johnson, National Centre for Health Statistics, États-unis
- **(A) Enquêtes utilisant les mesures directes de la santé en Finlande : l'expérience de la Carélie du Nord et au delà**  
Arpo Aromaa, KTL, Helsinki, Finlande

---

17h à 19h      **Soirée rencontre – Rive Gauche / Beau Rivage A**

---

## Vendredi le 3 novembre 2006

8h à 9h                    **Inscription – Foyer de Chaudière ABC**  
**Rafraîchissements – Foyer de Chaudière ABC**

---

9h à 10h                   **Session 8 – Discours du Prix Waksberg –**  
**Chaudière ABC**

Organisateur : John Kovar, Statistique Canada  
Président : Wayne Fuller, Iowa State University, États-Unis

- **(A) L'analyse des études avec témoins représentatives de la population**  
Alastair Scott, University of Auckland, Nouvelle-Zélande
- 

10h à 10h30               **Session d'affiches C – Foyer de Chaudière ABC**  
(Voir page 19 pour la liste des présentations)

---

**Pause-café – Foyer de Chaudière ABC**

---

10h30 à 12h               **Session 9 – Méthodes d'analyse des données**  
**sur la santé – Chaudière C**

Président : David Binder, Statistique Canada

- **(A) Méthodes bootstrap pour l'analyse de données d'enquêtes à plan de sondage complexe**  
J. N. K. Rao, Carleton University, Canada
- **(A) Estimation des paramètres de régression au moyen de données d'enquête**  
Wayne A. Fuller, Iowa State University, États-Unis  
Yu Wu, Iowa State University, IA, États-Unis
- **(A) Tenir compte de la dispersion des données en recherche contextuelle sur la santé de la population : effets de la petite taille des groupes et de l'analyse par grappes ciblée sur les résultats, le biais et l'efficacité des modèles multiniveaux**  
Philippa Clarke, University of Michigan, États-Unis  
Patricia O'Campo, University of Toronto, Canada  
Blair Wheaton, University of Toronto, Canada

## Vendredi le 3 novembre 2006

- **(A) Inférence causale dans les études par observation utilisant des données administratives sur la santé**  
Therese A. Stukel, ICES, Toronto, Canada

10h30 à 12h      **Session 10 – Effets du mode et collecte des données – Chaudière AB**

Président : Yves Béland, Statistique Canada

- **(A) Passage à la téléphonie sans fil aux États-Unis et au Canada : prévalence et effet sur les enquêtes téléphoniques sur la santé utilisant la composition aléatoire de numéros de téléphone**  
Stephen J. Blumberg, Centers for Disease Control and Prevention, National Center for Health Statistics, États-Unis  
Julian V. Luke, Centers for Disease Control and Prevention, National Center for Health Statistics, États-Unis
- **(A) Étude de l'effet du mode de collecte sur les estimations clés de l'état de santé dans la « National Health Interview Survey »**  
Catherine Simile, National Center for Health Statistics, États-Unis  
Barbara Stussman, National Center for Health Statistics, États-Unis  
James Dahlhamer, National Center for Health Statistics, États-unis
- **(A) Effet de l'interview téléphonique comparativement à l'interview sur place pour le rappel alimentaire de 24 heures dans les enquêtes sur l'alimentation et la nutrition**  
S. Hayward, Direction des aliments, Santé Canada  
E. Junkins, Direction des aliments, Santé Canada  
M. Vigneault, Direction des aliments, Santé Canada  
M. Villeneuve, Direction des aliments, Santé Canada
- **(F) Comment tenir compte d'un changement méthodologique important dans l'analyse des données de l'Enquête sur la santé dans les collectivités canadiennes?**  
Luc Côté, Institut de la statistique du Québec, Canada  
Robert Courtemanche, Institut de la statistique du Québec, Canada

---

12h à 13h30      **Dîner – Champions / Rive Gauche / Beau Rivage A**

---

## Vendredi le 3 novembre 2006

13h30 à 15h **Session 11 - Échantillonnage – théorie et applications – Chaudière AB**

Président : Mike Hidioglou, Statistique Canada

- **(A) Échantillonnage PPT systématique avec substitution d'unités basé sur la simulation**  
Mary E. Thompson, University of Waterloo, Canada  
Changbao Wu, University of Waterloo, Canada
- **(A) Réutilisation des grappes dans les études répétées**  
Stanislav Kolenikov, University of Missouri, États-Unis  
Gustavo Angeles, University of North Carolina, Carolina Population Center, Chapel Hill, États-Unis
- **(F) Plan d'échantillonnage de l'Enquête canadienne sur les mesures de la santé**  
Suzelle Giroux, Statistique Canada
- **(A) Les défis posés par la conception de la « National Health and Nutrition Examination Survey »**  
Leyla Mohadjer, Westat, Washington, DC, États-Unis

13h30 à 15h **Session 12 - Mesure de la santé : une perspective internationale - Chaudière C**

Organisateur et Président : Jean-Marie Berthelot, CIHI

- **(A) La mesure de la santé des populations: l'approche analytique et conceptuelle de l'étude "Global Burden of Disease"**  
Colin D. Mathers, Department of Measurement and Health Information, Organisation mondiale de la santé, Genève, Suisse
- **(A) Comparaisons internationales dans la mesure des états de santé: expériences provenant des enquêtes mondiales sur la santé**  
Somnath Chatterji, Department of Measurement and Health Information, Organisation mondiale de la santé, Genève, Suisse

## Vendredi le 3 novembre 2006

- **(A) La stratégie de recherche pour développer des mesures comparables d'un statut de santé fonctionnel à utiliser pour les enquêtes sur la population**  
Chris Murray, Harvard University, États-Unis
  - **Intervenant :** Michael Wolfson, Statistique Canada
- 

15h à 15h15      **Pause – Foyer de Chaudière ABC**

---

15h15 à 16h45      **Session 13 – Confidentialité et contrôle de la divulgation – Chaudière AB**

Président : Jean-Louis Tambay, Statistique Canada

- **(A) Application des méthodes de contrôle de la divulgation statistique à la base de données du Système canadien hospitalier d'information et de recherche en prévention des traumatismes**  
Ann Brown, Statistique Canada  
Margaret Herbert, Agence de santé publique du Canada
- **(A) Masquage dans le cas de variables discrètes**  
Myron J. Katzoff, National Center for Health Statistics, États-Unis  
Jay J. Kim, National Center for Health Statistics, États-Unis
- **(A) Création de fichiers de microdonnées à grande diffusion pour la « National Survey on Drug Use and Health »**  
Feng Yu, RTI International, États-unis  
Lanting Dai, RTI International, États-unis  
Moshe Feder, RTI International, États-unis  
James R. Chromy, RTI International, États-unis
- **(A) Confidentialité et estimation de la variance par rééchantillonnage**  
Wilson Lu, Simon Fraser University, Canada

## Vendredi le 3 novembre 2006

15h15 à 16h45 **Session 14 – Questions relatives à la conception et à la mise en œuvre des enquêtes sur la santé - Chaudière C**

Présidente : Julie McAuley, Statistique Canada

- **(A) Questions méthodologiques liées à la mesure de la santé mentale des enfants et des jeunes en Grande-Bretagne**  
Howard Meltzer, Office for National Statistics, Londres, Angleterre  
Amanda Wilmot, Office for National Statistics, Londres, Angleterre  
Abigail Dewar, Office for National Statistics, Londres, Angleterre
- **(A) Échantillonnage de la population Maori dans la « New Zealand Health Survey »**  
Robert Clark, University of Wollongong, NSW, Australie  
Sarah Gerritsen, Public Health Intelligence, Ministère de la Santé, Nouvelle Zélande
- **(A) Élaboration et mise à l'essai d'un questionnaire sur la santé de l'enfant avec réponse par procuration par un parent ou un gardien pour la « New Zealand Health Survey »**  
Sarah Gerritsen, conseillère (enquêtes), Public Health Intelligence, Ministère de la Santé, Nouvelle-Zélande
- **(A) Conception de questions pour le dépistage des personnes handicapées dans les enquêtes sur la population active : mesures prises en vue de déterminer le niveau d'emploi des adultes handicapés dans la « CPS »**  
Terence M. McMenamin, Division of Labour Force Statistics, Bureau of Labour Statistics, États-Unis

---

16h45 à 17h **(A,F) Mot de clôture – Chaudière C**

Don Royce, Statistique Canada

---

## Sessions d'affiches

### Session d'affiches A

- **(A) Recherche par simulation sur les méthodes d'imputation hot deck et d'imputation multiple à l'aide de données de la HCSDB**  
Donsig Jang, Mathematica Policy Research. Inc., États-Unis  
Among Sukasih, Mathematica Policy Research. Inc., États-Unis  
Xiaojing Lin, Mathematica Policy Research. Inc., États-Unis
  - **(A) Estimation de la taille d'échantillon pour les enquêtes complexes : Réalisation d'un consensus dans un environnement d'hypothèses multiples, d'intervenants multiples et de contraintes budgétaires**  
JC Victor<sup>1,2</sup>, L. Diemert<sup>1,2</sup>, S Bondy<sup>1,2</sup>, KS Brown<sup>1,3</sup>, J Cohen<sup>1,2</sup>, R Ferrence<sup>1,2</sup>, PW McDonald<sup>1,3</sup>, P Selby<sup>1,4</sup>, T Stephens<sup>1</sup>  
<sup>1</sup>Unité de recherche sur le tabac de l'Ontario; <sup>2</sup>Université de Toronto; <sup>3</sup>Université de Waterloo; <sup>4</sup>Centre de toxicomanie et de santé mentale
  - **(A) L'effet de la spécification du modèle sur les données résultant de l'imputation multiple : leçons tirées du projet DC-HOPE du NIH**  
Marie G. Gantz, , NIH-DC Initiative, RTI International, États-Unis  
M. Nabil El-Khorazaty, NIH-DC Initiative, RTI International, États-Unis
  - **(A) Validation d'une méthode qualitative appliquée à un instrument multidimensionnel comprenant une question ouverte**  
José Eduardo Corrente, UNESP, Botucatu, São Paulo, Brésil.  
Cinthia Esbrile Moraes, UNESP, Botucatu, São Paulo, Brésil.  
Tania Ruiz, UNESP, Botucatu, São Paulo, Brésil.
- 

### Session d'affiches B

- **(A) Élaboration et conception de l'enquête sur le tabagisme en Ontario**  
S. Bondy<sup>1,2</sup>, L. Diemert<sup>1,2</sup>, J.C. Victor<sup>1,2</sup>, K.S. Brown<sup>1,3</sup>, J. Cohen<sup>1,2</sup>, R. Ferrence<sup>1,2</sup>, P.W. McDonald<sup>1,3</sup>, P. Selby<sup>1,4</sup>, T. Stephens,<sup>1</sup> <sup>1</sup>Unité de recherche sur le tabac de l'Ontario; <sup>2</sup>Université de Toronto; <sup>3</sup>Université de Waterloo; <sup>4</sup>Centre de toxicomanie et de santé mentale
- **(F) Un modèle pour l'estimation de l'espérance de vie dans les petites villes canadiennes**  
Philippe Finès, Statistique Canada
- **(F) Infarctus du myocarde au Québec : les immigrants et les Canadiens de naissance sont-ils différents ?**  
Maria Gabriela Orzanco, Université de Sherbrooke, Canada  
Alain Vanasse, Université de Sherbrooke, Canada  
Théophile Niyonsenga, Université de Sherbrooke, Canada  
Josiane Courteau, Université de Sherbrooke, Canada

- **(A) Géovisualisation des données sur la santé et sur le capital social provenant des enquêtes réalisées par Statistique Canada**  
Daniel Rainham, Université d'Ottawa, Canada  
Daniel Krewski, Université d'Ottawa, Canada  
Ian McDowell, Université d'Ottawa, Canada  
Mike Sawada, Université d'Ottawa, Canada
- 

### Session d'affiches C

- **(A) Les effets de différentes méthodes de sélection des dossiers de sortie d'hôpital aux fins de l'établissement de rapports sur les blessures**  
Susan G. Mackenzie, Agence de santé publique du Canada
  - **(A) Remplacement de la CIM-9 par la CIM-10-CA dans les études sur les soins de santé hospitaliers**  
Christie Sambell, Statistique Canada  
Hude Quan, Statistique Canada  
Helen Johansen, Statistique Canada
  - **(A) Examen de l'incidence de la résistance du participant sur la qualité des données dans la National Health Interview Survey**  
James M. Dahlhamer, National Center for Health Statistics, Centers for Disease Control and Prevention, États-Unis  
Catherine M. Simile, National Center for Health Statistics, Centers for Disease Control and Prevention, États-Unis  
Net Taylor, National Center for Health Statistics, Centers for Disease Control and Prevention, États-Unis
  - **(A) Effet des incitatifs sans condition sur une enquête à plusieurs degrés : cas de la Health Survey for England**  
C. Deverill, National Centre for Social Research, Londres, Royaume-Uni  
H. Wardle, National Centre for Social Research, Londres, Royaume-Uni  
R. Craig, National Centre for Social Research, Londres, Royaume-Uni
  - **(A) Comparaison de l'utilisation autodéclarée des soins de santé mentale primaires dans l'Enquête sur la santé dans les collectivités canadiennes et des dossiers d'assurance-maladie provinciaux des répondants**  
JoAnne Palin, Université de Colombie-Britannique, Vancouver, C.-B., Canada
-

# RÉSUMÉS

## Ateliers

### **(A,F) Couplage d'enregistrements dans le cadre d'études sur la santé de la population — une vue d'ensemble**

Karla Fox, ministère de la Défense nationale et Patricia Whitridge, Élections Canada (Présentation en français et anglais avec traduction simultanée et avec le matériel en anglais et en français)

Le couplage d'enregistrements est simplement l'intégration d'informations provenant de deux sources indépendantes ou plus. Dans ce contexte, les enregistrements sont couplés sur la base de données communes. Le couplage d'enregistrements prend de plus en plus d'importance dans de nombreux domaines différents : tenue à jour de données de registres électroniques, administration des soins de santé, études démographiques et recherche médicale. Qu'ils suivent une cohorte au fil du temps ou qu'ils couplent des patients et des données de l'état civil afin de calculer des courbes de survie, les chercheurs ont besoin de comprendre le couplage d'enregistrements.

À partir d'exemples provenant de diverses applications du domaine de la santé, cet atelier portera sur les principes généraux du couplage d'enregistrements (appariement statistique et appariement exact), y compris la préparation des données, les techniques de couplage et l'évaluation du couplage. Des exemples pratiques serviront à illustrer différents points tout au long du cours. Les participants pourront prendre connaissance des différents concepts et principes touchés et comprendront la façon d'aborder un problème nécessitant le couplage d'enregistrements. À la fin, les participants recevront une liste exhaustive de références dans le domaine du couplage d'enregistrements.

**(A) Méthodes pour l'analyse des données d'enquêtes longitudinales sur la santé – Théorie et applications**

Mary Thompson, Université de Waterloo, Waterloo, Canada  
(Présentation en anglais avec traduction simultanée et avec le matériel en anglais et en français)

L'atelier débutera par une vue d'ensemble des modèles pour les données longitudinales avec des exemples issus des sciences de la santé. Les modèles de mesures répétées considérés incluront les modèles de croissance, les équations d'estimation généralisées EEG et les modèles de processus stochastiques. Les modèles à événements répétés incluront les modèles de survie et de récurrence. Les différents types de variables explicatives (fixes, variant dans les temps, internes, externes) ainsi que les types de données manquantes seront passés en revue. La vue d'ensemble sera poursuivie par une discussion des études observationnelles et de causalité à travers des exemples illustrant l'utilisation des données longitudinales ou quasi-longitudinales pour examiner des hypothèses de causalité. Le rôle des modèles graphiques pour modéliser les dépendances sera considéré. Les principes pour adapter les modèles et méthodes à des données d'enquêtes complexes seront brièvement présentés avec des exemples provenant des enquêtes canadiennes sur la santé. Les problèmes présents dans les modèles avec des vraisemblances complexes (particulièrement ceux associés aux données manquantes et aux variables latentes) seront illustrés. L'atelier se terminera par une discussion et une brève revue des logiciels disponibles, ou encore en développement, pour les données d'enquêtes longitudinales de la santé.

---

**Session 1 - Discours principal****(A) La mesure de la santé dans les enquêtes auprès de population**

Graham Kalton, Westat Inc., Washington, DC, États-Unis

Les données sur la santé de la population sont recueillies à la fois dans le cadre d'enquêtes axées sur la santé et dans le cadre de nombreuses enquêtes sur d'autres sujets en raison du rôle majeur que joue la santé auprès d'autres aspects de la vie. Les grands thèmes des enquêtes sur la santé de la population englobent la santé physique et mentale, la santé dentaire, les incapacités, la toxicomanie, les facteurs de risque pour la santé, l'alimentation, la promotion de la santé, l'utilisation et la qualité des soins de santé, leur portée et leurs coûts. Certaines enquêtes s'intéressent à des problèmes de santé particuliers, tandis que d'autres visent l'évaluation globale de la santé. La santé est souvent une composante importante des études longitudinales, particulièrement pour les cohortes de naissance et les cohortes sur le vieillissement. Des données sur la santé peuvent être recueillies au moyen de déclarations des répondants (pour eux-mêmes et parfois pour d'autres), d'exams médicaux et de prélèvements biologiques. L'exactitude des données sur la santé qui sont recueillies à partir des déclarations des répondants nous préoccupe sérieusement. Les problèmes de logistique et de coût ainsi que la collaboration des répondants figurent au premier plan des préoccupations concernant les données recueillies au moyen d'un examen médical. Les questions d'éthique et de protection des renseignements personnels sont souvent importantes, particulièrement en présence d'ADN et de biomarqueurs. La comparabilité internationale des mesures de la santé prend de plus en plus d'importance. Ce document passe en revue la méthodologie s'appliquant à toute une gamme d'enquêtes sur la santé et aborde les défis liés à l'obtention de données précises dans ce domaine.

**Session 2 – Enjeux actuels relatifs à l'estimation pour petits domaines avec applications pour les enquêtes sur la santé****(A) Modèles mixtes linéaires et non linéaires pour l'estimation pour petits domaines avec application à l'Enquête sur la santé dans les collectivités canadiennes**

Avi Singh et François Verret, Statistique Canada, Ottawa, ON, Canada

Nous examinons le problème de la modélisation d'estimations directes de dénombrements de population ayant des caractéristiques d'intérêt pour de petits domaines à l'aide de données d'enquête en utilisant l'information d'autres domaines afin de renforcer la précision des estimations. Les praticiens considèrent intéressant le modèle mixte

linéaire (MML) de Fay et Herriot (1979) utilisé couramment pour les données agrégées (ou au niveau des domaines), en raison de la simplicité des calculs et de l'interprétation des résultats. Lorsqu'on analyse des données multivariées au cours du temps, les avantages de la simplicité des calculs peuvent être considérables, parce que le MML permet d'appliquer des techniques de filtration récursive, comme celle de Kalman, en classant les domaines de manière adéquate et en traitant le rang comme la variable de temps. Les estimations résultantes ont la propriété d'être les meilleurs prédicteurs sans biais (BUP, pour *best unbiased predictors*) sous l'hypothèse de normalité des estimations directes. Cependant, l'optimalité peut être fragile si les paramètres d'intérêt doivent satisfaire à des contraintes d'intervalle de valeurs. Par exemple, pour des données discrètes, telles que les dénombrements ou les proportions, il est naturel d'utiliser un modèle mixte linéaire non linéaire (MMLNL) ou un modèle mixte linéaire généralisé (MMLG) pour faire en sorte que les estimations au niveau des domaines satisfassent aux contraintes d'intervalle appropriées. Ainsi, en suivant la méthode hiérarchique bayésienne (HB), on peut utiliser le MMLG au niveau agrégé pour l'analyse de données d'enquête en recourant à des méthodes MCMC pour calculer les estimations optimales, comme dans les travaux de You et Rao (2002). Dans la présente communication, nous proposons une solution de rechange plus simple, analogue au MML, qui évite le recours à la formulation HB. Nous utilisons à cette fin un modèle mixte linéaire non linéaire (MMLNL) où la partie aléatoire de la moyenne est linéaire et s'ajoute à la partie fixe non linéaire ou à la moyenne marginale. Nous pouvons justifier la structure additive des effets aléatoires en partant de la spécification de la structure de la covariance de l'erreur marginale en présence d'échantillonnage en grappes ou d'une loi a priori sur les moyennes des grappes, comme cela se fait habituellement pour définir la surdispersion. En présence de conditions de normalité approximative des estimations directes et en utilisant la modélisation MMLNL, nous pouvons définir les meilleurs prédicteurs linéarisés et approximativement sans biais (BLAUP pour *best linearized and approximately unbiased predictors*) comme étant une généralisation directe des BUP. Il est vraisemblable que le BLAUP satisfasse aux contraintes d'intervalle, parce que les moyennes marginales et les estimations directes le font et que l'effet d'autres estimations directes corrélées sur l'effet aléatoire estimé sera probablement faible. Nous constatons que la modélisation MMLNL peut produire des résultats sensiblement meilleurs que le MML tout en retenant la simplicité de ce dernier. Nous décrivons une application de la méthode aux données de l'Enquête sur la santé dans les collectivités canadiennes en vue d'estimer la proportion (prévalence) de certaines caractéristiques de la santé dans des sous-populations telles que les régions sociosanitaires selon le groupe d'âge et le sexe dans une province.

**(A) Traitement des observations influentes et des observations aberrantes dans l'estimation pour petits domaines**

William R. Bell et Elizabeth T. Huang, U.S. Census Bureau, Washington D.C., États-Unis

L'estimation de petits domaines au moyen de modèles linéaires de niveau du domaine donne des prédictions pour de petits domaines qui sont des moyennes pondérées d'estimations directes d'enquête et de prédictions par la régression. Dans ces conditions, il est intéressant d'examiner les données qui influent sur l'ajustement des droites de régression et sur les prédictions du modèle. Certaines mesures d'influence utilisées dans un contexte de régression conventionnelle par les moindres carrés, telles que la distance de Cook (1977), ont été généralisées et appliquées à des modèles linéaires mixtes utilisés dans le cas de l'estimation de petits domaines (Rao, 2003). Nous appliquons ces mesures à des modèles établis par le Small Area Income and Poverty Estimates Program (programme SAIPE) du Bureau du recensement des États-Unis, et appliquons parallèlement une mesure de distance de Mahalanobis directement à la matrice de régression. Les valeurs aberrantes peuvent aussi poser un problème, mais elles sont parfois difficiles à déceler en raison de l'importance de l'erreur d'échantillonnage des estimations directes d'enquête (ce qui a d'abord motivé l'estimation pour de petits domaines). Nous illustrons l'utilisation d'une loi  $t$  afin de tenir compte de l'existence éventuelle de valeurs aberrantes dans un modèle bivarié pour les taux de pauvreté d'États américains, en examinant les effets sur les prédictions du modèle et sur les estimations connexes des erreurs quadratiques moyennes des prédictions.

**(A) Inférence bayésienne prédictive robuste pour les quantiles de population finie d'un petit domaine**

Balgobin Nandram<sup>1</sup> et Jai Won Choi<sup>2,1</sup>, Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, États-Unis,<sup>2</sup> Office of Research and Methodology, National Center for Health Statistics, CDC, Hyattsville, Maryland, États-Unis

En suivant une méthode bayésienne robuste, nous examinons le problème de l'analyse des données en présence de non-réponse non-ignorable. Sachant les paramètres, nous supposons que les indicateurs de réponse sont des variables aléatoires de Bernoulli indépendantes et que la variable de réponse est dépourvue de distribution. Ici, une inférence prédictive est requise pour le nième quantile de la population d'un sous-domaine (un secteur dans un petit domaine). Nous supposons que nous disposons d'un échantillon aléatoire provenant d'une population finie, que des covariables sont également disponibles pour chaque unité de la population finie et que ces covariables peuvent expliquer adéquatement l'écart entre les

répondants et les non-répondants. Nous utilisons un modèle de régression logistique robuste pour établir le lien entre les indicateurs de réponse et les covariables pour l'échantillon. Nous obtenons ce modèle en élargissant le modèle de régression logistique conventionnel à un mélange de lois  $t$  de Student, ce qui nous fournit des scores de propension que nous utilisons pour construire des cellules d'ajustement. Nous complons les variables manquantes des non-répondants en tirant des échantillons des variables de réponse des répondants provenant des mêmes cellules d'ajustement. Nous fondons la prédiction sur une régression linéaire fondée sur les rangs de la variable de réponse sur les covariables selon le domaine, ce qui rend la méthode encore plus robuste. Nous utilisons des méthodes fondées sur les chaînes de Markov par simulation de Monte Carlo (MCMC) pour ajuster notre modèle. Dans chaque sous-domaine, nous faisons une inférence sur des quantiles de la variable de réponse en utilisant les statistiques d'ordre sur l'ensemble des individus (échantillonnés et non échantillonnés) et nous obtenons la distribution de cette variable en utilisant les itérations provenant de l'échantillonneur MCMC. Nous appliquons la méthode aux données du troisième cycle de la National Health and Nutrition Examination Survey en prenant pour variable de réponse l'indice de masse corporelle (IMC), et pour covariables le groupe ethnique et le sexe; les données pour ces variables sont disponibles pour trente-cinq comtés (petits domaines). Nous obtenons les quantités d'intérêt à partir de l'inférence prédictive bayésienne des quantiles d'IMC de la population finie (85<sup>e</sup> percentile pour l'embonpoint et 95<sup>e</sup> centile pour l'obésité) pour chaque domaine (sous-domaines) en fonction de l'âge, du groupe ethnique et du sexe dans chaque comté. Nous comparons notre méthode robuste à une méthode paramétrique proposée récemment.

**(A) Estimation robuste de l'erreur quadratique moyenne (EQM) du meilleur prédicteur linéaire sans biais empirique (MPLNB) de la moyenne pour petits domaines**

Partha Lahiri, University of Maryland, College Park, Maryland, États-Unis

Cette communication a pour but de présenter une méthode générale d'estimation de l'erreur quadratique moyenne (EQM) du meilleur prédicteur linéaire sans biais empirique (MPLSB) pour le modèle bien connu de Fay-Herriot pour de petits domaines. Nous obtenons d'abord une approximation de deuxième ordre de l'EQM, c'est-à-dire en omettant tous les termes d'ordre  $o(m-1)$ , où  $m$  représente le nombre de petits domaines. Contrairement aux approximations fondées sur l'hypothèse de normalité, la nôtre tient compte de l'aplatissement de la distribution des erreurs d'échantillonnage et de celle des erreurs de modélisation, et dépend de la méthode d'estimation des composantes de la variance. Nous constatons que, dans le cas de l'estimateur par la

méthode des moments (MM) des composantes de la variance, nous ne sommes pas obligés d'estimer l'aplatissement de la distribution des erreurs de modélisation pour obtenir un estimateur de deuxième ordre sans biais de l'EQM. En effet, le résultat antérieur de Lahiri et Rao (1995) est étendu à la situation où les erreurs d'échantillonnage ne suivent pas une loi normale. Cependant, l'estimation de l'aplatissement est nécessaire afin d'obtenir un estimateur de deuxième ordre sans biais de l'EQM si on utilise d'autres méthodes d'estimation des composantes de la variance, comme celle des équations d'estimation de Fay et Herriot (1979). Nous proposons une méthode d'estimation de l'aplatissement qui, à son tour, fournit une méthode d'estimation de l'EQM pour une classe très générale d'estimateurs des composantes de la variance dans des conditions de non-normalité des distributions des erreurs d'échantillonnage et des erreurs de modélisation. Nous empruntons des exemples aux domaines de la cartographie des maladies et des enquêtes sur la santé pour illustrer la méthodologie.

### Session 3 – Analyse des données sur la santé - applications

#### **(F) Les défis méthodologiques reliés à l'analyse des données nutritionnelles de l'Enquête sur la nutrition**

François Verret, Statistique Canada, Ottawa, ON, Canada

Le principal objectif de l'Enquête canadienne sur la nutrition de Statistique Canada est d'estimer, pour plusieurs domaines d'intérêt à l'échelle provinciale, les distributions de consommation habituelle pour différents éléments nutritifs requis pour une bonne santé. L'information nutritionnelle a été recueillie via un journal alimentaire de 24 heures avec second rappel chez un échantillon de plus de 36 000 Canadiens de tout âge.

La façon la plus courante de produire ces estimations est de traiter les données en ajustant un modèle à mesures répétées. Le modèle nous permet de déterminer l'importance relative de la variance de l'apport nutritionnel entre les individus (variabilité inter-individuelle) et de la variance journalière propre aux individus (variabilité intra-individuelle). Un des principaux problèmes qu'il est possible de rencontrer est d'obtenir des estimations de variance inter-individuelle négatives. Ceci représente évidemment un obstacle dans l'estimation des distributions d'intérêt.

De plus, étant donné que la variabilité d'échantillonnage est estimée avec la méthode de réplication du bootstrap, un modèle doit être ajusté pour chaque réplique. Ceci a pour effet d'engendrer ou d'amplifier certains problèmes. Entre autres, le temps d'exécution du logiciel

informatique Software for Intake Distribution (SIDE), logiciel commercial utilisé pour estimer les distributions, est non négligeable, ce qui représente un défi pour les utilisateurs et remet possiblement en question le nombre de répliques à utiliser. Également, on obtient des estimations de variance négatives pour certaines répliques bootstrap et positives pour d'autres.

Cet article décrira les défis méthodologiques ainsi que les solutions envisagées pour résoudre les problèmes engendrés par l'application du modèle à mesures répétées (à l'aide de SIDE) à des données d'enquête provenant d'un plan de sondage complexe où la méthode de réplication du bootstrap est utilisée pour mesurer la variabilité d'échantillonnage.

**(F) Estimation des courbes de croissance de l'IMC chez l'enfant au Canada**

Stéphane Tremblay et Rebecca Morrison, Statistique Canada, Ottawa, Canada

Les courbes de croissance sont utilisées par les professionnels de la santé afin de déterminer si la croissance d'un enfant ou d'un fœtus, par exemple, se situe dans les limites normales. Les graphiques de croissance employés à l'heure actuelle au Canada pour la taille, le poids et l'indice de masse corporelle (IMC) sont fondés sur des données américaines. Des courbes de croissance pour les enfants peuvent maintenant être établies d'après les données les plus récentes disponibles au Canada. L'une des méthodes qui seront étudiées en vue d'estimer et de tracer les courbes de croissance est la méthode Lambda-Mu-Sigma (LMS). Cette dernière a été utilisée dans le cadre de diverses études réalisées par l'Organisation mondiale de la santé, le Royaume-Uni et les États-Unis afin d'établir des courbes de croissance pour les enfants. La méthode LMS sera utilisée pour estimer les courbes de croissance en percentiles de l'IMC au moyen de données transversales pondérées. Des méthodes adaptées à des tailles d'échantillon plus petites seront également examinées. La communication portera sur l'IMC des enfants, l'une des mesures anthropométriques les plus couramment utilisées pour évaluer la croissance et l'obésité. Des mesures de l'IMC peuvent être obtenues des enquêtes telles que les cycles 2.2 et 3.1 de l'Enquête sur la santé dans les collectivités canadiennes et l'Enquête santé Canada de 1978-1979. Afin d'accroître le nombre d'observations, donc de rendre l'estimation des courbes plus précise, on combinera diverses sources pour la partie empirique de l'étude.

**(A) Écart entre les taux d'obésité chez les femmes américaines et canadiennes ainsi que chez les hommes américains et canadiens : résultats de l'Enquête conjointe Canada-États-Unis sur la santé**

Jane F. Gentleman, Debra L. Blackwell et Michael E. Martinez, National Center for Health Statistics, Division of Health Interview Statistics, Hyattsville, MD, États-Unis

L'Enquête conjointe Canada/États-Unis sur la santé de 2002-2003 est une enquête téléphonique qui a été réalisée de concert par Statistique Canada et le National Center for Health Statistics des États-Unis. Les deux pays ont utilisé essentiellement le même questionnaire au même moment, ce qui a fourni un ensemble de données assurant une comparabilité sans précédent des estimations nationales calculées pour les deux pays. Des résultats antérieurs obtenus au moyen de l'ensemble de données ont prouvé l'existence d'écarts notables entre les taux d'obésité chez les Américaines et les Canadiennes, mais non chez les hommes américains et canadiens. Plus précisément, 20 % d'Américaines étaient obèses, comparativement à 12 % seulement de Canadiennes, tandis que pour les hommes américains et canadiens, les taux étaient, respectivement, de 20 % et de 18 %. La présente étude vise à analyser plus en détail les facteurs associés à l'obésité, mesurée par l'indice de masse corporelle, dans les deux pays. Les résultats montrent que non seulement les taux d'obésité sont les mêmes pour les hommes dans les deux pays, mais aussi que les courbes de répartition de l'indice de masse corporelle sont presque identiques. L'analyse multivariée révèle que les facteurs associés significativement à l'obésité chez les Américaines, mais non significativement (ou moins significativement) chez les Canadiennes comprennent l'âge, la race ou l'ethnicité, le niveau de scolarité, le pays de naissance (immigrante ou non) et le fait d'avoir un médecin de famille. Par exemple, la cote exprimant le risque d'être obèse était 2,6 fois plus élevée pour les Américaines noires non hispaniques que pour leurs homologues de race blanche, alors qu'il n'existait aucun écart significatif comparable chez les Canadiennes. En outre, la cote exprimant le risque d'être obèse était 36 % plus faible pour les Américaines nées hors du pays que pour celles nées aux États-Unis, tandis que le pays de naissance n'était pas associé significativement à l'obésité chez les Canadiennes.

**(F) L'étude de scénarios d'intervention en nutrition : l'exemple des boissons gazeuses**

Didier Garriguet, Division de la statistique de la santé, Statistique Canada, Ottawa, ON, Canada

À l'aide des données de l'Enquête canadienne sur la nutrition, il est possible d'effectuer diverses analyses : mesurer la consommation quotidienne moyenne ou annuelle d'un aliment, mesurer la distribution habituelle de nutriments afin d'identifier les populations à risque qui ne respectent pas les recommandations de l'Institut de la Médecine ou mesurer la distribution habituelle des groupes alimentaires et les comparer avec le Guide alimentaire canadien pour manger sainement. Bref, il est possible d'évaluer l'état nutritionnel de la population.

Au niveau des aliments, la consommation habituelle est obtenue en multipliant la consommation habituelle de l'aliment à sa probabilité de consommation dans la population. Pour obtenir la consommation habituelle, on doit soustraire la variance intra-individuelle de la distribution de la consommation quotidienne. La probabilité de consommer un aliment est obtenue en regardant la fréquence de consommation sur plusieurs jours. Or, avec seulement deux mesures quotidiennes de données, il est difficile, voire impossible, d'évaluer cette probabilité pour certains aliments.

Dans cet article, nous explorerons d'autres avenues nous permettant de quantifier l'importance d'un aliment dans l'alimentation de la population. Nous étudierons divers scénarios de probabilité de consommation. Nous étudierons également des scénarios d'intervention, c'est-à-dire que nous tenterons de mesurer l'impact de la consommation d'un aliment soit en éliminant cet aliment de la diète des Canadiens, soit en le remplaçant par diverses alternatives et en comparant les apports en nutriments de ces divers scénarios.

Pour illustrer ces méthodes analytiques, l'exemple des boissons gazeuses sera utilisé. La consommation est suffisante pour obtenir une distribution de l'apport habituel des consommateurs alors qu'elle n'est pas suffisante pour obtenir la probabilité de consommation.

**Session 4 – Questions relatives à l'analyse des données longitudinales sur la santé****(A) Erreur de mesure dans les données sur le cycle de vie**

Grace Y. Yi, University of Waterloo, Waterloo, ON, Canada

Dans les études médicales, il arrive souvent que certaines données recueillies soient altérées par une erreur de mesure. Parfois, des covariables (ou facteurs de risque) d'intérêt sont difficiles à observer

avec précision en raison de l'emplacement physique ou du coût. De plus, il est quelquefois impossible de mesurer précisément les covariables en raison de leur nature. Dans d'autres situations, une covariable peut représenter la moyenne d'une certaine grandeur mesurable au fil du temps et tout moyen pratique de mesurer cette grandeur comporte nécessairement une erreur de mesure. Lorsqu'on procède à des inférences statistiques dans de telles conditions, il est important de tenir compte des effets des covariables mesurées incorrectement sinon les résultats risquent d'être incorrects, voire même trompeurs. Cette présentation portera sur les méthodes d'analyse statistique des données sur le cycle de vie en présence de covariables mesurées incorrectement. Plus précisément, elle sera axée sur les méthodes fonctionnelles qui sont très répandues en pratique. Certaines applications seront également présentées.

### **(A) Méthodes souples pour l'analyse des données longitudinales sur la santé de la population**

Joel A. Dubin, Departments of Statistics & Actuarial Science and Health Studies & Gerontology, University of Waterloo, Waterloo, ON, Canada

L'étude de données longitudinales (ou données recueillies au fil du temps) est essentielle si l'on veut observer correctement l'évolution au fil du temps des variables d'intérêt à l'égard des personnes, des collectivités et des populations plus importantes. Les modèles linéaires à effets mixtes (pour les réponses continues observées au fil du temps) et les modèles linéaires généralisés à effets mixtes et les équations d'estimation généralisées (pour les réponses plus générales, telles que les données binaires ou les dénombrements observés au fil du temps) sont les techniques les plus répandues pour analyser les données longitudinales provenant d'études sur la santé, même si, comme toute méthode de modélisation, elles ont leurs limites, dues en partie aux hypothèses sous-jacentes. La présentation comprendra un exposé des progrès relativement à ces techniques qui rendent la modélisation des données longitudinales plus flexible. Ces progrès comprennent notamment l'utilisation de méthodes fondées sur des courbes. Plusieurs exemples d'application des nouvelles méthodes plus flexibles tirés de publications sur la santé seront présentés afin de démontrer que ces méthodes permettent la résolution de certaines questions difficiles liées à l'analyse de données longitudinales complexes dans le contexte des études de la santé de la population.

**(A) Analyse de données familiales longitudinales sur la santé provenant d'enquêtes à plan de sondage complexe : une vue d'ensemble**

Brajendra Sutradhar, Memorial University of Newfoundland, St. John's, T.-N.-L., Canada

Dans le contexte des études sur la santé, il est assez fréquent de recueillir des réponses binaires ou numériques répétées au moyen d'un ensemble de covariables multidimensionnelles durant une courte période auprès d'un grand nombre de familles indépendantes. Ces familles sont sélectionnées dans une population finie, conformément à certains plans de sondage complexes. Il est intéressant d'examiner les effets des covariables sur les réponses familiales longitudinales après voir tenu compte de la variation des effets familiaux et des corrélations longitudinales des réponses répétées. Cette communication traitera des avantages et des inconvénients des méthodes existantes d'estimation des effets de régression, de la variance des effets familiaux et des corrélations longitudinales. On décrira ensuite les avantages d'une nouvelle méthode unifiée généralisée de quasi-vraisemblance pour analyser les données familiales longitudinales fondées sur un plan de sondage complexe. On analysera certaines études numériques existantes en vue d'illustrer les méthodes prises en considération dans la communication.

**(A) Estimation du nombre attribuable de décès et des erreurs-types d'après des cohortes échantillonnées selon des plans de sondage simples et complexes**

Barry I. Graubard<sup>1</sup>, Mitchell H. Gail<sup>1</sup>, Katherine M. Flegal<sup>2</sup>, David F. Williamson<sup>3,1</sup>, National Cancer Institute, Bethesda, MD,<sup>2</sup> National Center for Health Statistics, Hyattsville, MD,<sup>3</sup> Centers for Disease Control and Prevention, Atlanta, GA, États-Unis

On peut obtenir des estimations du nombre attribuable de décès (DA), toutes causes confondues, en commençant par estimer le risque attribuable corrigé (RA) au sein de la population pour tenir compte des covariables confusionnelles, puis en multipliant le RA par le nombre de décès, déterminé d'après les statistiques de l'état civil sur la mortalité, survenus pendant une période de référence précise. Nous obtenons par la méthode de régression à risques proportionnels des estimations des risques relatifs corrigés basées sur des données sur la mortalité provenant du suivi d'une cohorte ou d'une enquête que nous combinons à une distribution conjointe du facteur de risque et des covariables confusionnelles pour calculer un risque attribuable (RA) corrigé. Nous examinons deux estimateurs du RA corrigé, qui se distinguent par la population de référence d'après laquelle la distribution conjointe du facteur de risque et des variables confusionnelles est obtenue. Les deux populations de référence

considérées sont : i) la population qui est représentée par la cohorte de référence et : ii) une population extérieure à la cohorte. Nous appliquons des méthodes fondées sur la théorie de la fonction d'influence pour obtenir des expressions permettant d'estimer la variance de l'estimateur du nombre attribuable de décès (DA). Ces estimateurs de la variance peuvent être appliqués à des données recueillies auprès d'échantillons allant de l'échantillon aléatoire simple à l'échantillon en grappes stratifié à plusieurs degrés pondéré par les poids de sondage utilisés pour réaliser les enquêtes nationales auprès des ménages. Nous illustrons l'estimation de la variance du nombre attribuable de décès au moyen d'une analyse de la surmortalité liée à l'indice de masse corporelle non idéal dans une population réalisée sur des données provenant de la deuxième étude de la mortalité basée sur la National Health and Examination Survey (NHANES) et de la NHANES de 1999-2002. Ces méthodes peuvent également être utilisées pour estimer le nombre attribuable de décès par cause d'une maladie ou de nouveaux cas ainsi que leurs erreurs-types, lorsque la période de référence est brève.

## **Session 5 – Combinaison de données provenant de sources différentes**

### **(A) Études de qualité en vue d'évaluer le couplage de données aux données du Recensement de la population**

Glensy Bishop, Analytical Services Branch, ABS, Canberra, Australie

L'Australian Bureau of Statistics (ABS) entamera la création d'un fichier statistique de données longitudinales de recensement en choisissant un échantillon de 5 % des personnes qui composent la population du recensement de 2006 qui sera lié aux recensements subséquents. Les noms et les adresses ne seront pas conservés après la période de traitement du recensement. Cependant, au cours de cette période, nous effectuerons différentes opérations de couplage au moyen des noms et des adresses. Celles-ci ont été qualifiées d'études de la qualité.

Il y a deux types d'études de la qualité. Le premier a pour but d'évaluer la faisabilité et la qualité probable du couplage entre les fichiers de données lorsque les noms et les adresses ne sont plus disponibles. Le deuxième type consiste à un couplage entre les données du recensement et d'autres données afin d'améliorer les sorties statistiques de l'ABS. Cette communication portera sur le premier type et des méthodes d'étude de la qualité que nous avons utilisées seront présentées, ainsi que les résultats provisoires concernant les couplages réalisés avec les données de la répétition générale du recensement.

Un des objectifs à long terme est d'utiliser la puissance de la richesse des données démographiques longitudinales fournie par le fichier statistique de données longitudinales de recensement afin d'éclaircir divers points en faisant le couplage avec les registres des naissances, des décès, des maladies et des établissements pour immigrants. Les études initiales de la qualité détermineront dans quelle mesure nous pouvons effectuer ces couplages.

**(A) Appariement de données censitaires régionales avec des données d'enquête et des données administratives**

Denis Gonthier, Tina Hotton, Cynthia Cook et Russell Wilkins, Statistique Canada, Ottawa, Ontario, Canada

Dans cet article, nous expliquons comment joindre les données sommaires du recensement par région à des données d'enquête ou à des données administratives. Nous citons des exemples d'ensembles de données présents dans les Centres de données de recherche de Statistique Canada, mais les méthodes valent aussi pour des ensembles extérieurs, notamment les ensembles de données administratives. À l'aide de quatre exemples, nous illustrons des situations courantes dans lesquelles se trouvent les chercheurs : (1) cas où les données d'enquête (ou les données administratives) et les données du recensement contiennent des identificateurs géographiques de même niveau et sont codés selon la même année de référence du découpage géographique aux fins du recensement (par exemple, si les deux ont des données 2001 AD); (2) cas où les deux fichiers contiennent des identificateurs géographiques pour la même année de référence, mais pour des niveaux différents de découpage géographique du recensement (par exemple, 1996 SD dans l'enquête, mais 1996 SR dans les données du recensement); (3) cas où les deux fichiers contiennent des données codées pour des années de référence différentes de géographie telles que 1996 SD pour l'enquête, mais 2001 AD pour le recensement); (4) cas où les données d'enquête n'ont pas d'identificateurs géographiques, ceux-ci devant d'abord être produits à partir des codes postaux du fichier d'enquête. Ces exemples sont présentés en syntaxe SAS, mais les principes s'appliquent à d'autres langages de programmation ou progiciels statistiques.

**(A) Une étude de cas du recours à l'estimation assistée par modèle pour intégrer des données d'enquête et des données administratives**

Robert E. Fay, U.S. Census Bureau, Washington, DC, États-Unis

Statistique Canada est un chef de file de l'application des estimations basées sur un modèle aux statistiques publiques. Par exemple, l'introduction de cette démarche dans l'estimation de l'échantillon du

Recensement du Canada de 1991 a constitué une étape importante. Un certain nombre d'autres pays ont également utilisé ces méthodes qui gagnent également en popularité aux États-Unis.

Le présent document décrit la recherche en cours visant à intégrer l'estimation basée sur un modèle à l'American Community Survey (ACS), une grande enquête en cours devant remplacer les données du formulaire de recensement détaillé des recensements décennaux américains. Une des caractéristiques de l'application proposée est très certainement d'intérêt général. Il s'agit de la démarche selon laquelle on intègre à l'estimation des renseignements provenant de dossiers administratifs. Pour ce qui est de la qualité des données, l'utilisation des estimations basées sur un modèle permet au Census Bureau de mettre l'accent sur la qualité des données de l'ACS recueillies directement auprès des répondants, et ce, tout en ne comptant que sur les relations basées sur une régression avec les données administratives afin de réduire la variance. Ainsi, la qualité des données des dossiers administratifs ne doit être évaluée qu'en fonction de leur effet sur la variance. Bien que l'ACS soit une enquête générale qui n'est pas précisément liée à la santé, il est possible que cette étude de cas propose certaines applications possibles dans les secteurs de la statistique de la santé en vue de combiner les données d'enquête et les données administratives.

**(A) Couplage d'enregistrements exact et probabiliste**

Claude Nadeau, Marie P. Beudet, Jocelyne Marion, Christel Le Petit, Statistique Canada, Ottawa, Ontario, Canada

Le couplage d'enregistrements entre les bases de données peut servir à augmenter la gamme de mesures d'analyse, à permettre une vérification du niveau de correspondance entre deux mesures d'une construction et à inclure les personnes qui ne participent habituellement pas aux enquêtes en raison de contraintes de temps ou de problèmes de santé ou qui refusent d'y participer.

On présentera les résultats des deux études dans le cadre desquelles on a couplé les fichiers de données au moyen d'une méthode exacte et d'une méthode probabiliste. Une étude examine le couplage d'enregistrements entre un échantillon du Recensement de 1991, des enregistrements de l'Enquête sur la santé et les limitations d'activité (ESLA), des données fiscales sommaires (T1) de 1990 et de 1991 et des enregistrements de décès obtenus entre 1991 et 2001. L'étude vise à créer un ensemble d'indicateurs de base de la mortalité aux fins du suivi des inégalités relatives à la santé au Canada et des inégalités socioéconomiques en matière de santé. L'autre étude illustre le couplage de l'Enquête sur la santé dans les collectivités canadiennes de 2003 et de la Base de données de l'information sur la santé orientée vers la personne. Cette étude examine le rôle des facteurs de risques

modifiables en ce qui concerne la probabilité d'être hospitalisé en raison d'une maladie cardiaque.

Après un aperçu des règles du couplage et du nombre d'enregistrements couplés au moyen des différentes méthodes, on présentera une comparaison des caractéristiques des personnes figurant dans chaque base de données. Également, on examinera les résultats des analyses multivariées opposant les conclusions dans les cas où on a obtenu les enregistrements au moyen des méthodes de couplage exacte et probabiliste afin de décrire les effets d'une augmentation potentielle de la puissance de la base de données couplée et de l'étendue de son échantillon. On présentera également les limites liées à ces types de couplages.

### **Session 6 - Questions relatives à la pondération, à l'estimation et à la détection des observations aberrantes**

#### **(A) Modèles de sélection des variables pour l'élagage des coefficients de pondération**

Michael Elliott, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, États-Unis

Dans le cas de l'échantillonnage avec probabilités de sélection inégales, les corrélations entre les probabilités de sélection et les données échantillonnées peuvent induire un biais. Pour le corriger, on fait souvent appel à des poids de sondage correspondant à l'inverse de la probabilité de sélection. Les plans d'échantillonnage fortement disproportionnels donnent lieu à des poids élevés, susceptibles d'introduire une variabilité superflue dans des statistiques telles que l'estimation de la moyenne de la population. L'élagage des poids consiste à réduire ceux dont la valeur est élevée à une valeur seuil fixée et à rajuster ceux dont la valeur est inférieure à ce seuil afin que leur somme demeure égale à celle des poids non élagués. Cet élagage réduit la variabilité, mais introduit un certain biais. Les méthodes habituelles ne sont pas « guidées par les données », c'est-à-dire qu'elles ne se fondent pas sur les données pour arriver au compromis biais-variance approprié, ou bien elles le font de façon très inefficace. Cette communication décrit l'élaboration de méthodes bayésiennes de sélection des variables pour l'élagage des poids destinées à compléter les méthodes habituelles, ponctuelles, fondées sur le plan de sondage, lors de l'utilisation de plans à probabilités d'inclusion disproportionnelles où les variances dues aux poids de sondage excèdent la correction pour le biais. Ces méthodes sont utilisées pour estimer les paramètres de population provenant de modèles de régression linéaire et de régression linéaire généralisée dans le contexte de plans d'échantillonnage à probabilités de sélection connues stratifiés et stratifiés a posteriori. Des applications seront

décrites dans le cas de données provenant d'enquêtes sur les blessures causées par des accidents de circulation, pour lesquelles des plans de sondage fortement disproportionnels sont souvent utilisés.

**(A) Combinaison de renseignements provenant de deux enquêtes pour améliorer l'analyse des données autodéclarées utilisées pour estimer les mesures de la santé**

Nathaniel Schenker, Trivellore E. Raghunathan et Irina Bondarenko, National Center for Health Statistics, Hyattsville, MD, États-Unis

Malgré les progrès réalisés en vue d'améliorer la santé de la population des États-Unis, des disparités en matière de santé persistent entre divers groupes raciaux/ethniques et socioéconomiques. Les sources habituelles de données pour évaluer la santé d'une population d'intérêt comprennent les enquêtes à grande échelle, qui contiennent souvent des questions à réponse autodéclarée, comme « Un médecin ou un autre professionnel de la santé vous a-t-il déjà dit que vous souffrez de < problème de santé d'intérêt >? Les réponses à de telles questions pourraient ne pas refléter la prévalence réelle des problèmes de santé (p. ex., si le répondant n'a pas accès à un médecin ou à un autre professionnel de la santé). De même, les renseignements autodéclarés concernant des données quantitatives telles que la taille et le poids pourraient être entachés d'erreurs de déclaration. Ce genre d'« erreurs de mesure » dans les données sur la santé pourraient avoir une incidence sur les inférences relatives aux mesures de la santé et aux disparités en matière de santé. Dans la présente étude, nous ajustons des modèles d'erreur de mesure aux données de la National Health and Nutrition Examination Survey, qui comprend un volet d'interview durant lequel sont posées des questions à réponse autodéclarée et un volet d'examen durant lequel sont obtenues des mesures physiques. Puis, nous élaborons des méthodes permettant d'utiliser les modèles ajustés pour améliorer l'analyse des données autodéclarées provenant d'une autre enquête ne comportant pas de volet d'examen. Enfin, nous appliquons ces méthodes, qui comportent l'imputation multiples de valeurs fondées sur les données recueillies par examen à l'enquête ne produisant que des données autodéclarées, à la National Health Interview Survey dans des exemples ayant trait au diabète, à l'hypertension et à l'obésité. Les résultats préliminaires laissent entendre que les ajustements pour l'erreur de mesure peuvent causer des variations non négligeables des estimations des mesures de la santé.

**(A) Évaluation des méthodes de détection et de traitement des observations aberrantes dans la « Survey of Occupational Illnesses and Injuries » des États-Unis**

John L. Eltinge, Bureau of Labour Statistics, Washington, DC, États-Unis

La Survey of Occupational Illnesses and Injuries (SOII) des États-Unis est une enquête à grande échelle effectuée auprès des établissements et réalisée par le Bureau of Labor Statistics. Elle a pour but de mesurer les taux d'incidence et les conséquences des maladies et des blessures professionnelles dans certaines industries à l'échelle du pays et de l'État. À l'heure actuelle, cette enquête s'appuie sur des méthodes relativement simples pour la détection et le traitement des valeurs aberrantes. Les méthodes de détection des valeurs aberrantes reposent sur la comparaison des taux d'incidence déclarés en ce qui concerne l'établissement à la distribution correspondante des déclarations dans des cellules particulières définies par le croisement des classifications selon l'État et selon l'industrie. Les méthodes de traitement comportent le remplacement des poids probabilistes type par un poids dont la valeur est fixée à un, suivi par un étalonnage.

Des méthodes plus complexes pourraient être utilisées pour la détection et le traitement des valeurs aberrantes dans la SOII, par exemple des méthodes de détection qui s'appuient sur des fonctions d'influence, des poids probabilistes et des observations multivariées, ou des méthodes de traitement fondées sur la winsorisation ou l'estimation M. L'évaluation des avantages pratiques de ces méthodes plus complexes nécessite la prise en considération de trois facteurs importants. Premièrement, les valeurs très extrêmes sont relativement rares, mais lorsqu'elles se produisent, elles peuvent avoir un effet important sur les estimateurs de la SOII dans les cellules définies par le croisement des États et des industries. Par conséquent, l'évaluation pratique de l'effet des méthodes de détection des valeurs aberrantes se concentre principalement sur les queues des distributions des estimateurs, plutôt que sur les mesures de performance agrégées normalisées, comme la variance ou l'erreur quadratique moyenne. Deuxièmement, les évaluations analytiques et fondées sur des données sont axées sur l'amélioration progressive obtenue grâce à l'utilisation de méthodes plus complexes, comparativement aux résultats produits par les méthodes simples suivies à l'heure actuelle. Troisièmement, l'élaboration des outils susmentionnés nécessite le recours à une théorie asymptotique qui n'est pas tout à fait standard pour refléter les compromis en ce qui a trait aux effets associés à, respectivement, l'accroissement de la taille des échantillons, l'accroissement du nombre de cellules pour la publication et l'évolution des queues des distributions sous-jacentes des observations.

**(A) La combinaison de cycles de l'enquête sur la santé dans les collectivités canadiennes**

Steven Thomas, Statistique Canada

L'enquête sur la santé dans les collectivités canadiennes (ESCC) est une enquête par échantillon composée de deux enquêtes transversales menées sur un cycle annuel en alternance. Le premier cycle recueille les renseignements de plus de 130 000 répondants pour la production d'estimations sur la santé générale au niveau de la région sociosanitaire. Le deuxième cycle recueille les renseignements de plus de 30 000 répondants pour la production d'estimations sur un sujet relié à la santé spécifique à chaque enquête et ce, au niveau de la province. Même si les tailles des échantillons sont grandes, les utilisateurs sont intéressés à combiner les cycles de l'ESCC afin d'améliorer la qualité des estimations, de produire des estimations pour les petits domaines géographiques ou pour des populations rares. La méthodologie existe pour la combinaison d'enquêtes dans le but d'augmenter la taille de l'échantillon et donc la précision des estimations. Cependant, dans le contexte général, ces méthodes s'appliquent à la combinaison d'enquêtes qui mesurent la même chose. Hors, les cycles de l'ESCC ne mesurent pas exactement la même chose; sinon la répétition des cycles serait inutile. Par conséquent, la combinaison des cycles de l'ESCC ou de n'importe quelle série d'enquêtes transversales ne peut être exécutée aveuglément. Le chercheur doit avoir une compréhension claire de l'objectif de la combinaison des cycles de l'ESCC, ceci inclus une compréhension claire de la population couverte et du paramètre mesuré. Ce document mettra l'accent sur les enjeux reliés à la combinaison des cycles de l'ESCC incluant les interprétations possibles d'un résultat combiné. Les méthodes envisageables pour la combinaison des cycles seront exposées brièvement.

**Session 7– Enquêtes sur la santé utilisant les mesures directes : problèmes et possibilités****(A) L'apprentissage des défis uniques que posent les enquêtes utilisant les mesures directes de la santé : l'expérience canadienne**

Mark Tremblay, Statistique Canada, Ottawa, ON, Canada

L'Enquête canadienne sur les mesures de la santé (ECMS) est la première enquête d'importance portant sur la santé qui utilise des mesures physiques directes réalisée par Statistique Canada. L'ECMS, qui se déroulera en 2007-2008, permettra de recueillir des données sur la santé mesurées directement auprès d'un échantillon représentatif de la population composé de 5 000 Canadiens de 6 à 79 ans. Après une interview détaillée portant sur la santé et faite à domicile, les

répondants se présenteront à un centre d'examen mobile (CEM; semi-remorques d'une longueur de 53 pieds) où seront prises des mesures physiques directes. Le CEM sera installé pendant six semaines à chacun des 15 emplacements sélectionnés dans les diverses régions du pays. À chaque emplacement, 30 employés de terrain et intervieweurs, qui se déplaceront avec le CEM, prendront des mesures auprès d'environ 350 répondants. Les mesures comprendront des tests d'évaluation de la condition physique, des mesures anthropométriques, une évaluation objective de l'activité physique, des mesures spirométriques, des mesures de la pression artérielle, des mesures de la santé buccodentaire ainsi que des prélèvements d'échantillons de sang et d'urine qui seront analysés. Les analyses de sang et d'urine comporteront le dépistage des marqueurs de maladies chroniques, de maladies infectieuses et de l'état nutritionnel ainsi que des biomarqueurs d'exposition aux contaminants environnementaux. L'enquête pose un grand nombre de défis particuliers, rarement observés, dans le cadre de la plupart des autres enquêtes réalisées par Statistique Canada. Ces défis sont liés à un nouveau niveau de fardeau de réponse (déplacements, temps, frais, effort physique, douleur), à la complexité du transfert des données (intervieweur-clinique-laboratoire-Statistique Canada-rapport à l'intention du répondant), aux considérations relatives à la protection des renseignements personnels et à l'éthique (tranche d'âge pour les évaluations, consentements, confidentialité des données, anonymat des participants, entreposage des prélèvements biologiques en vue de futures analyses, collecte d'échantillons d'ADN), aux questions de communication (production d'un rapport à l'intention des répondants, maladies à déclaration obligatoire, médias, public) et à la possibilité d'événements indésirables (phlébite, accident cardiaque durant l'évaluation de la condition physique) ou de résultats indésirables (maladie infectieuse non diagnostiquée antérieurement, cancer de la bouche). Malgré ces défis nombreux, les données recueillies grâce à l'ECMS seront uniques et s'avèreront une ressource précieuse tant pour la surveillance de la santé que pour la recherche au Canada.

**(A) La « National Health and Nutrition Examination Survey » des États-Unis : Ce que 40 années d'expérience nous ont appris**  
Cliff Johnson, National Centre for Health Statistics, Hyattsville, MD, États-Unis

Depuis plus de 40 ans, la National Health and Nutrition Examination Survey (NHANES) est réalisée par le National Center for Health Statistics (qui fait partie des Centers for Disease Control and Prevention (CDC) des États-Unis). Cette enquête est unique, car elle est conçue pour mesurer l'état de santé et l'état nutritionnel de la population des États-Unis (sauf les pensionnaires d'établissements institutionnels) au moyen d'interviews et d'examens physiques directs.

L'expérience de ces quelque 40 années nous a permis de tirer de nombreuses leçons sur toute une gamme de questions méthodologiques relatives au plan de sondage, au contenu du questionnaire, aux opérations d'enquête, à la qualité des données et à l'analyse de ces dernières. Nous présenterons des exemples d'« apprentissage » pour chacun de ces domaines. Ces exemples toucheront : 1) à l'utilisation de centres d'examen mobiles par opposition à des « emplacements fixes » pour la collecte des données; 2) aux mesures qui doivent être prises avant de passer à une nouvelle méthodologie de collecte des données pour une composante d'une enquête en cours; 3) à la nécessité de décrire (pour la postérité) tout changement apporté au plan de sondage et aux processus de collecte des données; 4) au lien entre la qualité des données, la formation complète et répétée du personnel et la réalisation d'un essai pilote des nouvelles méthodes et 5) à la nécessité d'être souple et prêt à accepter le changement (ce n'est pas parce qu'on réalise l'enquête d'une certaine façon depuis 40 ans qu'il n'existe pas de meilleur moyen de le faire). Les défis ont été nombreux, des erreurs ont été (et seront) commises et on continuera de tirer d'autres leçons de l'expérience. Quoi qu'il en soit, l'importance du recours à des mesures physiques directes de la santé comme outils de surveillance et de recherche justifie largement les efforts qui ont été et qui seront déployés pour surmonter les défis méthodologiques.

**(A) Enquêtes utilisant les mesures directes de la santé en Finlande : l'expérience de la Carélie du Nord et au-delà**

Arpo Aromaa, KTL, Helsinki, Finlande

Les enquêtes nationales sur la santé par examen (HES, Health Examination Surveys) ont été lancées en Finlande par l'Unité clinique mobile (UCM) de l'Institut d'assurance sociale (IAS) en 1965. Réalisées dans les diverses régions du pays auprès d'environ 60 000 adultes, ces études comportaient diverses mesures, ainsi que des interviews sur le régime alimentaire. Une version plus restreinte de l'enquête a été réalisée dans le cadre du projet touchant la Carélie du Nord en 1972. Depuis, cette enquête est répétée tous les cinq ans.

Une enquête globale de portée nationale, appelée Mini-Finlande, a été réalisée de 1978 à 1980 par l'IAS. L'UCM s'est penchée sur plus de 7 200 personnes (parmi 8 000) sélectionnées dans 40 régions afin qu'elles soient représentatives de l'ensemble de la population. On a réalisé une large gamme d'interviews et de mesures, ainsi que des examens cliniques effectués par des dentistes et des médecins. On s'intéressait principalement aux maladies chroniques et à la capacité fonctionnelle. La population a été suivie par divers registres.

Coordonnée par KTL, Health 2000, une nouvelle enquête globale, a été réalisée en 2000 et en 2001 dans 80 régions représentatives de la Finlande. L'échantillon comprenait 10 000 personnes de 18 ans et plus. On a limité les examens de santé aux 8 028 personnes de 30 ans et plus. Les objectifs principaux étaient d'examiner l'état de santé et la capacité fonctionnelle, leurs déterminants et leur répartition, et d'évaluer les changements survenus après l'étude réalisée de 1978 à 1980. Après une interview approfondie, on a demandé aux participants de remplir un long questionnaire et de l'apporter lors de la séance d'examen physique. Le programme de base comprenait des mesures anthropométriques et de la pression artérielle, des mesures de la bio-impédance, le prélèvement d'échantillons de sang (plasma et sang total pour l'ADN) et d'échantillons de salive, une radiographie numérique des dents et un examen par un dentiste, diverses épreuves fonctionnelles (physiques et cognitives), un examen clinique structuré par un médecin et, enfin, une interview psychiatrique (M-CIDI). Le fonctionnement a été évalué au moyen de questionnaires, d'observations et de tests informatisés.

Il est possible de consulter un grand nombre de nos méthodes et de nos résultats sur notre site Web trilingue ([www.ktl.fi/health2000](http://www.ktl.fi/health2000)). La prochaine enquête semblable sera probablement réalisée dans cinq ans et appelée Santé 2012 et constitue un élément du système global d'enquêtes sur la santé de la Finlande.

## Session 8 – Discours du Prix Waksberg

### **(A) L'analyse des études avec témoins représentatives de la population**

Alastair Scott, University of Auckland, Auckland, Nouvelle-Zélande

L'utilisation de plans d'échantillonnage complexes est de plus en plus répandue dans les études de cas-témoins représentatifs de la population, habituellement afin d'obtenir les témoins, mais parfois les cas également. Le sujet est particulièrement pertinent à une session qui rend hommage à Joe Waksberg, car c'est grâce à sa persévérance devant l'importance de l'utilisation de témoins réellement représentatifs que l'on a adopté des normes en matière de méthodologie d'enquête de façon généralisée dans ce domaine.

Aujourd'hui, la plupart des progiciels statistiques contiennent des modules spécialistes pour les données d'enquête qui permettent d'exécuter couramment des analyses pondérées sur des données provenant de ce genre d'études. Toutefois, la pondération a tendance à être inefficace dans des situations telles que celle-ci, où les écarts entre les poids sont habituellement très grands. Des méthodes de

vraisemblance complètement efficaces peuvent être élaborées pour certains plans d'échantillonnage spéciaux. Toutefois, à l'exception du cas de l'échantillonnage stratifié simple, ces méthodes nécessitent un logiciel spécial et sont difficiles à mettre en œuvre. En outre, on a émis des doutes quant à leur robustesse à la défaillance du modèle. Nous examinons la question de la robustesse de façon plus approfondie et proposons des méthodes de rechange qui sont raisonnablement efficaces et qui peuvent être appliquées au moyen des logiciels standard d'analyse des données d'enquête.

## **Session 9 – Méthodes d'analyse des données sur la santé**

### **(A) Méthodes bootstrap pour l'analyse de données d'enquêtes à plan de sondage complexe**

J. N. K. Rao, Université Carleton, Ottawa, ON, Canada

Les enquêtes à plan de sondage complexe comportent souvent un échantillonnage en grappes à plusieurs degrés qui donne lieu à des dépendances entre les éléments échantillonnés, ainsi qu'à l'utilisation de probabilités de sélection inégales produisant des poids d'échantillonnage inégaux. En outre, les poids d'échantillonnage sont souvent calés sur des totaux de population connus pour des variables auxiliaires et ajustés pour tenir compte de la non-réponse. Par conséquent, l'application des méthodes classiques aux données provenant de ces enquêtes en omettant de tenir compte des caractéristiques du plan et des ajustements de la pondération peut aboutir à des inférences erronées. Les méthodes bootstrap constituent une option intéressante pour l'analyste qui veut tenir compte du plan de sondage. Le fichier de données comprend les poids de sondage finals pour l'échantillon complet et les poids bootstrap finals connexes pour un grand nombre de répliques bootstrap, ainsi que les données observées sur les unités de l'échantillon. Nous montrons comment ce genre de fichier peut être utilisé pour analyser les données d'enquête de façon simple en utilisant des équations d'estimation pondérées. Nous discutons aussi d'une méthode bootstrap à fonction d'estimation (FE) en une étape qui permet d'éviter certaines difficultés que pose le bootstrap. Nous montrons comment la méthode peut être appliquée à des données de réponses binaires transversales et longitudinales en utilisant des modèles de régression logistique linéaire. Nous présentons également des tests formels de la qualité de l'ajustement des modèles tenant compte du plan de sondage. Nous utilisons des données provenant de l'Enquête nationale sur la santé de la population du Canada pour illustrer les méthodes proposées.

**(A) Estimation des paramètres de régression au moyen de données d'enquête**

Wayne A. Fuller et Yu Wu, Iowa State University, IA, États-Unis

Les coefficients des équations de régression s'avèrent souvent des paramètres d'intérêt dans le cas des enquêtes sur la santé. Ces dernières sont parfois réalisées selon un plan de sondage complexe avec taux d'échantillonnage différentiels. Nous présentons des estimateurs des coefficients de régression applicables aux enquêtes complexes qui sont supérieurs aux estimateurs à facteur d'extension ordinaires à l'égard du modèle en question, mais qui retiennent également les propriétés souhaitables du plan. Nous présentons les priorités théoriques ainsi que celles de la méthode Monte Carlo.

**(A) Tenir compte de la dispersion des données en recherche contextuelle sur la santé de la population : effets de la petite taille des groupes et de l'analyse par grappes ciblée sur les résultats, le biais et l'efficacité des modèles multiniveaux**Philippa Clarke<sup>1</sup>, Patricia O'Campo<sup>2</sup> and Blair Wheaton<sup>2, 1</sup> University of Michigan, Ann Arbor, MI, USA,<sup>2</sup> University of Toronto, Toronto, ON, Canada

L'usage courant des modèles multiniveaux pour examiner les effets du contexte environnant (quartier) sur les résultats en matière de santé témoigne de leur valeur en tant que méthode statistique d'analyse de données groupées. Cependant, l'application de la modélisation multiniveaux à des données provenant d'enquêtes à l'échelle de la population est souvent limitée par le petit nombre de cas par unité de niveau 2, si bien que l'on relève dans les travaux antérieurs portant sur les effets du quartier une tendance récente à appliquer des méthodes d'analyse par grappes pour contourner le problème de la dispersion des données. Dans la présente communication, nous utilisons des simulations de Monte Carlo pour étudier les effets de tailles marginales de groupe sur les résultats, le biais et l'efficacité des modèles multiniveaux. Puis, nous appliquons des techniques d'analyse par grappes afin de minimiser la dispersion des données et examinons les conséquences dans les simulations. Nous déterminons la borne minimale à partir de laquelle la dispersion des données devient problématique si l'on veut produire des estimations valides au moyen de modèles contextuels. Nous constatons que les estimations des effets fixes sont robustes aux limites de dispersion des données et que l'analyse par grappes est une stratégie efficace en vue d'accroître la taille de groupe et d'éviter le biais vers le haut dans les composantes de la variance qui se manifeste lorsque s'accroît la dispersion des données. Cependant, les stratégies de mise en grappes introduisent une hétérogénéité intragroupe artificielle qui se traduit par des effets de groupe faussement faibles (erreur de type II) causés par une

minimisation des différences intergroupes. Par conséquent, il faut veiller, lors de l'utilisation des méthodes d'analyse par grappes, à maintenir l'équilibre entre la variance intragroupe et la variance intergroupes dans les données.

**(A) Inférence causale dans les études par observation utilisant des données administratives sur la santé**

Therese A Stukel, ICES, Toronto, ON, Canada

La recherche sur les services de santé s'appuie habituellement sur des données d'observation afin de comparer les résultats chez des patients recevant des traitements différents. La comparaison de groupes de patients participant à des études par observation peut être biaisée, car les résultats diffèrent à la fois en raison des effets du traitement et de ceux liés au pronostic fait sur le patient. Dans certains cas, particulièrement lorsque les données recueillies ont trait à des facteurs de risque cliniques précis, il est possible de tenir compte de ces différences en utilisant des méthodes statistiques ou épidémiologiques. Dans d'autres cas, où des caractéristiques non mesurées de la population de patients ont une incidence sur la décision de prodiguer le traitement ainsi que sur le résultat, il est impossible d'éliminer ces différences par des techniques courantes. L'utilisation de données administratives sur la santé pour entreprendre des études par observation demande des précautions particulières, puisque d'importants renseignements cliniques font défaut. Nous discutons de plusieurs méthodes statistiques et épidémiologiques visant à éliminer le biais manifeste (mesurable) et caché (non mesurable) dans les études par observation. Ces méthodes comprennent l'ajustement de la composition des groupes de cas, l'appariement fondé sur la propension, la redéfinition des variables d'exposition d'intérêt et la technique économétrique d'analyse avec variables instrumentales (VI). Nous illustrons ces méthodes à l'aide d'exemples extraits de publications médicales, dont la prédiction de la mortalité un an après une crise cardiaque, le rendement des dépenses en soins de santé, en ce qui a trait aux avantages cliniques et financiers, dans les régions des États-Unis où les dépenses sont élevées, ainsi que les avantages du traitement effectif des patients ayant fait une crise cardiaque liés à la survie à long terme. Il est possible d'utiliser des données administratives sur la santé pour réaliser des études par observation à condition de veiller soigneusement à résoudre les problèmes liés à la causalité inverse et au facteur de confusion non mesuré.

**Session 10– Effets du mode et collecte des données****(A) Passage à la téléphonie sans fil aux États-Unis et au Canada : prévalence et effet sur les enquêtes téléphoniques sur la santé utilisant la composition aléatoire de numéros de téléphone**

Stephen J. Blumberg et Julian V. Luke, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD, États-Unis

En 2004, on comptait près de 15 millions de téléphones cellulaires au Canada; donc presque un téléphone cellulaire pour deux personnes. Aux États-Unis, plus de la moitié de tous les téléphones sont maintenant des téléphones cellulaires et le nombre moyen de minutes d'utilisation par mois de ceux-ci est supérieur au nombre moyen de minutes d'utilisation des téléphones à fil résidentiels par personne, par mois. Il n'est peut-être pas surprenant que bon nombre d'utilisateurs de téléphones cellulaires des deux pays aient songé à « couper le fil » et à remplacer leur téléphone conventionnel par un téléphone cellulaire.

La plupart des organismes de recherche par enquête d'importance au Canada et aux États-Unis n'utilisent pas les numéros de téléphone cellulaire lorsqu'ils communiquent de façon aléatoire avec les ménages pour effectuer des enquêtes téléphoniques. L'incapacité de joindre des adultes qui ne possèdent que des téléphones cellulaires (ou qui ne sont pas abonnés à un service téléphonique) risque d'avoir des répercussions sur les recherches en santé menées par l'entremise d'enquêtes par composition aléatoire.

La présentation mettra d'abord l'accent sur les estimations les plus à jour obtenues auprès de Statistique Canada et du National Center for Health Statistics des États-Unis relativement à la taille et aux caractéristiques de la population ne possédant que des téléphones cellulaires. Des estimations pour le Canada pourront être tirées de l'Enquête sur le service téléphonique résidentiel qui est un complément de l'Enquête sur la population active. Les estimations pour les États-Unis seront obtenues du National Health Interview Survey.

Ensuite, pour mieux comprendre les répercussions du passage aux téléphones cellulaires sur les volets ménages des enquêtes sur la santé, les auteurs mettront l'accent sur les estimations du National Health Interview Survey sur la santé et l'accès aux soins de santé des adultes qui n'utilisent que des téléphones cellulaires. Finalement, nous discuterons des différences démographiques entre les adultes qui n'utilisent que des téléphones cellulaires et ceux utilisant des téléphones conventionnels pour évaluer si elles peuvent expliquer en grande partie les différences observées relativement à la santé et à l'accès aux soins de santé.

**(A) Étude de l'effet du mode de collecte sur les estimations clés de l'état de santé dans la « National Health Interview Survey »**

Catherine Simile, Barbara Stussman, James Dahlhamer, National Center for Health Statistics, Hyattsville, MD, États-Unis

Afin d'augmenter le taux de réponse et de réduire les coûts, un grand nombre d'établissements d'enquête ont commencé à utiliser différentes méthodes pour recueillir des données pertinentes. Bien que dans le cadre du National Health Interview Survey (NHIS), une enquête polyvalente sur la santé des ménages menée annuellement par le National Center for Health Statistics, les intervieweurs comptent parfois sur le téléphone pour effectuer leurs interviews, les enquêtes menées par le Centers for Disease Control and Prevention des États-Unis sont principalement des interviews sur place. Cette situation soulève des questions quant à la comparabilité des données obtenues. Par exemple, les questions d'un questionnaire pour lesquelles des fiches sont généralement nécessaires, dans le cadre d'interviews sur place, génèrent-elles des réponses différentes lorsqu'elles sont adaptées afin d'être utilisées au téléphone? Bon nombre d'études ont aussi révélé que le fait d'utiliser une méthode différente pouvait générer des résultats différents pour des questions considérées comme délicates. Dans le cadre du NHIS, les estimations pour des sujets dits délicats sont-elles différentes selon que ces questions sont recueillies par téléphone ou sur place?

En vue d'aborder ces questions, des analyses sont effectuées sur l'historique des tentatives de contact infructueuses et les données de base provenant du NHIS 2005. Nous comparons les estimations obtenues à partir de deux types précis de questions : 1) les questions qui sont propices aux interviews sur place, mais qui peuvent s'avérer problématiques au téléphone (p. ex. celles qui comprennent des fiches) et 2) les questions considérées délicates (p. ex. le revenu, les maladies transmissibles sexuellement). Les répercussions de nos résultats sur les opérations courantes sur le terrain et sur la qualité des estimations du NHIS seront traitées.

**(A) Effet de l'interview téléphonique comparativement à l'interview sur place pour le rappel alimentaire de 24 heures dans les enquêtes sur l'alimentation et la nutrition**

S. Hayward, E. Junkins, M. Vigneault, M. Villeneuve, Direction des aliments, Santé Canada, Ottawa, ON, Canada

Beaucoup d'enquêtes sur la population qui permettent de recueillir des données sur la consommation d'aliments utilisent des formulaires de rappel alimentaire de 24 heures pour consigner de façon détaillée la quantité d'aliments ingérés dans une journée. Afin d'estimer l'ingestion d'aliments et de nutriments à plus long terme à partir de ces données,

on a élaboré des méthodes qui requièrent une deuxième interview de rappel auprès d'un sous-ensemble de répondants, et ce, afin d'estimer la variabilité d'un jour à l'autre. Durant l'Enquête thématique sur la nutrition comprise dans l'Enquête sur la santé dans les collectivités canadiennes (cycle 2.2), la plupart des premières interviews ont été effectuées en personne et les interviews de rappel, au téléphone. Dans le cadre de la conception de l'enquête, les interviews de rappel ont été réalisées en personne pour un sous-ensemble de répondants afin de permettre l'évaluation des effets de la méthode d'interview. Le présent article examine les effets de la méthode utilisée pour les interviews de rappel sur les aliments et les nutriments ingérés déclarés le jour de l'interview et l'estimation de la variabilité chez une même personne entre la première et la deuxième interview.

**(F) Comment tenir compte d'un changement méthodologique important dans l'analyse des données de l'Enquête sur la santé dans les collectivités canadiennes?**

Luc Côté et Robert Courtemanche, Institut de la statistique du Québec, Québec, QC, Canada

L'Enquête sur la santé dans les collectivités canadiennes (ESCC) est une enquête générale sur la santé, couvrant environ 130 régions sociosanitaires du Canada. En 2000-2001, Statistique Canada réalisait le cycle 1.1 de l'ESCC. Celle-ci était reconduite en 2003 (cycle 2.1).

L'ESCC utilise deux types de bases de sondage : l'un est aréolaire et l'autre est téléphonique. Au cycle 1.1, environ 95 % de l'échantillon pondéré du Québec provenait de la base aréolaire. Au cycle 2.1, cette proportion était inférieure à 50 %. Une collecte en face-à-face est principalement utilisée pour l'échantillon provenant de la base aréolaire, alors que le mode de collecte de la base téléphonique est associé uniquement à une collecte téléphonique. Entre ces deux cycles, la proportion de l'échantillon provenant de la base téléphonique a augmenté.

Les résultats de l'étude de St-Pierre et Béland (congrès de l'ASA, 2004) sur l'effet des modes de collecte indiquent des différences touchant plusieurs variables de l'enquête, faisant ainsi état de problèmes de comparabilité des données entre ces deux cycles. Il est alors recommandé d'être prudent dans l'utilisation des données. La question de la comparabilité a donc surgi rapidement de la part des utilisateurs québécois : Peut-on ou non comparer les données des deux cycles? Quelles variables peut-on comparer?

Pour répondre aux questions des utilisateurs, et non pour trouver la cause des différences entre les données de ces deux cycles, l'Institut

de la statistique du Québec (ISQ) s'est penché sur la principale modification méthodologique apportée à la dernière enquête, soit le changement dans la répartition de l'échantillon selon la base de sondage. En ce qui concerne certaines variables, ce changement méthodologique compromet la comparabilité entre les deux cycles. L'ISQ a produit des lignes directrices pour que les utilisateurs québécois puissent utiliser les mêmes méthodes d'analyse, selon que la variable est ou n'est pas comparable.

Cette communication exposera la méthodologie retenue et les résultats obtenus.

## **Session 11 – Échantillonnage — théorie et applications**

### **(A) Échantillonnage PPT systématique avec substitution d'unités basé sur la simulation**

Mary E. Thompson et Changbao Wu, University of Waterloo, Waterloo, ON, Canada

L'enquête réalisée par la Chine, dans le cadre du programme de lutte internationale contre le tabagisme (ITC, *International Tobacco Control*), comprend un plan d'échantillonnage à plusieurs degrés avec probabilités inégales où les grappes du niveau supérieur sont sélectionnées par la méthode d'échantillonnage PPT systématique. Au cours de l'exécution de l'enquête, il faut résoudre le problème suivant : plusieurs grappes de niveau supérieur sélectionnées refusent de participer et doivent être remplacées par des unités de substitution, sélectionnées à partir des unités restantes, en utilisant de nouveau une méthode d'échantillonnage PPT systématique. Selon de telles conditions (en supposant que le refus soit aléatoire), les probabilités d'inclusion de premier ordre des unités finales sélectionnées sont très difficiles à calculer et la détermination des probabilités d'inclusion de deuxième ordre devient virtuellement impossible. Nous élaborons une méthode fondée sur la simulation pour calculer les probabilités d'inclusion de premier et de deuxième ordre, et appliquons la méthode au moyen du progiciel statistique R/SPLUS. Nous démontrons l'efficacité et l'efficacité de la méthode que nous proposons au moyen d'exemples numériques. La méthode peut être étendue à des situations de refus plus complexes susceptibles de survenir en pratique.

**(A) Réutilisation des grappes dans les études répétées**

Stanislav Kolenikov<sup>1</sup> et Gustavo Angeles<sup>2</sup>, <sup>1</sup> University of Missouri, Columbia, MS, <sup>2</sup> University of North Carolina, Carolina Population Center, Chapel Hill, NC, États-Unis

Supposons que les données d'une enquête avec plan de sondage à plusieurs degrés doivent être recueillies pour deux périodes de référence. La présente communication décrit l'évaluation des mérites relatifs de la méthode qui consiste à garder les mêmes grappes dans l'échantillon en comparaison avec celles qui consistent à échantillonner de nouvelles grappes, selon divers scénarios statistiques (corrélation entre les grappes et au cours du temps) et logistiques (coûts de l'enquête). L'effet du plan lié à la réutilisation au cours du temps des mêmes grappes tirées de l'échantillon principal est de la forme  $1 - A\rho\pi/n$ , où  $\rho$  est la corrélation intertemporelle des totaux des grappes,  $n$  est le nombre de grappes,  $\pi$  est la proportion de grappes du cycle précédent qui sont retenues, et  $A > 0$  est une constante fixe. Tant que les gains d'efficacité paraissent mineurs, la valeur des plans comportant la réutilisation des grappes provient de considérations logistiques (coûts de l'enquête). Nous présenterons des démonstrations empiriques au moyen de données provenant de l'Enquête sur la démographie et la santé au Bangladesh réalisée en 1996 et en 2000 ainsi qu'une simulation de Monte Carlo.

**(F) Plan d'échantillonnage de l'Enquête canadienne sur les mesures de la santé**

Suzelle Giroux, Statistique Canada, Ottawa, ON, Canada

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, les facteurs de risque pour les conditions chroniques ainsi que l'exposition aux agents environnementaux polluants et transmissibles.

Afin d'atteindre ces objectifs ambitieux, un échantillon représentatif à l'échelle nationale composé d'environ 5 000 Canadiens, hommes et femmes âgés de 6 à 79 ans, sera invité à participer à l'ECMS. Les répondants seront priés de remplir un questionnaire à la maison, mais aussi de se présenter aux cliniques de l'ECMS, où des professionnels 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. Le début de la collecte est prévu en février 2007 et s'effectuera sur une période de 2 ans.

La présentation se concentrera sur le plan d'échantillonnage élaboré pour répondre aux objectifs et à la logistique de l'enquête. 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.

**(A) Les défis posés par la conception de la « National Health and Nutrition Examination Survey »**

Leyla Mohadjer, Westat, Washington, DC, États-Unis

La National Health and Nutrition Examination Surveys (NHANES) fait partie d'une série de programmes liés à la santé parrainés par le National Center for Health Statistics des États-Unis. Ce qui caractérise la NHANES est que tous les répondants de l'échantillon sont soumis à un examen médical complet. Afin de normaliser l'administration des examens médicaux, ceux-ci se déroulent dans des centres d'examen mobiles (CEM). L'examen comprend la prise de mesures physiques, des tests comme des examens de la vue et des dents ainsi que le prélèvement d'échantillons de sang et d'urine pour l'exécution d'analyses biologiques.

La NHANES est une enquête annuelle continue sur la santé réalisée auprès de la population civile des États-Unis ne résidant pas en établissement. Les objectifs analytiques comprennent l'estimation du nombre et du pourcentage de personnes, dans la population des États-Unis et dans des sous-groupes désignés qui présentent certaines maladies et certains facteurs de risque. Aux fins de l'enquête, on effectue la surveillance des tendances de certaines maladies, de la sensibilisation à ces dernières, aux traitements et à leur contrôle, la surveillance des tendances concernant les comportements à risque et les expositions environnementales ainsi que l'analyse des facteurs de risque pour certaines maladies. L'étude de la relation entre le régime alimentaire, la nutrition et la santé, et l'exploration des nouvelles questions de santé publique et des nouvelles technologies s'avèrent également des aspects importants de l'enquête.

Le plan d'échantillonnage de la NHANES doit permettre d'établir un juste équilibre entre les exigences liées à l'obtention d'échantillons annuels et pluriannuels efficaces et la souplesse requise pour pouvoir modifier les paramètres essentiels du plan afin de rendre l'enquête plus adaptée aux besoins des chercheurs et des décideurs qui élaborent les politiques en matière de santé. La présente communication décrit les

défis associés à la conception et à la mise en œuvre d'un processus d'échantillonnage permettant d'atteindre les objectifs de la NHANES.

## **Session 12 – Mesure de la santé : une perspective internationale**

**(A) La mesure de la santé des populations : l'approche analytique et conceptuelle de l'étude « Global Burden of Disease »**  
Colin D. Mathers, Department of Measurement and Health Information, Organisation mondiale de la santé, Genève, Suisse

Au cours des deux dernières décennies, d'importants efforts ont été déployés sur la scène internationale en vue d'élaborer des mesures sommaires de la santé des populations intégrant l'information sur la mortalité et sur les problèmes de santé non mortels. L'intérêt des spécialistes internationaux des politiques pour ce genre d'indicateurs augmente. Il existe deux grandes classes de mesures sommaires de la santé des populations : les différences en matière de santé et les attentes en matière de santé. L'année de vie corrigées de l'incapacité (AVCI) représente la mesure des différences en matière de santé la mieux connue et quantifie l'écart entre la santé réelle d'une population et un objectif de santé normatif, défini en fonction d'une table de survie normalisée mondiale précisant le nombre d'années de vie en bonne santé perdues à la suite d'un décès survenu à un âge donné.

Cette communication donnera un aperçu du cadre conceptuel du fardeau global de la maladie (en anglais, Global Burden of Disease (GBD)), de la relation entre l'AVCI et d'autres mesures de la santé de la population, et de la démarche analytique relative au fardeau global de la maladie, en accordant une attention particulière aux problèmes que posent 1) le traitement de données biaisées ou manquantes, 2) le traitement de l'incertitude et 3) la résolution des questions techniques spécifiques dont il faut tenir compte pour assurer la comparabilité entre les populations. Ce dernier point englobe la prise en compte des variations de la qualité et de l'intégralité de l'information sur la cause du décès, l'usage explicite d'un cadre général et de contrôles de la cohérence interne en vue d'améliorer la comparabilité des estimations de l'incidence, de la prévalence et de la mortalité par cause, l'évaluation de la pondération de l'incapacité et des techniques visant à rendre plus comparables les évaluations du fardeau de la maladie imputable aux facteurs de risque.

**(A) Comparaisons internationales dans la mesure des états de santé : expériences provenant des enquêtes mondiales sur la santé**

Somnath Chatterji, Department of Measurement and Health Information, Organisation mondiale de la santé, Genève, Suisse

Au cours des trois dernières décennies, on a généralement reconnu la démarche consistant à décrire les états de santé des individus en fonction de multiples aspects de la santé, et à élaborer des questionnaires d'auto-évaluation visant à recueillir des renseignements sur chacun de ces aspects. Un état de santé peut donc être considéré comme un attribut pluridimensionnel qui reflète le niveau de la personne pour diverses composantes, ou aspects, de la santé. Donc, l'état de santé est un concept qui diffère de l'état pathologique, des facteurs de risque, de l'étiologie, ainsi que des contacts avec les services de santé ou des interventions de ces derniers.

Choisir la façon de décrire les états de santé est une question essentielle dans le contexte de la mesure de la santé. La relation entre les états de santé et d'autres aspects de la santé, comme de futurs problèmes de santé non mortels ou le risque de mortalité, doit être examinée. La façon dont les gens déclarent leur propre état de santé varie en fonction du niveau de scolarité, du sexe, de l'âge ou d'autres facteurs culturels. Les valeurs seuils dans les catégories de réponse varient selon la culture ou le sous-groupe de population, et ce « déplacement de la réponse » indique que les données catégoriques autodéclarées ne sont pas comparables d'un individu à l'autre. Les réponses ne peuvent pas être utilisées directement, sans correction, pour mesurer la santé.

Compte tenu de ce qui précède, l'Organisation mondiale de la santé (OMS) a utilisé une série de questions pour les enquêtes mondiales sur la santé (EMS), qui couvraient un ensemble de domaines de base en vue de mesurer les états de santé, ainsi que des vignettes pour dépister et corriger les biais dans les réponses autodéclarées, afin d'apporter une correction pour tenir compte des déplacements des valeurs seuils dans les catégories de réponse. Cette communication décrira l'instrument utilisé pour les EMS et les méthodes appliquées pour produire des données comparables entre les populations. On présentera les résultats des EMS démontrant l'existence de biais de déclaration systématiques, la capacité des répondants à évaluer les vignettes et l'utilisation de ces dernières pour corriger les biais afin de rendre les données plus comparables. Les futures stratégies en vue de régler ces problèmes seront également exposées.

**(A) La stratégie de recherche pour développer des mesures comparables d'un statut de santé fonctionnel à utiliser pour les enquêtes sur la population**

Chris Murray, Harvard University, États-Unis

**Session 13 - Confidentialité et contrôle de la divulgation**

**(A) Application des méthodes de contrôle de la divulgation statistique à la base de données du Système canadien hospitalier d'information et de recherche en prévention des traumatismes**

Ann Brown<sup>1</sup> et Margaret Herbert<sup>2</sup>, <sup>1</sup> Statistique Canada, <sup>2</sup> Agence de santé publique du Canada, Ottawa, Ontario, Canada

Dans le présent document, nous décrivons les méthodes d'élaboration d'un fichier de microdonnées qui convient à la diffusion publique de la base de données du Système canadien hospitalier d'information et de recherche en prévention des traumatismes (SCHIRPT). Le SCHIRPT est une base de données nationale de surveillance des traumatismes gérée par l'Agence de santé publique du Canada (ASPC) qui vise à réduire le nombre et la gravité des traumatismes au Canada. Le SCHIRPT passe des contrats avec dix hôpitaux pédiatriques et quatre hôpitaux généraux dans les principales villes de sept provinces et d'un territoire pour recueillir des renseignements relatifs aux traumatismes fournis par les patients qui font un rapport à leur service d'urgence. Le SCHIRPT contient des renseignements détaillés sur les caractéristiques des patients et les circonstances où les accidents sont survenus. Le SCHIRPT est la première base de données de l'ASPC à élaborer un fichier de microdonnées se fondant sur les méthodes de contrôle de divulgation statistique.

Les méthodes de contrôle de divulgation ont été élaborées en vue de : maximiser la quantité de données diffusées, protéger rigoureusement les renseignements personnels et mettre au point une méthode systématique rentable pour les futures mises à jour. Parmi les défis posés par le SCHIRPT, on peut citer sa nature asymétrique, le niveau de détail et le caractère sensible des renseignements relatifs aux accidents causant une blessure (parfois, les accidents sont largement connus dans la collectivité). Les résultats incluent des mesures de suppression au sein des enregistrements et au niveau global.

Le document s'achève sur une analyse concernant les répercussions de cette étude sur le domaine de l'information sur la santé dans le contexte de la hausse de la demande de microdonnées, les progrès de la technologie, le raffinement de plus en plus marqué des utilisateurs et le nombre croissant de fichiers de données publics et privés.

**(A) Masquage dans le cas de variables discrètes**

Myron J. Katzoff et Jay J. Kim, National Center for Health Statistics, Hyattsville, MD, États-Unis

L'utilisation de variables discrètes dont les distributions statistiques sont connues pour le masquage de données sur des variables discrètes est à l'étude depuis quelque temps. Cette communication présente d'autres résultats de nos recherches sur ce sujet. Les effets de l'échantillon avec et sans remplacement des populations finies constituent l'un des principaux intérêts. Les estimations des moments de premier et de deuxième ordre qui atténuent ou tiennent compte de la variation supplémentaire due au masquage de type connu sont mises en œuvre. Les répercussions du masquage des données originales sur la structure de corrélation des variables mesurées simultanément et sur les tests de signification statistique présentent un intérêt particulier. On soulève le besoin de mises au point ultérieures en vue d'entreprendre les analyses de données multivariées.

**(A) Création de fichiers de microdonnées à grande diffusion pour la « National Survey on Drug Use and Health »**

Feng Yu, Lanting Dai, Moshe Feder et James R. Chromy, RTI International, Research Triangle Park, NC, États-Unis

La National Survey on Drug Use and Health est une enquête-ménage nationale annuelle transversale sur la prévalence et les corrélats de l'usage de la drogue et sur la santé mentale aux États-Unis. Le risque de divulgation constitue une préoccupation majeure dans la création de fichiers de microdonnées à grande diffusion (FMGD), car les données de l'enquête contiennent de l'information de nature hautement délicate sur le comportement des répondants quant aux drogues et sur des questions liées à la santé. Par conséquent, une forte assurance de la confidentialité est nécessaire afin de protéger les renseignements personnels des répondants et de réduire la perception qu'ils ont du risque de divulgation.

Nous analysons et évaluons les risques d'intrusion à la fois interne et externe. Notre présentation se concentre sur notre processus de création de fichiers de microdonnées à grande diffusion (FMGD). Le processus se fonde principalement sur la méthode MASSC qui permet d'introduire une incertitude suffisante sur la présence et l'identité d'un répondant figurant dans la base de données en vue d'une protection contre une intrusion interne. Pour assurer une protection renforcée contre l'intrusion externe, nous réduisons les catégories comprenant des réponses rares, nous reprogrammons les variables ayant des valeurs extrêmes, nous supprimons les variables sans valeur analytique présentant un risque de divulgation élevé et nous traitons les variables de stratification.

Le processus de traitement se termine par l'évaluation de l'utilité analytique des bases de données traitées en comparant les FMGD (avec environ 54 000 à 59 000 enregistrements) avec les données originales sur une gamme de statistiques univariées et multivariées et leurs erreurs types.

**(A) Confidentialité et estimation de la variance par rééchantillonnage**

Wilson Lu, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Colombie-Britannique, Canada

La protection contre la divulgation de données par inadvertance dans les fichiers de données diffusées publiquement à partir d'enquêtes complexes devient de plus en plus importante. La majorité de la documentation sur le sujet tient compte du masquage des données diffusées en tant que tel. La suppression des identificateurs des strates et des fichiers de microdonnées à grande diffusion (FMGD) est une mesure simple prise en vue d'éviter la divulgation des données. Cette mesure peut rendre difficile l'estimation directe de la variance étant donné que les identificateurs des strates et des FMGD sont généralement nécessaires à l'obtention d'estimateurs de variance asymptotiquement sans biais.

On pourrait croire que la diffusion d'une série de poids de rééchantillonnage avec les données qui possèdent également les identificateurs des FMGD et de strates contournerait le problème. Ce n'est pas le cas. Nous montrons qu'en visualisant le problème lié à la reconstruction des identificateurs des FMGD ou des strates à partir des poids de rééchantillonnage comme un problème de grappes de grande dimension, on peut facilement reconstruire les identificateurs des FMGD en employant une méthode de répétition usuelle sans connaître la méthode de rééchantillonnage qui a été employée. Ainsi, même les utilisateurs sans grande expérience peuvent très facilement obtenir les identificateurs des FMGD à partir de la série de poids de rééchantillonnage si on ne prend pas toutes les précautions nécessaires.

Une méthode courante de masquage direct des données dans les enquêtes complexes consiste à modifier les données ou à permuter les valeurs des données d'un cas à un autre. Dans ce contexte, on pourrait permuter les identificateurs des strates ou des FMGD avant de créer la série de poids de rééchantillonnage. Cette opération n'aurait de répercussions que sur l'estimation de la variance et pourrait être fait de manière à limiter l'incidence sur les estimations de la variance résultante pour les caractéristiques mesurées dans l'enquête. Nous présentons deux algorithmes automatiques pour permuter les identificateurs des FMGD et évaluer le rendement.

**Session 14 – Questions relatives à la conception et à la mise en œuvre des enquêtes sur la santé**

**(A) Questions méthodologiques liées à la mesure de la santé mentale des enfants et des jeunes en Grande-Bretagne**

Howard Meltzer, Amanda Wilmot, Abigail Dewar, Office for National Statistics, Londres, Angleterre

En 1999, la première enquête représentative à l'échelle nationale de la santé mentale des enfants et des jeunes âgés de 5 à 15 ans a été menée en Grande-Bretagne. La deuxième enquête a été réalisée en 2004. Ces enquêtes avaient trois objectifs : l'estimation de la prévalence des troubles mentaux chez les jeunes personnes, l'examen de leur utilisation des services de santé, des services sociaux et des services éducatifs et l'analyse des facteurs de risque liés aux troubles mentaux. On a mené respectivement 10 500 et de 8 000 interviews.

Certaines questions essentielles concernant un grand nombre d'enjeux méthodologiques ont dûes être abordées.

Comment peut-on obtenir un échantillon représentatif des enfants âgés de 5 à 15 ans?

Doit-on utiliser un plan d'échantillonnage à un ou à deux degrés?

Faut-il utiliser un instrument d'évaluation épidémiologique normalisé ou créer un nouvel instrument?

Les renseignements devaient provenir de trois sources : les enfants, leurs parents et les enseignants. Quelles méthodes d'enquête devait-on suivre pour chaque source?

Quel parent faut-il interviewer?

Dans quel ordre faut-il mener les interviews?

Quels sont les enjeux précis liés aux interviews menées auprès des enfants et des jeunes?

Comment doit-on procéder en présence de répondants qui ne parlent pas l'anglais?

Quelle est la meilleure façon d'organiser les questionnaires pour obtenir des données quantitatives et qualitatives?

Comment doit-on recueillir les données quantitatives et qualitatives auprès de ces trois sources?

Comment faut-il analyser les données lorsqu'il n'y a qu'une ou deux sources?

Dans quelle mesure et sous quelle forme doit-on utiliser des données cliniques?

Peut-on utiliser les données afin de produire des estimations régionales?

On analyse des facteurs pris en compte afin de parvenir à une décision sur ces enjeux.

**(A) Échantillonnage de la population Maorie dans la « New Zealand Health Survey »**

Robert Clark<sup>1</sup> et Sarah Gerritsen<sup>2,1</sup>, University of Wollongong, Wollongong, NSW, Australie,<sup>2</sup> Public Health Intelligence, Ministère de la Santé, Wellington, Nouvelle Zélande

Atteindre un échantillon de taille suffisante auprès des sous-populations ethniques est une difficulté courante dans la réalisation d'un plan d'échantillonnage d'enquête sur la population. Environ 11 % de la population de la Nouvelle-Zélande appartient au groupe ethnique Maori. Ce groupe est relativement regroupé, mais la plupart des Maoris habitent des secteurs où leur proportion est plutôt faible, ce qui rend difficile l'obtention d'un échantillon de façon efficace. Dans le plan de la New Zealand Health Survey (enquête sur la santé de la population de la Nouvelle-Zélande) de 2006-2007, on a utilisé deux stratégies d'échantillonnage des Maoris, des peuples du Pacifique et des Asiatiques afin d'obtenir un échantillon de taille suffisante tout en limitant les effets du plan. La première stratégie consiste en un ciblage léger selon lequel on accorde une plus grande probabilité de sélection aux logements des régions comptant une proportion élevée de Maoris. La deuxième stratégie consiste à sélectionner (a) un échantillon de base dont les membres sont sélectionnés sans égard à leur ethnicité et (b) un grand échantillon ayant fait l'objet d'un tri où l'on tient compte de l'ethnicité de chaque membre et où l'on ne soumet que les membres des groupes ethniques d'intérêt au questionnaire complet. L'équilibre entre le tri et le ciblage sera décrit pour la New Zealand Health Survey, et on présentera une théorie relativement à l'optimisation générale de cet équilibre.

**(A) Élaboration et mise à l'essai d'un questionnaire sur la santé de l'enfant avec réponse par procuration par un parent ou un gardien pour la « New Zealand Health Survey »**

Sarah Gerritsen, conseillère (enquêtes), Public Health Intelligence, Ministère de la Santé, Wellington, Nouvelle-Zélande

Le ministère de la santé de la Nouvelle-Zélande a modifié son enquête sur la santé de la population, la New Zealand Health Survey (NZHS), afin d'y inclure un questionnaire sur la santé des enfants. Outre la collecte de renseignements en personne auprès des adultes des ménages figurant dans l'échantillon, la NZHS de 2006-2007 prévoit également la sélection aléatoire d'un membre âgé de moins de 15 ans, le cas échéant, dans chaque ménage sélectionné et la collecte des renseignements supplémentaires à son égard auprès du soignant principal.

Le questionnaire pour enfants de la NZHS vise principalement à recueillir de nouvelles données sur la santé qui seront utilisées afin

d'évaluer la santé des enfants à l'échelle de la population, l'utilisation des services de santé ainsi que les comportements à risque pour la santé et ceux qui sont destinés à la protéger, et qui ont été adoptés pendant l'enfance. Auparavant, seules les données recueillies après que l'enfant a eu recours au système de santé permettaient une évaluation de l'état de santé des enfants de la Nouvelle-Zélande, que ce soit au moyen de dossiers administratifs d'hôpitaux et de bases de données sur les maladies et les blessures.

Le présent document examine la création du questionnaire sur la santé des enfants faisant partie de la New Zealand Health Survey de 2006-2007, notamment en ce qui a trait au choix des sujets, à la création des questions, aux tests cognitifs, à la conception de l'échantillon préliminaire, à la rédaction du questionnaire final et à la répétition générale. Le produit final intéressera les chercheurs en santé du gouvernement qui recueillent actuellement, dans le cadre d'enquêtes, les données sur la santé des enfants fournies au moyen d'une procuration par un soignant. Il intéressera également les personnes qui songent à ajouter cet élément important à leur programme d'évaluation de la santé de la population.

**(A) Conception de questions pour le dépistage des personnes handicapées dans les enquêtes sur la population active : mesures prises en vue de déterminer le niveau d'emploi des adultes handicapés dans la « CPS »**

Terence M. McMenemy, Division of Labour Force Statistics, Bureau of Labour Statistics, Washington, DC, États-Unis

En 1998, le décret-loi 13078 a exigé la création d'une mesure exacte et fiable du taux d'emploi des personnes ayant des incapacités. D'ailleurs, cette mesure devrait être publiée le plus souvent possible. Le Bureau of Labor Statistics (BLS) [bureau de la statistique du travail], en collaboration avec le groupe de travail interorganismes sur la méthodologie de mesure du taux d'emploi, a défini l'objectif qui consiste à insérer un petit ensemble de questions dans la Current Population Survey (CPS) [enquête sur l'état de la population]. Une série de questions possibles a été tirée d'enquêtes existantes, soumise à des tests cognitifs et intégrée à la National Comorbidity Survey [enquête nationale sur la comorbidité] aux fins d'un essai sur le terrain. Le BLS a analysé les données d'essai afin de déterminer la situation des répondants vis-à-vis de l'incapacité. On a utilisé la méthode Delphi pour les cas difficiles. On a réalisé différents tests afin de déterminer l'ensemble de questions qui correspondait le mieux aux résultats de l'analyse initiale. Le BLS a ensuite soumis les questions à des tests cognitifs afin de repérer tout problème de compréhension qui pourrait survenir au cours de la CPS. En février 2006, l'ensemble de questions

portant sur l'incapacité a été ajouté dans les questionnaires comportant un panel fractionné de répondants de la CPS. Une analyse de ces données permettra de déterminer si les questions fonctionnent dans le cadre d'une enquête sur la main-d'œuvre et si l'ajout de cet ensemble de questions nuira à la CPS.

---

## Session A des présentations par affiche

**(A) Recherche par simulation sur les méthodes d'imputation hot deck et d'imputation multiple à l'aide de données de la HCSDB**  
Donsig Jang, Amang Sukasih et Xiaojing Lin, Mathematica Policy Research, Inc., Washington, DC, États-Unis

À l'heure actuelle, lors de l'analyse des données de l'Enquête sur les soins de santé auprès des bénéficiaires du régime de santé du département de la Défense (HCSDB), aucune imputation n'est faite pour corriger la non-réponse partielle. On exclut plutôt de l'analyse tous les cas de non-réponse partielle, ce qui réduit l'efficacité des tailles d'échantillons disponibles pour l'analyse. Bien que les régions géographiques correspondant aux régions desservies sont considérées depuis longtemps comme des domaines d'analyse importants par les chercheurs et les décideurs qui utilisent les données de la HCSDB, la petite taille des échantillons, les faibles taux de réponse et les cas de non-réponse partielle pourraient empêcher ces derniers de produire des estimations statistiquement fiables à l'égard de la région desservie. L'imputation pour la non-réponse partielle peut contribuer non seulement à réduire le biais dans l'estimation, mais aussi à accroître la capacité d'estimation en ce qui a trait aux domaines couverts.

Toutefois, le choix d'une méthode d'imputation parmi les méthodes existantes peut tenir à des facteurs de nature non technique. Nous examinons deux méthodes d'imputation : une méthode d'imputation unique, c.-à-d., l'imputation hot deck par la méthode du plus proche voisin dans les classes et une méthode d'imputation multiple, c.-à-d., l'imputation multivariée par régression séquentielle (IMRS). Nous prévoyons réaliser une étude par simulation qui nous permettra de quantifier le biais causé par la non-réponse à un petit nombre de questions choisies, puis de comparer les résultats donnés par les deux méthodes d'imputation susmentionnées. Les valeurs déclarées pour ces questions seront traitées comme les valeurs « vraies » et nous simulerons les non-réponses aux questions en générant de 10 % à 15 % de valeurs manquantes parmi ces valeurs « vraies » en utilisant divers mécanismes de non-réponse. Les résultats de cette étude par simulation pourraient être utilisés pour décider de la méthode qu'il conviendrait d'appliquer pour l'imputation des données de la HCSDB.

**(A) Estimation de la taille d'échantillon pour les enquêtes complexes : Réalisation d'un consensus dans un environnement d'hypothèses multiples, d'intervenants multiples et de contraintes budgétaires**

JC Victor<sup>1,2</sup>, L Diemert<sup>1,2</sup>, S Bondy<sup>1,2</sup>, KS Brown<sup>1,3</sup>, J Cohen<sup>1,2</sup>, R Ferrence<sup>1,2</sup>, PW McDonald<sup>1,3</sup>, P Selby<sup>1,4</sup>, T Stephens,<sup>1</sup> <sup>1</sup>Unité de recherche sur le tabac de l'Ontario; <sup>2</sup>Université de Toronto; <sup>3</sup>Université de Waterloo; <sup>4</sup>Centre de toxicomanie et de santé mentale

Justification : L'estimation de la taille de l'échantillon est l'une des étapes essentielles de la conception d'une étude. Dans le cas d'études où le nombre de variables d'intérêt dépendantes et indépendantes est faible, l'estimation se fait facilement au moyen de logiciels du commerce. Par contre, comment estime-t-on la taille de l'échantillon lorsqu'il faut réaliser un consensus entre de nombreux intervenants pour des projets polyvalents, comme les enquêtes complexes représentatives de la population? Nous présentons un modèle de démonstration produit au moyen d'un tableau ordinaire en vue de fournir des estimations en temps réel de la puissance, de la précision et des coûts financiers sous divers scénarios de taille d'échantillon, tel qu'il a été utilisé dans l'élaboration de l'Enquête sur le tabagisme en Ontario.

Méthodes : L'Enquête sur le tabagisme en Ontario est une enquête longitudinale et transversale réalisée auprès des fumeurs et des non-fumeurs. Le plan de sondage, les hypothèses et les questions ont été établis par réalisation d'un consensus entre les divers chercheurs, dont les besoins variaient de l'estimation de tendances transversales à l'analyse de mesures répétées.

Résultats : Des estimations stratifiées selon la région de la prévalence de l'usage du tabac et de l'incidence des résultats en matière de santé liés à l'usage du tabac produites d'après les données d'enquêtes transversales ont été utilisées pour calculer les nombres résultants d'événements pour l'étude sous diverses stratégies d'échantillonnage. Des formules de taille d'échantillon complexes appropriées, intégrant les effets des cas perdus de vue au suivi et les effets de plan, ont été imbriquées dans un tableau Excel. L'utilitaire affichait la puissance pour plusieurs analyses transversales et longitudinales et le coût total d'enquête basé sur les coûts d'intrant-utilisateur par interview.

Conclusion : Une approche polyvalente de l'estimation de la taille d'échantillon est nécessaire pour que l'élaboration d'une enquête complexe, la diffusion d'estimations de population multiples et la vérification d'hypothèses multiples demeurent rentables. L'utilisation de tableurs ordinaires et d'affichages graphiques en plus des formules

complexes ont facilité considérablement le dialogue entre les méthodologistes et les intervenants. L’utilitaire a permis de démystifier la notion de « boîte noire relative à la taille d’échantillon » et a facilité les compromis lors des prises de décision concernant les éléments du plan de sondage.

**(A) L’effet de la spécification du modèle sur les données résultant de l’imputation multiple : leçons tirées du projet DC-HOPE du NIH**

Marie G. Gantz, Ph.D. et M. Nabil El-Khorazaty, Ph.D. pour la NIH-DC Initiative, RTI International, Research Triangle Park, Caroline du Nord, États-Unis

Contexte : On fait souvent appel à l’imputation multiple pour combler les lacunes dues aux données manquantes dans les études longitudinales. Cependant, les hypothèses faites dans le modèle d’imputation peuvent influencer fortement sur les valeurs imputées. Nous comparons deux versions d’un ensemble de données ayant subi des imputations multiples générées au moyen de modèles qui diffèrent sur un aspect important. Les résultats de l’étude fondés sur les deux ensembles de données imputées sont comparés à ceux obtenus d’après les données avant l’imputation.

Méthodes : L’ensemble de données provient d’une étude réalisée auprès de femmes enceintes afro-américaines recrutées à Washington DC. On les a affectées aléatoirement à un groupe recevant les soins habituels ou à un groupe bénéficiant d’une intervention intégrée d’éducation et de counselling en vue de réduire les risques liés aux résultats indésirables de la grossesse. Les femmes ont été interviewées à quatre moments dans le temps. L’imputation multiple a été effectuée deux fois : 1) pour toutes les participantes regroupées et 2) séparément pour les femmes appartenant à chaque groupe de soins, c’est-à-dire en tenant compte explicitement de l’effet du traitement obtenu et des interactions faisant intervenir le groupe de soins. Dans les deux imputations, le nombre de séances d’intervention auxquelles les participantes à l’étude ont assisté faisait partie des variables prédictives, afin de tenir compte de la prestation de l’intervention.

Résultats : Les analyses des deux ensembles de données multi-imputées ont révélé un plus petit nombre d’effets significatifs du traitement que l’analyse des données avant l’imputation. Comparativement à la méthode d’imputation pour l’ensemble des participantes à l’étude, l’imputation séparée pour chaque groupe de soins produit un plus grand nombre d’effets significatifs du traitement. Cependant, elle pourrait donner lieu à une surestimation de l’effet du traitement reçu, car elle repose sur l’hypothèse que les femmes

comprises dans le groupe qui reçoit l’intervention pour lesquelles des données manquent se comportaient comme les autres femmes de ce groupe et non comme celles du groupes recevant les soins habituels.

Conclusion : Le choix du modèle d’imputation peut avoir une incidence importante sur les valeurs des données imputées, donc sur les conclusions tirées de l’étude.

**(A) Validation d’une méthode qualitative appliquée à un instrument multidimensionnel comprenant une question ouverte**

José Eduardo Corrente, Cinthia Esbrile Moraes et Tania Ruiz, UNESP, Botucatu, São Paulo, Brésil.

La population du Brésil a vieilli puisque le nombre de personnes âgées a augmenté. Toutefois, cette population devrait pouvoir jouir d’une bonne qualité de vie en vieillissant, malgré la perte d’autonomie et l’apparition de limites à un âge avancé. Par conséquent, on a élaboré des instruments pour mesurer la qualité de vie des personnes âgées. Ainsi, un questionnaire comprenant divers instruments validés et une question ouverte a été administré à un groupe de citoyens âgés dans la ville de Botucatu, SP, Brésil. L’analyse des réponses à la question ouverte, évaluées selon des méthodes qualitatives, a permis d’établir onze catégories d’opinions des personnes âgées sur la qualité de vie. On a procédé à une analyse par grappe de ces réponses et transformé les réponses en variables binaires, ce qui a permis d’établir trois groupes de personnes âgées. Dans le cadre de l’analyse par grappe, on a sélectionné subjectivement d’autres variables de l’instrument comme celles qui sont les plus étroitement liées à la qualité de vie. Par conséquent, cette étude visait à valider les catégories obtenues à l’aide de la question ouverte par les questions fermées de l’instrument au moyen d’associations et de l’application de tests du khi carré au niveau de signification de 5 %. On a observé que l’analyse qualitative permet de déterminer les phénomènes sans égard à la saturation d’une catégorie. La méthode quantitative, en revanche, montre le poids de chaque catégorie dans un ensemble, c’est-à-dire globalement. On a également observé que des variables comme la spiritualité, la rectitude, la charité, les connaissances et le travail s’avèrent très subjectives et difficiles à évaluer au moyen d’un instrument multidimensionnel. En pareils cas, il faut concevoir des instruments plus spécifiques.

**Session B des présentations par affiche****(A) Élaboration et conception de l'enquête sur le tabagisme en Ontario**

S. Bondy<sup>1,2</sup>, L. Diemert<sup>1,2</sup>, J.C. Victor<sup>1,2</sup>, K.S. Brown<sup>1,3</sup>, J. Cohen<sup>1,2</sup>, R. Ferrence<sup>1,2</sup>, P.W. McDonald<sup>1,3</sup>, P. Selby<sup>1,4</sup>, T. Stephens<sup>1</sup>

<sup>1</sup>Unité de recherche sur le tabac de l'Ontario; <sup>2</sup>Université de Toronto;

<sup>3</sup>Université de Waterloo; <sup>4</sup>Centre de toxicomanie et de santé mentale

**Raison d'être :**

Au cours des 15 dernières années, la prévalence de l'usage de la cigarette chez les adultes a diminué en Ontario, passant à environ 20 %. L'analyse de sous-groupes de fumeurs s'appuyant sur des données d'enquêtes transversales fondées sur la population donne souvent des résultats statistiquement instables ou des estimations non déclarables. Ces résultats font obstacle à l'étude des effets des campagnes et des politiques publiques antitabac. Les enquêtes longitudinales, même si elles sont plus utiles à cet égard, ne peuvent souvent fournir des estimations pouvant être généralisées à l'ensemble de la population. Nous examinons une méthode qui permet de surmonter ces obstacles en combinant des composantes transversales et longitudinales dans l'Enquête sur le tabagisme en Ontario (OTS pour Ontario Tobacco Survey).

**Méthodes :**

Pour répondre aux besoins des chercheurs qui s'intéressent à l'usage du tabac en Ontario, il fallait une nouvelle étude permettant de suréchantillonner les fumeurs tout en obtenant des résultats en temps opportun sans dépasser le budget. Le plan d'enquête était fondé sur la méthode de l'ENSP et comprenait une composante longitudinale cumulative et une composante transversale répétée.

**Résultats :**

L'OTS est une enquête transversale par composition aléatoire réalisée auprès de fumeurs et de non-fumeurs et comprenant trois interviews de suivi auprès des fumeurs; elle a débuté en juillet 2005. On pose aux fumeurs et aux non-fumeurs des questions identiques sur leurs opinions et sur les politiques publiques. Le comportement et les attitudes des fumeurs sont évalués tous les six mois. Ce plan d'enquête permet de produire des estimations de population générales relativement aux politiques publiques tout en fournissant de nombreuses données pouvant servir à vérifier les hypothèses concernant le renoncement au tabac.

**Conclusion :**

Une étude mixte transversale et par panel produit des complications sur le plan de l'analyse et de la pondération et peut toujours produire

des estimations axées expressément sur les résultats, qui sont peu fiables ou non déclarables. Toutefois, en limitant la participation à la composante longitudinale à un sous-groupe d'intérêt, soit, aux fumeurs dans le cas qui nous intéresse, nous montrons que ce plan d'enquête assure l'efficacité tant statistique que financière.

**(F) Un modèle pour l'estimation de l'espérance de vie dans les petites villes canadiennes**

Philippe Finès, Statistique Canada, Ottawa, ON, Canada

L'espérance de vie à la naissance (EV) est fréquemment utilisée comme indicateur de santé. Pour les grandes villes, lorsque les décès sont comptés pour trois années consécutives, le calcul de l'EV produit un intervalle de confiance à 95% (IC) dont l'étendue est inférieure à une demi-année. Cet estimé est suffisamment précis pour permettre des comparaisons entre villes et pour évaluer le gradient revenu-mortalité à l'intérieur d'une ville. Cependant, lorsque la population d'une ville est inférieure à 100 000 personnes, l'estimé de l'EV devient imprécis et invalide. L'objectif de ce projet est d'obtenir en même temps un bon estimé de l'EV et un IC pas trop grand, même pour ces petites villes. Nous pouvons atteindre cet objectif si nous tenons compte de l'information disponible hors de ces petites villes. Par exemple, en utilisant la forme générale de la courbe de mortalité – qui est caractéristique de toute ville canadienne –, nous pouvons réduire l'étendue de l'IC. La prise en compte d'information externe est le fondement de la théorie de l'« estimation dans les petites régions ». Dans cette conférence, nous montrerons des techniques spécifiques que nous avons développées pour réaliser notre objectif d'améliorer la précision de l'EV dans les petites villes canadiennes. En particulier, nous présenterons les équations essentielles de même que les résultats intéressants.

**(F) Infarctus du myocarde au Québec : les immigrants et les Canadiens de naissance sont-ils différents ?**

Maria Gabriela Orzanco, Alain Vanasse, Théophile Niyonsenga et Josiane Courteau, Université de Sherbrooke, Sherbrooke, QC, Canada

Dans le cadre du programme « Système d'information spatio-temporel sur les maladies chroniques (SIST-MC) » financé par le réseau Geoïde, cette recherche vise à élaborer un filtre « immigration » pour le SIST-MC, car les immigrants présentent des particularités en matière de santé. L'étude cherche à vérifier l'existence d'une relation entre les caractéristiques de la population immigrante à Montréal et l'infarctus du myocarde (IM) en termes d'incidence et d'issue clinique. Les questions de recherche sont : (1) Les taux d'incidence d'IM, de revascularisation et de survie des populations des secteurs de recensement (SR) où les immigrants sont nombreux diffèrent-ils de ceux des populations des

autres SR? (2) Ces différences se maintiennent-elles dans le temps? Quels sont les facteurs associés aux changements? (3) Quelles caractéristiques s’associent plus fortement à ces issues de santé ?

Unités d’analyse de l’étude : populations des SR de 1991, 1996 et 2001. Variables à l’étude : (1) taux d’incidence d’IM, (2) taux de revascularisation et (3) taux de survie. Covariables : âge, sexe, comorbidité et utilisation des médicaments cardioprotecteurs. Sources de données : (1) les recensements de population et (2) le registre médico-administratif du Québec MED-ECHO. Les populations des SR sont décrites en termes d’immigration, de démographie et du statut socioéconomique et classifiées en « classes d’immigration-défavorisation » (CID). Les CID sont comparées aux issues de santé pour les trois périodes. Les variables les plus importantes pour ces issues de santé seront identifiées à l’aide de modèles de régression.

**(A) Géovisualisation des données sur la santé et sur le capital social provenant des enquêtes réalisées par Statistique Canada**

Daniel Rainham, Daniel Krewski, Ian McDowell et Mike Sawada, Université d’Ottawa, Ottawa, ON, Canada

La classification et l’identification des emplacements où se trouvent les personnes qui déclarent être en plus ou en moins bonne santé ou posséder plus ou moins de capital social dans une région particulière, comme une région sociosanitaire, contribue considérablement à comprendre les associations entre lieu et santé. L’objectif de l’étude proposée est de classer et de cartographier divers secteurs de la région sociosanitaire de la Zone 6 (figure 1) de la Nouvelle-Écosse (municipalité régionale de Halifax et régions de la vallée de l’Annapolis) en fonction de l’état de santé (dimension 1) et du capital social (dimension 2). Nous avons procédé au résumé analytique des réponses autodéclarées aux questions sur l’état de santé, l’état mental et le capital social figurant dans les fichiers maîtres de l’Enquête sur la santé dans les collectivités canadiennes (cycles 1.1, 1.2 et 2.1), de l’Enquête nationale sur la santé de la population (cycle 5) et de l’Enquête sociale générale (cycles 13, 14, 17 et 18). Les réponses ont été géocodées en utilisant le Fichier de conversion des codes postaux (FCCP+) de Statistique Canada et importées dans un système d’information géographique (SIG) afin que le code postal associé à la réponse soit assigné à une latitude et à une longitude dans la région sociosanitaire de la Zone 6 de la Nouvelle-Écosse. Nous avons utilisé des estimateurs à noyau de la densité et des interpolateurs spatiaux supplémentaires pour élaborer des surfaces statistiquement lissées de la distribution des valeurs des réponses pour chaque question. Le lissage rend impossible la détermination de l’emplacement d’un

répondant particulier et la divulgation des renseignements confidentiels contenus dans la base de sondage de Statistique Canada. L'utilisation de réponses à des questions semblables recueillies dans le cadre de plusieurs enquêtes accroît la probabilité de déceler une hétérogénéité entre les réponses obtenues dans la région sociosanitaire et rend la classification cartographique lissée plus exacte.

## **Session C des présentations par affiche**

### **(A) Les effets de différentes méthodes de sélection des dossiers de sortie d'hôpital aux fins de l'établissement de rapports sur les blessures**

Susan G. Mackenzie, Agence de santé publique du Canada, Ottawa, ON, Canada

#### **Introduction :**

D'importantes différences existent entre l'Agence de santé publique du Canada (ASPC) et l'Institut canadien d'information sur la santé (ICIS) quant à leurs méthodes de sélection des dossiers de sortie d'hôpital aux fins de l'établissement de rapports sur les blessures. Trois autres pays utilisent des méthodes qui diffèrent les unes des autres et des deux méthodes canadiennes.

#### **Objectif :**

Comparer les effets des cinq méthodes.

#### **Méthodes :**

À partir des dossiers de sortie d'hôpital de l'ICIS, on a établi un ensemble de 14 772 dossiers indiquant une cause extérieure de blessure ou comprenant un diagnostic de blessure tirés des dossiers de sorties d'hôpital déclarés par une province canadienne en 2000-2001. Les cinq méthodes ont été appliquées à la sélection des données.

#### **Résultats :**

Parmi les dossiers sélectionnés, 87 % comprenaient une cause extérieure, 70 %, un diagnostic de blessure et 56 %, l'un et l'autre. Les pourcentages des 14 772 dossiers retenus selon les cinq méthodes étaient les suivants : Pays-Bas : 87 %, Canada (critères de l'ASPC) : 73 %, Canada (critères de l'ICIS) : 64 %, Australie : 56 % et États-Unis : 53 %. Les cinq méthodes ont également abouti à des différences importantes relativement à la fréquence de certains types de blessures. Ces dernières seront décrites dans la présentation.

**Conclusion :**

Des méthodes normalisées de sélection des dossiers aux fins de l'établissement de rapports sur les causes extérieures de traumatisme seraient utiles.

**(A) Remplacement de la CIM-9 par la CIM-10-CA dans les études sur les soins de santé hospitaliers**

Christie Sambell, Hude Quan et Helen Johansen, Statistique Canada, Ottawa, ON, Canada

L'utilisation de données administratives sur la santé pour les recherches en santé au Canada a été facilitée par leur disponibilité, leur large couverture géographique et leur saisie relativement complète des contacts avec le système de santé dans le cas d'une population déterminée. Toutefois, il faut normaliser les définitions afin de permettre les comparaisons entre les provinces et les années. Un outil méthodologique essentiel, la Classification internationale des maladies (CIM) de l'Organisation mondiale de la santé (OMS), fournit un système uniforme de codage des diagnostics, permettant la réalisation d'études longitudinales et comparatives. En 1992, on a adopté la 10<sup>e</sup> révision de la CIM (CIM-10) et, à compter de 2001, les provinces canadiennes ont entrepris la mise en œuvre graduelle d'une version canadienne de la CIM-10 (CIM-10-CA) pour le codage des diagnostics hospitaliers. Toutefois, sa mise en œuvre entraînera le remaniement, pour application dans la CIM-10-CA, de plusieurs outils méthodologiques établis applicables dans la CIM-9-MC, notamment l'indice de Charlson et les comorbidités d'Elixhauser qui permettent d'estimer la gravité de la maladie. Notre étude porte sur les taux des comorbidités mentionnés ci-dessus selon la province avant et après le passage de la CIM-9 à la CIM-10-CA et propose la façon la plus uniforme de déterminer les taux au moyen de nouveaux algorithmes de codage en examinant différentes combinaisons de correction.

**(A) Examen de l'incidence de la résistance du participant sur la qualité des données dans la National Health Interview Survey**

James M. Dahlhamer, Catherine M. Simile et Net Taylor, National Center for Health Statistics, Centers for Disease Control and Prevention, États-Unis

Au cours des vingt dernières années, le taux de réponse aux enquêtes gouvernementales menées aux États-Unis a diminué de façon graduelle, mais constante. Cette situation a poussé les responsables de la collecte de données à consacrer d'importantes ressources pour optimiser la réponse par la conversion des cas de refus. Bien qu'en général on présume que de telles mesures réduisent le biais de non-réponse et les autres formes d'erreur dans les enquêtes, la réalité révèle que la conversion des refus peut causer des effets inattendus,

voire nuisibles, la qualité des données. Par exemple, les répondants réticents peuvent adopter une attitude minimaliste, c.-à-d. qu’ils ont tendance à donner les réponses minimales nécessaires à la poursuite de l’interview, ce qui peut donner lieu à des niveaux élevés de points manquants et d’interviews partiellement complétées. Les répondants réticents fournissent-ils autant de données que les répondants qui collaborent? Existe-t-il des différences selon la justification de la réticence (p. ex., « trop occupé », « pas intéressé », « questions de vie privée »)?

Afin de répondre à ces questions, on a procédé à des analyses de l’historique des tentatives de contact, du cadre et des données de base provenant de la National Health Interview Survey de 2005, une enquête sur la santé, volet ménage, menée en personne chaque année par le National Center for Health Statistics et les Centers for Disease Control and Prevention des États-Unis. Plus précisément, on a effectué des comparaisons entre les taux d’interview partiels, de refus de réponse et de réponses « ne sais pas » des participants qui collaborent et des participants réticents, par rapport à un ensemble d’indicateurs socioéconomiques et critiques de la santé. L’étude se penche ensuite sur l’incidence de la résistance sur ces indicateurs de qualité dans un contexte multidimensionnel, tout en contrôlant les mesures sociodémographiques et socioenvironnementales. Les conséquences de nos résultats sur les opérations actuelles des enquêtes, les indicateurs de qualité des données et les mécanismes par lesquels nous communiquons la qualité des données aux utilisateurs publics font l’objet de discussions.

**(A) Effet des incitatifs sans condition sur une enquête à plusieurs degrés : cas de la Health Survey for England**

C. Deverill, H. Wardle, R. Craig, National Centre for Social Research, Londres, Royaume-Uni

La Health Survey for England (HSE) est une enquête-ménages continue qui fournit des données sur la santé et le mode de vie de la population. Elle comprend des interviews sur place et la visite d’une infirmière. Ainsi que pour d’autres enquêtes réalisées en Grande-Bretagne, les taux de réponse des ménages ont baissé progressivement au cours des dix dernières années. Le taux de réponse à l’HSE a diminué de six points de pourcentage et les taux de réponse des particuliers, à divers degrés de l’enquête, ont baissé également.

La HSE comporte la présélection d’un échantillon de haute qualité d’adresses et l’envoi préalable d’une lettre à toutes les adresses sélectionnées. Une étude antérieure a montré que le taux de collaboration était de 3 à 5 points de pourcentage plus élevé si les

personnes recevaient un carnet de timbres-poste avec la lettre envoyée préalablement. L'étude a pris la forme d'une enquête à un degré. Nous avons conçu l'expérience de façon à déterminer si les incitatifs sans condition sous forme de timbres ou de signets envoyés avec la lettre auraient un effet sur le taux de réponse des ménages, ainsi que sur les taux de réponse à d'autres degrés de l'enquête (c.-à-d. visite de l'infirmière, mesures anthropométriques). À partir de juillet 2005, les adresses ont été réparties aléatoirement entre trois catégories de conditions : la réception d'un carnet de timbres, d'un signet avec les résultats pertinents de l'HSE ou d'une lettre préalable seulement (groupe témoin).

Selon les résultats préliminaires, le taux de réponse des ménages est de 2 à 3 points de pourcentage plus élevé pour le groupe ayant reçu les timbres que pour ceux ayant reçu le signet ou la lettre uniquement. Cependant, les taux de collaboration aux étapes ultérieures de l'enquête (mesures anthropométriques et visite d'une infirmière) les plus faibles ont été observés chez les personnes appartenant aux ménages ayant reçu les timbres.

Ces résultats donnent à penser que les incitatifs sans condition peuvent améliorer les taux de réponse des ménages, mais que la collaboration ne se poursuit pas jusqu'aux étapes de suivi de l'enquête. Les résultats complets de cette expérience seront disponibles au cours de l'été 2006.

**(A) Comparaison de l'utilisation autodéclarée des soins de santé mentale primaires dans l'Enquête sur la santé dans les collectivités canadiennes et des dossiers d'assurance-maladie provinciaux des répondants**

JoAnne Palin, Université de Colombie-Britannique, Vancouver, C.-B., Canada

Des renseignements exacts sur le moment de l'accès aux soins de santé mentale primaires sont d'importance capitale pour permettre de cerner les facteurs éventuellement modifiables qui pourraient faciliter la gestion permanente et en temps opportun des soins. Comme il n'existe pas de mesure « idéale » de l'utilisation des soins de santé mentale, il est utile de savoir comment les points forts, les lacunes et les limites des différentes sources de données influent sur les résultats des études. La présente étude compare deux mesures des données sur l'utilisation des soins de santé mentale primaires à l'échelle de la population, soit le volet Santé mentale et bien-être de l'Enquête sur la santé dans les collectivités canadiennes (ESCC, cycle 1.2) et les dossiers d'assurance-maladie dans la province de la Colombie-Britannique. Elle porte sur les quatre questions suivantes :

1) La prévalence sur 12 mois des contacts avec les omnipraticiens concernant des problèmes de santé mentale est-elle la même d'après les données d'enquête et les données administratives? (2) Quel est le niveau de concordance entre les données d'enquête et les données administratives en ce qui concerne les contacts avec un omnipraticien pour des problèmes de santé mentale durant les 12 mois précédant l'interview d'enquête? (3) Le niveau de concordance est-il constant au cours de la période de 12 mois ou bien baisse-t-il au cours de sous-périodes plus éloignées durant la période de 12 mois? (4) Quelles sortes de caractéristiques des répondants, y compris les troubles mentaux, sont associées à la concordance ou à l'absence de concordance?

Les résultats de cette étude fourniront des renseignements utiles sur la façon d'utiliser et d'interpréter chaque mesure de l'utilisation des soins de santé. En outre, cette étude contribuera de façon utile aux recherches sur le plan d'enquête et à celles qui visent à améliorer les méthodes d'utilisation des données administratives aux fins de recherches sur les services de santé mentale.

---

## Mot de la fin

Nous désirons vous remercier chaleureusement d'avoir participé au 23<sup>e</sup> Symposium international sur les questions de méthodologie de Statistique Canada, «**Enjeux méthodologiques reliés à la mesure de la santé des populations**». Nous espérons que vous avez trouvé ce Symposium très utile et bénéfique.

Veuillez compléter le formulaire d'évaluation avant de quitter et le laisser dans l'une des boîtes situées dans le foyer près de la sortie.

Les actes du Symposium seront disponibles sur CD-ROM approximativement en octobre 2007. Tous les participants du Symposium recevront une copie gratuite.

L'année prochaine, à la place du Symposium, Statistique Canada co-organise La Troisième Conférence internationale sur les enquêtes d'établissements (CIEE III), 18-21 juin 2007, à Montréal, Québec, Canada. Nous espérons vous compter à nouveau parmi nous. Nous reprendrons la série des symposiums en 2008.

Le comité organisateur du Symposium 2006



Statistique  
Canada

Statistics  
Canada

Canada