Welcome

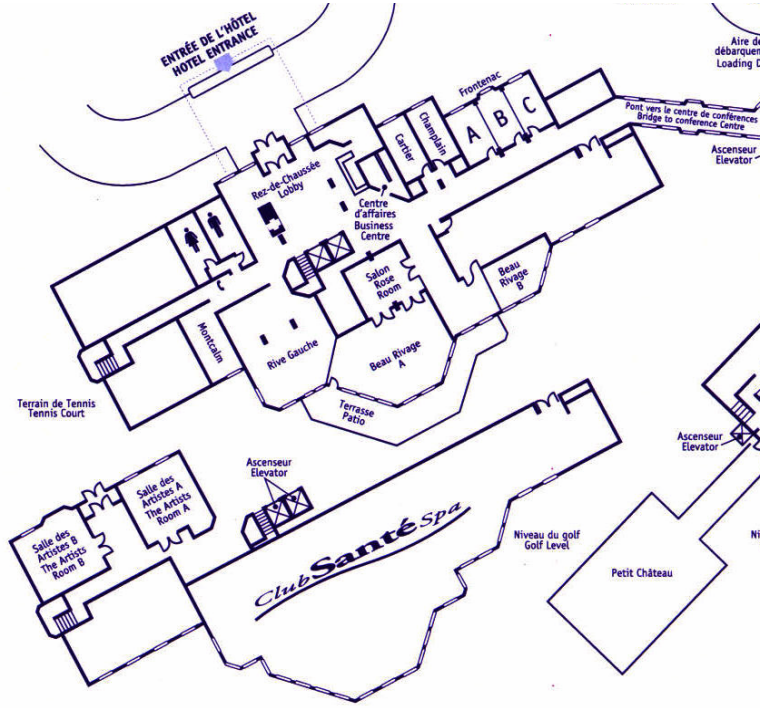
The Organizing Committee of Statistics Canada's 23rd International Symposium on Methodological Issues **"Methodological Issues in Measuring Population Health"** welcomes you. We hope that you will find the Symposium's program to be interesting, relevant, stimulating, and useful, and that you will enjoy meeting and discussing the subject with your colleagues over the next three days.

Symposium 2006 Organizing Committee

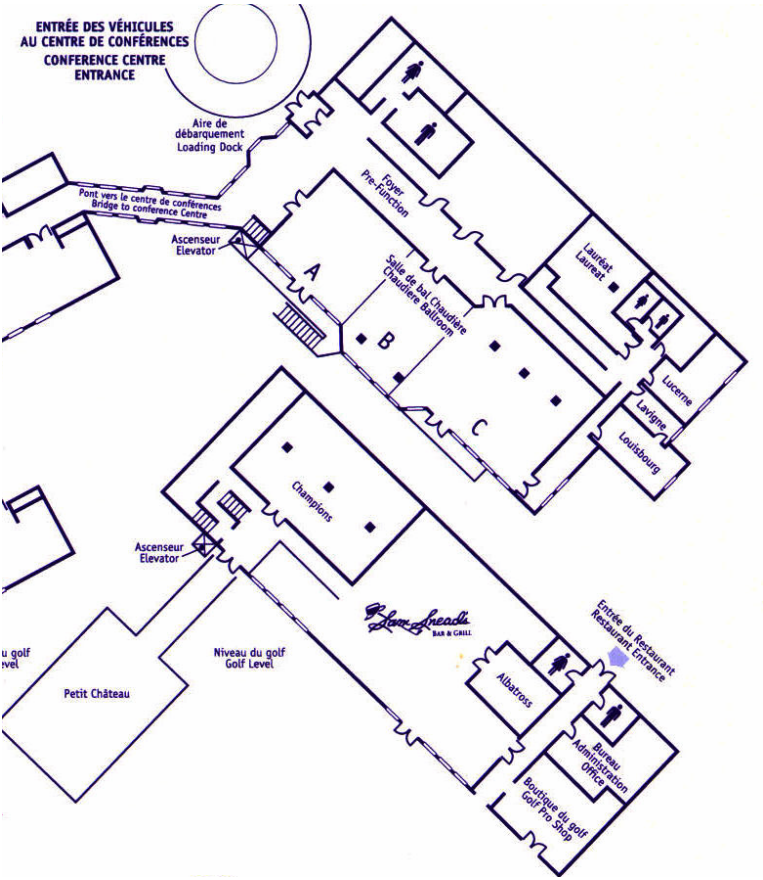
Milorad Kovacevic, Chair
Colin Babyak
Marie P. Beaudet
Yves Béland
Chris Mohl

Table of Contents:

Château Cartier
RELAIS · RESORT



Château Cartier
RELAIS · RESORT



Schedule

Day 1 – Wednesday, November 1, 2006

08:00 – 09:00	Workshop Registration, Refreshments (Chaudière ABC Lobby)	
09:00 – 16:30 Concurrent Workshops	**Workshop 1** - Record linkage in studies of population health — an overview (Chaudière B)	**Workshop 2** - Methods for analyzing longitudinal health survey data - Theory and applications (Chaudière A)

Day 2 – Thursday, November 2, 2006

07:45 – 08:45	Registration, Refreshments (Chaudière ABC Lobby)	
08:45 – 09:00	**Opening Remarks** (Chaudière ABC)	
09:00 – 10:00 Plenary Session	**Session 1**: Keynote Address – (Chaudière ABC)	
10:00 – 10:30	**Poster Session A**, Coffee Break (Chaudière ABC Lobby)	
10:30 – 12:00 Concurrent Sessions	**Session 2**: Current Issues in Small area Estimation with Application to Health Surveys (Chaudière AB)	**Session 3:** Analysis of Health Data – Applications (Chaudière C)
12:00 – 13:30	**Lunch** (Champions / Rive Gauche/Beau Rivage A)	
13:30 – 15:00 Concurrent Sessions	**Session 4**: Issues in Analysis of Longitudinal Health Data (Chaudière C)	**Session 5:** Combining Data from Different Sources (Chaudière AB)
15:00 – 15:30	**Poster Session B**, Coffee Break (Chaudière ABC Lobby)	
15:30 – 17:00 Concurrent Sessions	**Session 6**: Issues in Weighting, Estimation and Outlier Detection (Chaudière AB)	**Session 7**: Direct Health Measures Surveys: Problems and Possibilities (Chaudière C)
17:00 – 19:00	**Evening Reception** (Rive Gauche/Beau Rivage A)	

Day 3 – Friday, November 3, 2006

08:00 – 09:00	Registration, Refreshments (Chaudière ABC Lobby)	
09:00 – 10:00 Plenary Session	**Session 8**: Waksberg Award Address (Chaudière ABC)	
10:00 – 10:30	**Poster Session C**, Coffee Break (Chaudière ABC Lobby)	
10:30 – 12:00 Concurrent Sessions	**Session 9**: Methods for Analysis of Health Data (Chaudière C)	**Session 10**: Collection and Mode Effects (Chaudière AB)
12:00 – 13:30	**Lunch** (Champions / Rive Gauche/Beau Rivage A)	
13:30 – 15:00 Concurrent Sessions	**Session 11**: Sampling: Theory and Applications (Chaudière AB)	**Session 12**: Health Measurement: An International Perspective (Chaudière C)
15:00 – 15:15	**Break** (Chaudière ABC Lobby)	
15:15 – 16:45 Concurrent Sessions	**Session 13**: Confidentiality and Disclosure Control (Chaudière AB)	**Session 14**: Issues in Designing and Implementing Health Surveys (Chaudière C)
16:45 – 17:00	**Closing Remarks** (Chaudière C)	

PROGRAM

Wednesday, November 1, 2006

8:00-9:00 **Workshop Registration – Chaudière ABC Lobby**
Refreshments - Chaudière ABC Lobby

9:00-16:30 **Concurrent Workshops**

10:00-10:30 **Refreshments**
12:00-13:30 **Lunch break** – (Lunch not included in registration fee)
15:00-15:30 **Refreshments**

- **(E,F) Record Linkage in Studies of Population Health – An Overview –Chaudière B**
  Karla Fox, Department of National Defence
  Patricia Whitridge, Elections Canada

- **(E) Methods for Analyzing Longitudinal Health Survey Data – Theory and Applications – Chaudière A**
  Mary Thompson, University of Waterloo, Canada

* The capital letter in front of the title indicates the language of the presentation
E=English, F=French

Thursday, November 2, 2006

7:45-8:45 **Registration – Chaudière ABC Lobby**
Refreshments – Chaudière ABC Lobby

8:45-9:00 **(E,F) Opening Remarks – Chaudière ABC**

François Maranda, Statistics Canada

9:00-10:00 **Session 1 - Keynote Address – Chaudière ABC**

- **(E) Measuring Health in Population Surveys**
  Graham Kalton, Westat Inc., Washington, DC, USA

10:00-10:30 **Poster Session A – Chaudière ABC Lobby**
(See page 18 for list of presenters)

Coffee Break – Chaudière ABC Lobby

10:30-12:00 **Session 2 – Current Issues in Small Area Estimation with Application to Health Surveys – Chaudière AB**

Organizer: Avi Singh, Statistics Canada
Chair: Robert Fay, U.S. Census Bureau

- **(E) Mixed Linear-Nonlinear Models for Small Area Estimation with Application to the Canadian Community Health Survey**
  Avi Singh, Statistics Canada
  François Verret, Statistics Canada

- **(E) Dealing with Influential Observations and Outliers in Small Area Estimation**
  William R. Bell, U.S. Census Bureau
  Elizabeth T. Huang, U.S. Census Bureau

- **(E) Robust Bayesian Predictive Inference for the Finite Population Quantile of a Small Area**
  Balgobin Nandram, Worcester Polytechnic Institute, USA
  Jai Won Choi, National Center for Health Statistics, CDC, USA

- **(E) Robust Estimation of the Mean Squared Error of an EBLUP of a Small Area Mean**
  Partha Lahiri, University of Maryland, USA

10:30-12:00 **Session 3 – Analysis of Health Data - Applications – Chaudière C**

Chair: Dave Dolson, Statistics Canada

- **(F) The Methodological Challenges Related to the Analysis of Nutritional Data from the Nutrition Survey**
  François Verret, Statistics Canada

- **(F) Estimating Child BMI Growth Curves for Canada**
  Stéphane Tremblay, Statistics Canada
  Rebecca Morrison, Statistics Canada

- **(E) Obesity Rate Differences Between U.S. and Canadian Women and Between U.S. and Canadian Men: Findings from the Joint Canada / United States Survey of Health**
  Jane F. Gentleman, National Center for Health Statistics, USA
  Debra L. Blackwell, National Center for Health Statistics, USA
  Michael E. Martinez, National Center for Health Statistics, USA

- **(F) Studying Scenarios of Nutrition Intervention: The Example of Soft Drinks**
  Didier Garriguet, Statistics Canada

12:00-13:30 **Lunch – Champions / Rive Gauche / Beau Rivage A**

Thursday, November 2, 2006

13:30-15:00 **Session 4 – Issues in Analysis of Longitudinal Health Data – Chaudière C**

Chair: Georgia Roberts, Statistics Canada

- **(E) Measurement Error in Life History Data**
  Grace Y. Yi, University of Waterloo, Canada

- **(E) Flexible methods for analyzing longitudinal data in population health**
  Joel A. Dubin, University of Waterloo, Canada

- **(E) Complex Sampling Design Based Familial Longitudinal Health Data Analysis: An Overview**
  Brajendra Sutradhar, Memorial University of Newfoundland, Canada

- **(E) Estimation of Attributable Number of Deaths and Standard Errors from Simple and Complex Sampled Cohorts**
  Barry I. Graubard, National Cancer Institute, USA
  Mitchell H. Gail, National Cancer Institute, USA
  Katherine M. Flegal, National Center for Health Statistics, USA
  David F. Williamson, Centers for Disease Control and Prevention, USA

13:30-15:00 **Session 5 – Combining Data from Different Sources – Chaudière AB**

Chair: Jillian Oderkirk, Statistics Canada

- **(E) Quality Studies to Evaluate Linking Data with the Population Census**
  Glenys Bishop, Analytical Services Branch, ABS, Australia

- **(E) Merging Area-Level Census Data With Survey And Administrative Data**
  Denis Gonthier, Statistics Canada
  Tina Hotton, Statistics Canada
  Cynthia Cook, Statistics Canada
  Russell Wilkins, Statistics Canada

- **(E) A Case Study in Using Model-Assisted Estimation to Integrate Survey and Administrative Data**
  Robert E. Fay, U.S. Census Bureau

- **(E) Exact And Probabilistic Record Linkage**
  Claude Nadeau, Statistics Canada
  Marie P. Beaudet, Statistics Canada
  Jocelyne Marion, Statistics Canada
  Christel Le Petit, Statistics Canada

15:00-15:30 **Poster Session B – Chaudière ABC Lobby**
(See page 18 for list of presenters)

Coffee Break – Chaudière ABC Lobby

15:30 – 17:00 **Session 6 – Issues in Weighting, Estimation and Outlier Detection – Chaudière AB**

Chair: Jack Gambino, Statistics Canada

- **(E) Variable Selection Models for Weight Trimming**
  Michael Elliott, University of Michigan School of Public Health, USA

- **(E) Combining Information from Two Surveys to Improve on Analyses of Self-Reported Data in Estimating Measures of Health**
  Nathaniel Schenker, National Center of Health Statistics, USA
  Trivellore E. Raghunathan, National Center of Health Statistics, USA
  Irina Bondarenko, National Center for Health Statistics, USA

- **(E) Evaluation of Methods for Outlier Detection and Treatment in the U.S. Survey of Occupational Illnesses and Injuries**
  John L. Eltinge, Bureau of Labour Statistics, USA

- **(E) Combining Cycles of the Canadian Community Health Survey**
  Steven Thomas, Statistics Canada

Thursday, November 2, 2006

15:30 – 17:00 **Session 7 – Direct Health Measures Surveys: Problems and Possibilities – Chaudière C**

Organizer: Mark Tremblay, Statistics Canada
Chair: Michael Wolfson, Statistics Canada

- **(E) Learning the Unique and Peculiar Challenges of Direct Health Measures Surveys: The Canadian Experience**
  Mark Tremblay, Statistics Canada

- **(E) The U.S. National Health and Nutrition Examination Survey: What 40 Years of Experience has Taught Us**
  Cliff Johnson, National Centre for Health Statistics, USA.

- **(E) Direct Health Measures Surveys in Finland: From North Karelia and Beyond**
  Arpo Aromaa, KTL, Helsinki, Finland

17:00-19:00 **Evening Reception – Rive Gauche / Beau Rivage A**

Friday, November 3, 2006

8:00-9:00 **Registration – Chaudière ABC Lobby**
Refreshments – Chaudière ABC Lobby

9:00-10:00 **Session 8 – Waksberg Award Address – Chaudière ABC**

Organizer: John Kovar, Statistics Canada
Chair: Wayne Fuller, Iowa State University, USA

- **(E) The Analysis of Population-based Case Control Studies**
  Alastair Scott, University of Auckland, New Zealand

10:00-10:30 **Poster Session C – Chaudière ABC Lobby**
(See page 19 for list of presenters)

Coffee Break – Chaudière ABC Lobby

10:30-12:00 **Session 9 – Methods for Analysis of Health Data – Chaudière C**

Chair: David Binder, Statistics Canada

- **(E) Bootstrap Methods for Analyzing Complex Sample Survey Data**
  J. N. K. Rao, Carleton University, Canada

- **(E) Estimation of Regression Parameters with Survey Data**
  Wayne A. Fuller, Iowa State University, USA
  Yu Wu, Iowa State University, USA

- **(E) Addressing Data Sparseness in Contextual Population Health Research: Effects of Small Group Size and Targeted Cluster Analysis On Multilevel Model Performance, Bias And Efficiency Cohorts**
  Philippa Clarke, University of Michigan, USA
  Patricia O'Campo, University of Toronto, Canada
  Blair Wheaton, University of Toronto, Canada

Friday, November 3, 2006

- **(E) Causal Inference in Observational Studies using Health Administrative Data**
  Therese A. Stukel, ICES, Toronto, Canada

10:30-12:00 **Session 10 – Collection and Mode Effects – Chaudière AB**

Chair: Yves Béland, Statistics Canada

- **(E) Wireless Substitution in the United States and Canada: Prevalence and Impact on Random-Digit-Dialed Health Surveys**
  Stephen J. Blumberg, Centers for Disease Control and Prevention, National Center for Health Statistics, USA
  Julian V. Luke, Centers for Disease Control and Prevention, National Center for Health Statistics, USA

- **(E) Exploring the Impact of Mode on Key Health Estimates in the National Health Interview Survey**
  Catherine Simile, National Center for Health Statistics, USA
  Barbara Stussman, National Center for Health Statistics, USA
  James Dahlhamer, National Center for Health Statistics, USA

- **(E) Impact of Telephone versus Face to Face Repeat 24 Hour Recall Interviews on Food and Nutrition Surveys**
  S. Hayward, Food Directorate, Health Canada
  E. Junkins, Food Directorate, Health Canada
  M. Vigneault, Food Directorate, Health Canada
  M. Villeneuve, Food Directorate, Health Canada

- **(F) How To Account For An Important Change In Methodology When Analysing Data From The Canadian Community Health Survey?**
  Luc Côté, Institut de la statistique du Québec, Canada
  Robert Courtemanche, Institut de la statistique du Québec, Canada

---

12:00-13:30 **Lunch – Champions/Rive Gauche/Beau Rivage A**

Friday, November 3, 2006

13:30-15:00 **Session 11 – Sampling – Theory and Applications – Chaudière AB**

Chair: Mike Hidiroglou, Statistics Canada

- **(E) Simulation-based Systematic PPS Sampling With Unit Substitution**
  Mary E. Thompson, University of Waterloo, Canada
  Changbao Wu, University of Waterloo, Canada

- **(E) On Reuse Of Clusters In Repeated Studies**
  Stanislav Kolenikov, University of Missouri, USA
  Gustavo Angeles, University of North Carolina, Carolina Population Center, Chapel Hill, USA

- **(F) Sampling Design for the Canadian Health Measures Survey**
  Suzelle Giroux, Statistics Canada

- **(E) Challenges in the Design of the National Health and Nutrition Examination Survey**
  Leyla Mohadjer, Westat Inc.,USA

13:30-15:00 **Session 12 – Health Measurement: An International Perspective– Chaudière C**

Organizer and Chair: Jean-Marie Berthelot, CIHI

- **(E) Measuring the Health of Populations: The Conceptual and Analytic Approach of The Global Burden of Disease Study**
  Colin D. Mathers, Department of Measurement and Health Information, World Health Organization, Geneva, Switzerland

- **(E) International Comparisons In Measuring Health States: Experiences From The World Health Surveys**
  Somnath Chatterji, Department of Measurement and Health Information, World Health Organization, Geneva, Switzerland

- **(E) A Research Agenda for Developing Comparable Measures of Functional Health Status for use in Population Surveys**
  Chris Murray, Harvard University, USA

- **Discussant:** Michael Wolfson, Statistics Canada

Friday, November 3, 2006

15:00-15:15 **Break – Chaudière ABC Lobby**

15:15-16:45 **Session 13 – Confidentiality and Disclosure Control – Chaudière AB**

Chair: Jean-Louis Tambay, Statistics Canada

- **(E) Application of Statistical Disclosure Control Methods to the Canadian Hospitals Injury Reporting and Prevention Program Database**
  Ann Brown, Statistics Canada
  Margaret Herbert, PHA Canada, Ottawa, Canada

- **(E) Masking for Discrete Variables**
  Myron J. Katzoff, National Center for Health Statistics, USA
  Jay J. Kim, National Center for Health Statistics, USA

- **(E) Creation of Public Use Micro-Data Files for the National Survey on Drug Use and Health (NSDUH)**
  Feng Yu, RTI International, USA
  Lanting Dai, RTI International, USA
  Moshe Feder, RTI International, USA
  James R. Chromy, RTI International, USA

- **(E) Confidentiality and Replication-based Variance Estimation**
  Wilson Lu, Simon Fraser University, Canada

Friday, November 3, 2006

15:15-16:45 **Session 14 – Issues in Designing and Implementing Health Surveys – Chaudière C**

Chair: Julie McAuley, Statistics Canada

- **(E) Methodological Issues in Measuring the Mental Health Of Children and Young People in Great Britain**
  Howard Meltzer, Office for National Statistics, London, England
  Amanda Wilmot, Office for National Statistics, London, England
  Abigail Dewar, Office for National Statistics, London, England

- **(E) Sampling the Maori Population in the New Zealand Health Survey**
  Robert Clark, University of Wollongong, NSW, Australia
  Sarah Gerritsen, Public Health Intelligence, Ministry of Health, New Zealand

- **(E) Development And Testing Of A Caregiver-Proxy Child Health Questionnaire For The New Zealand Health Survey**
  Sarah Gerritsen, Public Health Intelligence, Ministry of Health, New Zealand

- **(E) Designing Questions to Identify People with Disabilities in Labor Force Surveys: The Effort to Measure the Employment Level of Adults with Disabilities in the CPS**
  Terence M. McMenamin, Division of Labour Force Statistics, Bureau of Labour Statistics, USA

16:45-17:00 **(E,F) Closing Remarks – Chaudière C**

Don Royce, Statistics Canada

Poster Sessions

Poster Session A

- **(E) Simulation Research on Hot Deck and Multiple Imputation Methods Using HCSDB Data**
  Donsig Jang, Mathematica Policy Research. Inc., USA
  Amang Sukasih, Mathematica Policy Research. Inc., USA
  Xiaojing Lin, Mathematica Policy Research. Inc., USA

- **(E) Estimating Sample Size for Complex Surveys: Building Consensus in an Environment of Multiple Hypotheses, Multiple Stakeholders, and Limited Budgets**
  JC Victor[1,2], L Diemert[1,2], S Bondy[1,2], KS Brown[1,3], J Cohen[1,2], R Ferrence[1,2], PW McDonald[1,3], P Selby[1,4], T Stephens[1]
  [1]Ontario Tobacco Research Unit; [2]University of Toronto;
  [3]University of Waterloo; [4]Centre for Addiction and Mental Health

- **(E) The Effect of Model Specification on Multiply Imputed Data: Lessons Learned from the NIH DC-HOPE Study**
  Marie G. Gantz, NIH-DC Initiative, RTI International, USA
  M. Nabil El-Khorazaty, NIH-DC Initiative, RTI International, USA

- **(E) Validation of Qualitative Methodology Applied to a Multidimensional Instrument with an Open Question**
  José Eduardo Corrente, UNESP, Botucatu, São Paulo, Brazil.
  Cinthia Esbrile Moraes, UNESP, Botucatu, São Paulo, Brazil.
  Tania Ruiz, UNESP, Botucatu, São Paulo, Brazil.

Poster Session B

- **(E) Development and Design of the Ontario Tobacco Survey**
  S. Bondy[1,2], L Diemert[1,2], JC Victor[1,2], KS Brown[1,3], J Cohen[1,2], R Ferrence[1,2], PW McDonald[1,3], P Selby[1,4], T Stephens[1]
  [1]Ontario Tobacco Research Unit; [2]University of Toronto;
  [3]University of Waterloo; [4]Centre for Addiction and Mental Health

- **(F) A model for the estimation of life expectancy in small Canadian cities**
  Philippe Finès, Statistics Canada

- **(F) Myocardial infarction (MI) in Quebec: Are there differences between immigrants and native-born Canadians?**
  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

- **(E) Geovisualization of Health and Social Capital Data Derived from Statistics Canada Surveys**
  Daniel Rainham, University of Ottawa, Canada
  Daniel Krewski, University of Ottawa, Canada
  Ian McDowell, University of Ottawa, Canada
  Mike Sawada, University of Ottawa, Canada

Poster Session C

- **(E) The effects of different approaches to selecting hospital separation records for use in reporting on injuries**
  Susan G. Mackenzie, Public Health Agency of Canada

- **(E) Going from ICD 9 to ICD 10-CA in hospital health care studies**
  Christie Sambell, Statistics Canada
  Hude Quan, Statistics Canada
  Helen Johansen, Statistics Canada

- **(E) Exploring the Impact of Participant Reluctance on Data Quality in the National Health Interview Survey**
  James M. Dahlhamer, National Center for Health Statistics, U.S. Centers for Disease Control and Prevention, USA
  Catherine M. Simile, National Center for Health Statistics, U.S. Centers for Disease Control and Prevention, USA
  Net Taylor, National Center for Health Statistics, U.S. Centers for Disease Control and Prevention, USA

- **(E) Impact of unconditional incentives on a multistage survey: The Health Survey for England**
  C. Deverill, National Centre for Social Research, London, UK
  H. Wardle, National Centre for Social Research, London, UK
  R. Craig, National Centre for Social Research, London, UK

- **(E) A comparison of self-reported primary mental health care utilization in the Canadian Community Health Survey with respondents' provincial health insurance records**
  JoAnne Palin, University of British Columbia, Vancouver, BC, Canada

---

ABSTRACTS

Workshops

(E,F) Record Linkage in Studies of Population Health – An Overview

Karla Fox, Department of National Defense and Patricia Whitridge, Elections Canada
(French and English presentation with simultaneous translation and with English and French materials)

Record linkage is simply the integration of information from two or more independent sources. Under this framework, records are linked on the basis of common data. Record linkage has become increasingly important in many different domains: maintenance of electronic registry information, health care administration, demographic studies and medical research. Whether following a cohort over time or linking patients to vital statistics in order to calculate survival curves, researchers need to understand record linkage.

Using examples from different applications in the health field, this workshop will cover the general principles of record linkage (statistical matching and exact matching), including data preparation, linkage techniques and linkage evaluation. Practical examples will be used to illustrate different points throughout the course. Participants will gain an appreciation for the different concepts and principles involved and an understanding of how to approach a problem requiring record linkage. At the end, students will be provided with a comprehensive list of references in the field of record linkage.

(E) Methods for Analyzing Longitudinal Health Survey Data – Theory and Applications

Mary Thompson, University of Waterloo, Waterloo, Canada (English presentation with simultaneous translation and with French and English materials)

The workshop will begin with an overview of models for longitudinal data, with examples from the health sciences. Repeated measures models considered will include growth curve, GEE, and stochastic process models. Event history models will include survival and recurrent models. Types of explanatory variate (fixed, time dependent, internal, external) will be reviewed, as well as types of missing data. The overview will be followed by a discussion of observational studies and causality, through examples illustrating the use of longitudinal or quasi-longitudinal data in examining causal hypotheses. The role of path diagrams in modeling dependencies will be considered. Principles of adapting the models and methods to complex survey data will be outlined, with examples from Canadian health survey data. Problems presented by models with complex likelihoods (particularly those arising from missingness or latency) will be illustrated. The workshop will end with a discussion and limited environmental scan of software, available or known to be under development, for longitudinal health survey data.

Session 1 - Keynote Address

(E) Measuring Health in Population Surveys
Graham Kalton, Westat Inc., Washington, DC, USA

Information on the population's health is collected both in surveys that focus on health and, because of the key role of health in other aspects of life, in many surveys on other topics. The subject-matter of population health surveys broadly encompasses physical and mental heath, dental health, disabilities, substance abuse, health risk factors, food intake, health promotion, health care utilization and quality, coverage and costs. Some surveys focus on specific health conditions, whereas others aim to obtain an overall health assessment. Health is often an important component in longitudinal studies, particularly in birth and aging cohorts. Information about health can be collected by respondents' reports (for themselves and sometimes for others), by medical examinations, and by collecting biological specimens. There is a serious concern about the accuracy of health information collected by respondents' reports. Logistical issues, cost considerations, and respondent cooperation feature prominently when the information is collected by medical examination. Ethical and privacy issues are often important, particularly when DNA and biomarkers are involved. International comparability of health measures is of growing importance. This paper will review the methodology for a range of health surveys and will discuss the challenges in obtaining accurate data in this field.

Session 2 – Current Issues in Small Area Estimation with Application to Health Surveys

(E) Mixed Linear-Non-linear Models for Small Area Estimation with Application to the Canadian Community Health Survey
Avi Singh and François Verret, Statistics Canada, Ottawa, ON, Canada

We consider the problem of modeling direct estimates of population counts with characteristics of interest for small domains or areas from survey data by borrowing strength over areas to improve precision. The commonly used linear mixed model (LMM) at the aggregate (or domain) level of Fay and Herriot (1979) is appealing to practitioners because of its simplicity in terms of both computation and interpretation. In dealing with multivariate data over time, computational benefits may be marked because LMM allows for recursive filtering techniques such as Kalman by ranking the domains in an ad hoc manner and treating the rank as the time dimension. The resulting estimates have the property of being best unbiased predictors (BUP) under normality of direct estimates. However, the optimality may be

tenuous if the parameters of interest need to satisfy range restrictions. For example, for discrete data such as counts or proportions, it is natural to use a nonlinear or generalized LMM to ensure that the domain estimates satisfy suitable range restrictions. To this end, using the hierarchical Bayes HB approach, GLMM at the aggregate level for survey data can be used with the help of MCMC for computing optimal estimates as in You and Rao (2002). In this paper, we propose a simpler alternative, analogous to LMM, by avoiding the need of HB formulation. This is done through a mixed linear-non-linear (MLNL) model where the random part of the mean is linear and additive to the fixed nonlinear part or the marginal mean. The additive structure of random effects can be justified from specification of the marginal error covariance structure in the presence of cluster sampling or a prior on the cluster means as is typically done in defining over-dispersion. Under approximate normality of the direct estimates and using MLNL modeling, best linearized and approximately unbiased predictors (BLAUP) can be defined as a direct generalization of BUP. The BLAUP is likely to satisfy range restrictions as the marginal means and the direct estimates do, and the impact of other correlated direct estimates on the estimated random effect is likely to be small. It is observed that MLNL modeling can provide substantial improvements over LMM while retaining its simplicity. An application to Canadian Community Health Survey is considered for estimating the proportions with certain health characteristics for subpopulations such as health regions by age-gender within a province.

(E) Dealing with Influential Observations and Outliers in Small Area Estimation

William R. Bell and Elizabeth T. Huang, U.S. Census Bureau, Washington, DC, USA

Small area estimation using linear area level models gives small area predictions that are weighted averages of the direct survey estimates and the regression predications. In this setting it is of interest to examine data points that are influential in determining the regression fits and model predictions. Influence measures for ordinary least squares regression, such as Cook's (1977) distance, have been generalized to mixed linear models as used in small area estimation (Rao 2003). We apply such measures to models from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program, along with a Mahalanobis distance measure applied directly to the regression matrix. Outliers can also be a concern, but may be difficult to detect given the high level of sampling error in the direct survey estimates (that motivated small area estimation in the first place). We illustrate use of a t-distribution to allow for possible outliers in a bivariate model for U.S. state poverty rates, examining the effects on

model predictions and associated estimates of prediction mean squared errors.

(E) Robust Bayesian Predictive Inference for the Finite Population Quantile of a Small Area

Balgobin Nandram[1] and Jai Won Choi[2,1] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, USA,[2] Office of Research and Methodology, National Center for Health Statistics, CDC, Hyattsville, MD, USA

Using a robust Bayesian method, we consider the problem of analyzing data with nonignorable nonresponse. Given parameters, we assume that the response indicators are independent Bernoulli random variables, and the response variable is distribution-free. Here predictive inference is required for the finite population quantile of a sub-area (an area within a small area). It is assumed that a random sample is available from a finite population, covariates are also available for everyone in the finite population, and these covariates can adequately explain the difference between respondents and nonrespondents. A robust logistic regression model is used to relate the response indicators to the covariates for the sample. This is obtained by expanding the standard logistic regression model to a mixture of Student's t distributions, thereby providing propensity scores which are used to construct adjustment cells. The nonrespondents are filled in by drawing samples from the response variable of the respondents within the adjustment cells. Prediction uses a linear rank-based regression of the response variable on the covariates by areas, thereby further robustifying our method. We use Markov chain Monte Carlo (MCMC) methods to fit our model. Within each sub-area inference about the quantiles of the response variable is made using the order statistics over all the individuals (sampled and nonsampled), and its distribution is obtained using the iterates from the MCMC sampler. Our application is on the third National Health and Nutrition Examination survey in which the response variable is body mass index (BMI), and age, race and sex are the covariates; these quantities are available for thirty five counties (small areas). Bayesian predictive inference of the finite population BMI quantiles (85th percentile for overweight and the 95th percentile for obesity) for each age-race-sex domain (sub-area) within each county is of interest. We compare our robust method with a recent parametric method.

(E) Robust Estimation of the Mean Squared Error of an EBLUP of a Small Area Mean

Partha Lahiri, University of Maryland, College Park MD, USA

In this talk, we present a general method for estimating the mean square error (MSE) of EBLUP for the well-known Fay-Herriot small

area model. We first obtain a second-order approximation to the MSE, i.e., an approximation which ignores all terms of the order o(m-1), where m denotes the number of small areas. Unlike the normality-based approximations, our approximation involves the kurtosis of both the sampling and model errors and depends on the method of estimating variance component. We note that for the method of moments (MOM) estimator of the variance component, we do not require the estimation of kurtosis of the model error in order to obtain a second-order unbiased MSE estimator. This extends an earlier result of Lahiri and Rao (1995) to the situation when the sampling errors are non-normal. However, in order to obtain a second-order unbiased MSE estimator, the estimation of the kurtosis is necessary for other methods of estimating variance components, e.g., the method of estimating equation proposed by Fay and Herriot (1979). We propose a method for estimating the kurtosis which in tern provides a method of MSE estimation for a very general class of variance component estimators for the non-normality of both the sampling and model errors. Examples from disease mapping and health surveys will be used to illustrate the methodology.

Session 3 – Analysis of Health Data - Applications

(F) The Methodological Challenges Related to the Analysis of Nutritional Data From The Nutrition Survey

François Verret, Statistics Canada, Ottawa, ON, Canada

The main objective of the Canadian Nutrition Survey is to estimate, for a number of fields of interest at the provincial level, the distributions of usual consumption of different nutrients required for good health. Nutritional information was collected from a sample of more than 36,000 Canadians of all ages, via a 24-hour food logbook with a second recall.

The most common way to produce these estimates is to process the data by adjusting a repeated measures model. The model allows us to determine the relative size of the variance of the nutritional contribution between individuals (inter-individual variability) and the daily variance for individuals (intra-individual variability). One of the main problems is the possibility of obtaining negative estimates of inter-individual variance. Clearly, this is an obstacle to estimating the distributions of interest.

Also, since the sampling variability is estimated with the bootstrap replication method, a model must be adjusted for each replication. The effect of this is to cause or amplify certain problems. Among other things, the run time for the Software for Intake Distribution (SIDE), a

commercial software for estimating distributions, is substantial. This poses a challenge for users and may call into question the number of replications to use. Also, negative estimates of variance are obtained for some bootstrap replications while positive ones are obtained for others.

This article will describe the methodological challenges and the solutions considered for solving the problems caused by applying the repeated measures model (using SIDE) to survey data from a complex sampling plan where the bootstrap replication method is used to measure sampling variability.

(F) Estimating Child BMI Growth Curves for Canada

Stéphane Tremblay and Rebecca Morrison, ON, Statistics Canada

Growth curves are used by Health Care providers to determine if the growth of a child or foetus (for example) is within normal boundaries. The current charts used in Canada for height, weight and body mass index (BMI) are based on United States data. Child growth curves can now be developed based on the most recent data available in Canada. One of the methods that will be explored is called the Lambda-Mu-Sigma (LMS) method for estimating and plotting growth curves. This method has been used in different studies by the World Health Organisation, the United Kingdom and the United States to develop growth curves for children. The LMS method will be used to estimate percentile growth curves for BMI using weighted cross-sectional data. Adapted methods for smaller sample sizes will also be explored. This presentation will focus on the BMI of children, one of the most common anthropometric measures to assess growth and obesity. BMI measurements are available from surveys such as the Canadian Community Health Survey cycles 2.2 and 3.1, and the 1978-79 Canada Health Survey. To increase the number of observations, hence increase the precision of curve estimates, different sources of data will be combined for the empirical part of this study.

(E) Obesity Rate Differences Between U.S. and Canadian Women and Between U.S. and Canadian Men: Findings From The Joint Canada / United States Survey of Health

Jane F. Gentleman, Debra L. Blackwell, and Michael E. Martinez, National Center for Health Statistics, Division of Health Interview Statistics, Hyattsville, MD, USA

The 2002-03 Joint Canada/United States Survey of Health was a telephone survey conducted jointly by Statistics Canada and the U.S. National Center for Health Statistics. Essentially the same questionnaire was administered in both countries at the same time, yielding a data set that provided unprecedented comparability of

national estimates from the two countries. Previous findings from that data set included evidence of appreciable differences in obesity rates between U.S. and Canadian women but not between U.S. and Canadian men. Specifically, 20% of U.S. women were obese, compared with only 12% of Canadian women, whereas the rates for U.S. and Canadian men were, respectively, 20% and 18%. This study further investigates factors associated with obesity in the two countries, as measured by body mass index. Results show that not only are obesity rates similar for men in the two countries, but the distributions of body mass index are almost identical. Multivariate analysis indicates that factors significantly associated with being obese for U.S. women but not significant (or less so) for Canadian women include age, race/ethnicity, education, nativity (immigrant or not), and having a regular medical doctor. For example, the odds of Non-Hispanic black U.S. women being obese were 2.6 times the odds of Non-Hispanic white U.S. women being obese, whereas there was no comparable significant difference for Canadian women. Also, the odds of non-native-born U.S. women being obese were 36% lower than the odds of native-born U.S. women being obese, whereas nativity was not significantly related to obesity for Canadian women.

(F) Studying Scenarios of Nutrition Intervention: The Example of Soft Drinks

Didier Garriguet, Health Statistics Division, Statistics Canada, Ottawa, ON, Canada

With data from the Canadian Nutrition Survey, it is possible to conduct a number of analyses: measuring the average daily or annual consumption of a food item, measuring the usual distribution of nutrients so as to identify at-risk populations that do not follow the recommendations of the Institute of Medicine, or measuring the usual distribution of food groups and comparing them with Canada's Food Guide to Healthy Eating. In short, it is possible to assess the nutritional status of the population.

For foods items, usual consumption is obtained by multiplying the total consumption of the food item by its probability of consumption in the population. To obtain usual consumption, the intra-individual variation must be removed from the daily consumption distribution. The probability of consuming a food item is obtained by looking at the frequency of consumption over several days. However, with only two daily measures of data, it is difficult, if not impossible, to evaluate this probability for certain foods.

In this article, we will explore other approaches for quantifying the importance of a food item in the diet of the population. We will study various scenarios for the probability of consumption. We will also study

intervention scenarios; in other words, we will attempt to measure the impact of the consumption of a food item by either eliminating this item from Canadians' diets, or by replacing it with various alternatives and comparing nutritional contributions under these various scenarios.

To illustrate these analytical methods, the example of soft drinks will be used. Their consumption is sufficient to obtain a distribution of the usual contribution by consumers, although it is not sufficient to obtain the probability of consumption.

Session 4 – Issues in Analysis of Longitudinal Health Data

(E) Measurement Error in Life History Data

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

In medical studies it often happens that some collected data are subject to measurement error. Sometimes covariates (or risk factors) of interest may be difficult to observe precisely due to physical location or cost. Sometimes it is impossible to measure covariates accurately due to the nature of the covariates. In other situations, a covariate may represent an average of a certain quantity over time, and any practical way of measuring such a quantity necessarily features measurement error. When carrying out statistical inference in such settings, it is important to account for the effects of mismeasured covariates; otherwise, erroneous or even misleading results may be produced. In this talk, I will discuss statistical analysis methods for life history data with mismeasured covariates. Specifically, I will focus on functional methods that are widely used in practice. Some applications will also be reported.

(E) Flexible methods for analyzing longitudinal data in population health

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

The study of longitudinal data (or data collected over time) is vital in terms of accurately observing changes in responses of interest for individuals, communities, and larger populations over time. Linear mixed effects models (for continuous responses observed over time) and generalized linear mixed effects models and generalized estimating equations (for more general responses such as binary or count data observed over time) are the most popular techniques used for analyzing longitudinal data from health studies, though, as with all modeling techniques, these approaches have limitations, partly due to their underlying assumptions. In this presentation, I will discuss some advances in these areas which makes modeling longitudinal data more

flexible. Some of these advances involve using curve-based techniques. I will present some examples from the health literature utilizing these more flexible procedures, with the goal of demonstrating that some otherwise difficult questions can be answered in terms of analyzing complex longitudinal data in population health studies.

(E) Complex Sampling Design Based Familial Longitudinal Health Data Analysis: An Overview

Brajendra Sutradhar, Memorial University of Newfoundland, St. John's, NL, Canada

In health studies, it is quite common to collect binary or count repeated responses along with a set of multi-dimensional covariates over a small period of time from a large number of independent families, where the families are selected from a finite population by using certain complex sampling designs. It is of interest to examine the effects of the covariates on the familial longitudinal responses after taking the variation in the family effects as well as the longitudinal correlations of the repeated responses into account. In this paper, I review the advantages and drawbacks of the existing methodologies for the estimation of the regression effects, the variance of the family effects and the longitudinal correlations. We then outline the advantages of a new unified generalized quasilikelihood approach in analyzing the complex design based familial longitudinal data. Some existing numerical studies are discussed as illustrations of the methodologies considered in the paper.

(E) Estimation of Attributable Number of Deaths and Standard Errors from Simple and Complex Sampled Cohorts

Barry I. Graubard[1], Mitchell H. Gail[1], Katherine M. Flegal [2], David F. Williamson[3],[1] National Cancer Institute, Bethesda, MD,[2] National Center for Health Statistics, Hyattsville, MD,[3] Centers for Disease Control and Prevention, Atlanta, GA, USA

Estimates of the attributable number of deaths (AD) from all-causes can be obtained by first estimating population attributable risk (AR) adjusted for confounding covariates, and then multiplying the AR by the number of deaths determined from vital mortality statistics that occurred for a specific time period. Proportional hazard regression estimates of adjusted relative hazards obtained from mortality follow-up data from a cohort or a survey is combined with a joint distribution of risk factor and confounding covariates to compute an adjusted AR. Two estimators of adjusted AR are examined, which differ according to the reference population that the joint distribution of risk factor and confounders is obtained. The two types of reference populations considered: (i) the population that is represented by the baseline cohort and (ii) a population that is external to the cohort. Methods based on influence

function theory are applied to obtain expressions for estimating the variance of the AD estimator. These variance estimators can be applied to data that range from simple random samples to (sample) weighted multi-stage stratified cluster samples from national household surveys. The variance estimation of AD is illustrated in an analysis of excess deaths due to having a non-ideal body mass index using data from the second National Health and Examination Survey (NHANES) Mortality Study and the 1999-2002 NHANES. These methods can also be used to estimate the attributable number of cause-specific deaths or incident cases of a disease and their standard errors when the time period for the accrual of is short.

Session 5 – Combining Data from Different Sources

(E) Quality Studies to Evaluate Linking Data with the Population Census

Glenys Bishop, Analytical Services Branch, ABS, Canberra, Australia

The Australian Bureau of Statistics (ABS) will begin the formation of a Statistical Longitudinal Census Data Set (SLCD) by choosing a 5% sample of people from the 2006 population census to be linked with subsequent censuses. Names and addresses will not be retained beyond the census processing period. However, during this period we will perform various linking exercises using names and addresses. These have been termed Quality Studies.

There are two types of Quality Studies. The first type is to assess the feasibilty and likely quality of linkage between data sets when names and addresses are no longer available. The second type is link Census data with other data to improve ABS statistical outputs. This paper will focus on the first type and will present quality study methods that we have used and preliminary results from linking with Census Dress Rehearsal data.

A long-term aim is to use the power of the rich longitudinal demographic data provided by the SLCD to shed light on a variety of issues by linking it with births, deaths, immigration settlements or disease registers. The initial quality studies will determine how well we can perform those linkages.

(E) Merging Area-Level Census Data with Survey and Administrative Data

Denis Gonthier, Tina Hotton, Cynthia Cook, and Russell Wilkins, Statistics Canada, Ottawa, ON, Canada

This paper explains how to append census area-level summary data to survey or administrative data. It uses examples from survey datasets present in Statistics Canada Research Data Centres, but the methods also apply to external datasets, including administrative datasets. Four examples illustrate common situations faced by researchers: (1) when the survey (or administrative) and census data both contain the same level of geographic identifiers, coded to the same year standard ("vintage") of census geography (for example, if both have 2001 DA); (2) when the two files contain geographic identifiers of the same vintage, but at different levels of census geography (for example, 1996 EA in the survey, but 1996 CT in the census data); (3) when the two files contain data coded to different vintages of census geography (such as 1996 EA for the survey, but 2001 DA for the census); (4) when the survey data are lacking in geographic identifiers, and those identifiers must first be generated from postal codes present on the file. The examples are shown using SAS syntax, but the principles apply to other programming languages or statistical packages.

(E) A Case Study in Using Model-Assisted Estimation to Integrate Survey and Administrative Data

Robert E. Fay, U.S. Census Bureau, Washington, DC, USA

Statistics Canada has been a world leader in the application of model-assisted estimation to government statistics. For example, introduction of this approach into sample estimation for the 1991 Census of Canada was an important milestone. A number of other countries have also applied these methods, and they are now also gaining some traction in the U.S.

This paper reports research in progress intended to introduce model-assisted estimation into the American Community Survey (ACS), a large-scale ongoing survey intended to replace the long-form sample data in the U.S. decennial censuses. One feature of the proposed application is quite possibly of general interest—the approach to integrate information from administrative records into the estimation. From the perspective of data quality, the use of model-assisted estimation allows the Census Bureau to focus on the quality of the ACS data collected directly from respondents, while depending only on regression relationships with the administrative data to reduce variance. Thus, the issues of data quality of the administrative records need only be assessed through their effect on variance. Although the ACS is a general-purpose survey not specifically tied to health, this

case study may suggest possible applications in areas of health statistics to combine survey and administrative data.

(E) Exact and Probabilistic Record Linkage

Claude Nadeau, Marie P. Beaudet, Jocelyne Marion, Christel Le Petit, Statistics Canada, Ottawa, ON, Canada

Record linkage across databases can be used to increase the range of measures available for analysis, allow for a check on the degree of agreement between two measures of a construct and may allow for the inclusion of individuals who tend to be non-participants in surveys because of time constraints, ill health or refusals.

The results of two studies in which data files have been linked using both an exact and a probabilistic approach will be presented. One study explores record linkage between a sample of 1991 Census and Health and Activity Limitation Survey (HALS) records, 1990 and 1991 summary T1 tax file data, and death registrations from 1991 to 2001.

The purpose of the study is to develop a set of baseline indicators of mortality for monitoring health disparities in Canada and socio-economic inequalities in health. The other illustrates the linkage between the 2003 Canadian Community Health Survey and the Health Person Oriented Information database. This study examines the role of modifiable risks factors on the likelihood of hospitalization for heart disease.

Following an overview of the rules for linking and the number of records linked with each method, a comparison of the characteristics of the individuals included in each database will be presented. As well, the results of multivariate analyses contrasting the findings when records were obtained with exact and probabilistic record linkage methods will be examined to document the effect of a potential increase in power and in the sample's range of the resulting linked database. Limitations associated with these types of linkages will be discussed.

Session 6 – Issues in Weighting, Estimation and Outlier Detection

(E) Variable Selection Models for Weight Trimming

Michael Elliott, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA

In unequal-probability-of-selection sample, correlations between the probability of selection and the sampled data can induce bias. Weights equal to the inverse of the probability of selection are often used to

counteract this bias. Highly disproportional sample designs have large weights, which can introduce unnecessary variability in statistics such as the population mean estimate. Weight trimming reduces large weights to a fixed cutpoint value and adjusts weights below this value to maintain the untrimmed weight sum. This reduces variability at the cost of introducing some bias. Standard approaches are not "data-driven": they do not use the data to make the appropriate bias-variance trade-off, or else do so in a highly inefficient fashion. This presentation develops Bayesian variable selection methods for weight trimming to supplement standard, ad-hoc design-based methods in disproportional probability-of-inclusion designs where variances due to sample weights exceeds bias correction. These methods are used to estimate linear and generalized linear regression model population parameters in the context of stratified and poststratified known-probability sample designs. Applications will be considered in the context of traffic injury survey data, in which highly disproportional sample designs are often utilized.

(E) Combining Information from Two Surveys to Improve on Analyses of Self-Reported Data in Estimating Measures of Health
Nathaniel Schenker, Trivellore E. Raghunathan, and Irina Bondarenko, National Center for Health Statistics, Hyattsville, MD, USA

Despite advances that have improved the health of the United States population, disparities in health remain among various racial/ethnic and socio-economic groups. Common data sources for assessing the health of a population of interest include large-scale surveys that often pose questions requiring a self-report, such as, "Has a doctor or other health professional ever told you that you have <health condition of interest>?" Answers to such questions might not always reflect the true prevalences of health conditions (for example, if a respondent does not have access to a doctor or other health professional). Similarly, self-reported data on quantities such as height and weight might be subject to reporting errors. Such "measurement error" in health data could affect inferences about measures of health and health disparities. In this work, we fit measurement-error models to data from the National Health and Nutrition Examination Survey, which asks self-report questions during an interview component and also obtains physical measurements during an examination component. We then develop methods for using the fitted models to improve on analyses of self-reported data from another survey that does not include an examination component. The methods, which involve multiply imputing examination-based data values for the survey that has only self-reported data, are applied to the National Health Interview Survey in examples involving diabetes, hypertension, and obesity. Preliminary results suggest that the adjustments for measurement error can result in non-negligible changes in estimates of measures of health.

(E) Evaluation of Methods for Outlier Detection and Treatment in the U.S. Survey of Occupational Illnesses and Injuries

John L. Eltinge, Bureau of Labour Statistics, Washington, DC, USA

The U.S. Survey of Occupational Illnesses and Injuries (SOII) is a large-scale establishment survey conducted by the Bureau of Labor Statistics to measure incidence rates and impact of occupational illnesses and injuries within specified industries at the national and state levels. This survey currently uses relatively simple procedures for detection and treatment of outliers. The outlier-detection methods center on comparison of reported establishment-level incidence rates to the corresponding distribution of reports within specified cells defined by the intersection of state and industry classifications. The treatment methods involve replacement of standard probability weights with a weight set equal to one, followed by a benchmark adjustment.

One could use more complex methods for detection and treatment of outliers for the SOII, e.g., detection methods that use influence functions, probability weights and multivariate observations; or treatment methods based on Winsorization or M-estimation. Evaluation of the practical benefits of these more complex methods requires one to consider three important factors. First, severe outliers are relatively rare, but when they occur, they may have a severe impact on SOII estimators in cells defined by the intersection of states and industries. Consequently, practical evaluation of the impact of outlier methods focuses primarily on the tails of the distributions of estimators, rather than standard aggregate performance measures like variance or mean squared error. Second, the analytic and data-based evaluations focus on the incremental improvement obtained through use of the more complex methods, relative to the performance of the simple methods currently in place. Third, development of the abovementioned tools requires somewhat nonstandard asymptotics the reflect trade-offs in effects associated with, respectively, increasing sample sizes; increasing numbers of publication cells; and changing tails of underlying distributions of observations.

(E) Combining Cycles of the Canadian Community Health Survey

Steven Thomas, Statistics Canada

The Canadian Community Health Survey (CCHS) is a sample survey that consists of two cross-sectional surveys conducted on an alternating annual cycle. The first cycle collects information from over 130,000 respondents for production of general health information at the health region level. The second cycle collects information from over 30,000 respondents for production of survey specific health estimates

at the provincial level. Even with such large sample sizes, users are interested in combining the cycles of the CCHS to improve the quality of the estimates, create estimates for small geographical domains, or to estimate for rare populations. Methods exist for combining surveys to increase sample size and hence the precision of estimates. However, in the general context, this is to combine surveys that measure the exact same thing. The CCHS cycles do not measure the exact same thing otherwise the repetition of cycles would be redundant. Therefore, the combining of cycles of the CCHS or any series of cross-sectional surveys can not be performed blindly. The researcher must have a clear understanding of the goal in combining cycles of the CCHS. This includes a clear understanding of the population being covered and the parameter to be measured. This paper will focus on some of the issues related to combining cycles of the CCHS including some possible interpretations of the combined result. Possible methods to combine cycles will also be outlined.

Session 7 – Direct Health Measures Surveys: Problems and Possibilities

(E) Learning the Unique and Peculiar Challenges of Direct Health Measures Surveys: The Canadian Experience

Mark Tremblay, Statistics Canada, Ottawa, ON, Canada

The Canadian Health Measures Survey (CHMS) represents Statistics Canada's first substantial health survey employing direct physical measurements of health. The CHMS will be collecting directly measured health data on a representative sample of 5000 Canadians aged 6-79 in 2007-08. After a comprehensive in-home health interview, respondents will report to a mobile examination centre (MEC; 53 foot tractor trailers) where direct health measures will be performed. The MEC will be set-up for 6 weeks in each of the 15 cluster sites selected. Approximately 350 respondents will be measured per site by 30 field and interview staff that will travel around the country with the MEC. Measures will include fitness tests, anthropometry, objective physical activity monitoring, spirometry, blood pressure measurements, oral health measures and blood and urine sampling. Blood and urine will be analysed for measures of chronic disease, infectious disease, nutritional indicators and environmental biomarkers. This survey has many unique and peculiar challenges rarely experienced by most Statistics Canada surveys. These include: a new level of respondent burden (travel, time, expense, physical exertion, pain); data transfer complexity (interviewer-clinic-lab-Statistics Canada-respondent report); privacy and ethical considerations (age range for testing, consent, data confidentiality, participant anonymity, storing biospecimens for future analysis, collecting DNA samples); communications (reporting to

respondents, reportable diseases, media, public); and the potential for adverse events (phlebitis, cardiac event during fitness testing) or adverse findings (previously undiagnosed infectious disease, oral cancer). Though the challenges are many, the data collected through the CHMS will be unique and will provide a valuable health surveillance and research resource for Canada.

(E) The U.S. National Health and Nutrition Examination Survey: What 40 Years of Experience has Taught Us

Cliff Johnson, National Centre for Health Statistics, Hyattsville, MD, USA

The National Health and Nutrition Examination Survey (NHANES), has been conducted by the National Center for Health Statistics (part of the U.S. Centers for Disease Control and Prevention (CDC)) for over forty years. The survey is unique in that it measures the health and nutritional status of the U.S. non-institutionalized population through interviews and direct physical examinations. During the course of those forty plus years, many lessons have been learned. These lessons include a variety of methodological issues related to survey design, survey content, survey operations, data quality, and data analysis. Examples of "lessons learned" will be presented for each of these topic areas. These examples include: (1) the use of mobile examination centers versus "fixed sites" for data collection; (2) what needs to be done before changing to a new data collection methodology for an ongoing survey component; (3) the need to document (for posterity) every change in survey design and data collection procedures; (4) that data quality is a direct result of thorough and repetitive training of staff and pilot testing of new procedures; and (5) the need to be flexible and willing to change (just because you have been doing the survey one way for forty years does not mean there isn't a better way to do it). The challenges have been many, mistakes have (and will) occur, and the lessons learned are not finished. Regardless, the significance of employing direct physical measurements of health as a surveillance and research tool is well worth all the methodological challenges that have been and will be confronted.

(E) Direct Health Measures Surveys in Finland : From North Karelia and Beyond

Arpo Aromaa, KTL, Helsinki, Finland

National health examination surveys (HES) were initiated in Finland by the Mobile Clinic Unit (MCU) of the Social Insurance Institution (SII) in 1965. Carried out in various parts of the country on close to 60 000 adults, these studies comprised a wide variety of measurements as

well as dietary interviews. A more restricted version of the survey was carried out as part of the North Karelia project in 1972 and has since been continued every five years.

A national comprehensive survey, called Mini-Finland, was carried out in 1978 – 1980 by SII. The MCU examined over 7200 (of 8000) persons in 40 regions selected to represent the whole country. A wide range of interviews and measurements as well as clinical examinations by dentists and physicians was carried out. Major chronic diseases and functional capacity were central objects. The population has been followed-up by various registers.

Coordinated by KTL a new comprehensive survey, Health 2000, was carried out in 2000 and 2001 in 80 regions representative of Finland. The sample comprised 10 000 people aged 18 or over. Health examinations were restricted to those 8 028 aged 30 and over. The main aims were to examine health status and functional capacity, their determinants and distribution, and to assess changes after the previous 1978 – 1980 study. After an extensive interview the participants were asked to fill in and bring with them to the examination a lengthy questionnaire. The core programme comprised anthropometrics and blood pressure measurement, heel ultrasound, bioimpedance, blood samples (plasma and whole blood for DNA) and saliva samples, digital dental x-ray and a dentist's examination, various tests of functioning (physical and cognitive), a physician's structured clinical examination, and finally a psychiatric interview (M-CIDI). Functioning was assessed by questionnaires, observation and computerized tests.

Many of the findings and methods can be viewed on our trilingual web-site (www.ktl.fi/health2000). As part of the Finnish comprehensive health survey system, the next similar survey will probably be carried out in five years time as Health 2012.

Session 8 - Waksberg Award Address

(E) The Analysis of Population-based Case Control Studies

Alastair Scott, University of Auckland, New Zealand

The use of complex sampling designs is becoming increasingly common in population-based case-control studies, usually for obtaining the controls but sometimes for the cases as well. The topic is particularly appropriate for a session honouring Joe Waksberg, since it was his insistence on the importance of having genuinely representative controls that led to the widespread adoption of standard survey methodology in this area.

Most statistical packages now have specialist survey modules that can be used to carry out weighted analyses of data from such studies routinely. However, weighting tends to be inefficient in situations like this where differences among the weights are usually very large. Fully efficient likelihood methods can be developed for some special designs but, apart from the case of simple stratified sampling, these methods require special software and are difficult to implement. Moreover, there are questions about their robustness to model breakdown. We examine the question of robustness in more detail and suggest alternative procedures that are reasonably efficient and can be implemented using standard survey software.

Session 9 – Methods for Analysis of Health Data

(E) Bootstrap Methods for Analyzing Complex Sample Survey Data

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

Data obtained from complex sample surveys often involve multi-stage cluster sampling, leading to dependencies among sample elements, and unequal probabilities of selection, leading to unequal sampling weights. Also, sampling weights are often calibrated to known population totals of auxiliary variables as well as subject to non-response adjustments. As a result, application of standard methods to survey data without accounting for the design features and weight adjustments can lead to erroneous inferences. Bootstrap methods offer an attractive option to the analyst for taking account of the survey design. The data file consists of the full-sample final weights and associated bootstrap final weights for a large number of bootstrap replicates as well as the observed data on the sample elements. We show how such data files can be used to analyze survey data in a straightforward manner using weighted estimating equations. A one-step estimating function (EF) bootstrap method that avoids some difficulties with the bootstrap is also discussed. We show how the methodology can be applied to both cross-sectional and longitudinal binary response data using logistic linear regression models. Formal goodness-of-fit tests that take account of the survey design are also presented. Data from the National Public Health Survey of Canada are used to illustrate the proposed methods.

(E) Estimation of Regression Parameters with Survey Data

Wayne A. Fuller and Yu Wu, Iowa State University, Ames, IA, USA

The coefficients of regression equations are often parameters of interest for health surveys. Such surveys are sometimes of complex design with differential sampling rates. We give estimators for the regression coefficients for complex surveys that are superior to ordinary

expansion estimators under the subject matter model, but also retain desirable design properties. Theoretical and Monte Carlo properties are presented.

(E) Addressing Data Sparseness in Contextual Population Health Research: Effects of Small Group Size and Targeted Cluster Analysis On Multilevel Model Performance, Bias And Efficiency Cohorts

Philippa Clarke[1], Patricia O'Campo[2] and Blair Wheaton[2,1] University of Michigan, Ann Arbor, MI, USA,[2] University of Toronto, Toronto, ON, Canada

The current use of multilevel models to examine the effects of the surrounding context (neighborhood) on health outcomes attest to their value as a statistical method for analyzing grouped data. But the use of multilevel modeling with data from population-based surveys is often limited by the small number of cases per level-2 unit, prompting a recent trend in the neighborhood literature to apply cluster analysis techniques to address the problem of data sparseness. In this paper we use Monte Carlo simulations to investigate the effects of marginal group sizes on multilevel model performance, bias, and efficiency. We then employ cluster analysis techniques to minimize data sparseness and examine the consequences in the simulations. We identify the minimum boundary at which data sparseness becomes a concern for valid estimates in contextual models. We find that estimates of the fixed effects are robust at the extremes of data sparseness, and cluster analysis is an effective strategy to increase group size and prevent the upwards bias in the variance components that occurs with increasing sparseness. However, clustering strategies introduce artificial within-group heterogeneity, resulting in spuriously weak group effects (Type II error) due to a minimization of between-group differences. Caution should therefore be exercised when using clustering techniques in order to maintain the balance of within-group to between-group variance in the data.

(E) Causal Inference in Observational Studies using Health Administrative Data

Therese A Stukel, ICES, Toronto, ON, Canada

Health services research generally relies on observational data to compare outcomes of patients receiving different therapies. Comparisons of patient groups in observational studies may be biased, in that outcomes differ due to both the effects of treatment and the effects of patient prognosis. In some cases, especially when data are collected on detailed clinical risk factors, these differences can be controlled for using statistical or epidemiological methods. In other cases, when unmeasured characteristics of the patient population

affect both the decision to provide therapy and the outcome, these differences cannot be removed using standard techniques. Use of health administrative data requires particular cautions in undertaking observational studies since important clinical information does not exist. We discuss several statistical and epidemiological approaches to remove overt (measurable) and hidden (unmeasurable) bias in observational studies. These include regression model-based case-mix adjustment, propensity-based matching, redefining the exposure variable of interest, and the econometric technique of instrumental variable (IV) analysis. These methods are illustrated using examples from the medical literature including prediction of one-year mortality following heart attack; the return to health care spending in higher spending U.S. regions in terms of clinical and financial benefits; and the long-term survival benefits of invasive cardiac management of heart attack patients. It is possible to use health administrative data for observational studies provided careful attention is paid to addressing issues of reverse causation and unmeasured confounding.

Session 10 – Collection and Mode Effects

(E) Wireless Substitution in the United States and Canada : Prevalence and Impact on Random-Digit-Dialed Health Surveys

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

Canada had nearly 15 million wireless telephones in 2004. That is nearly one wireless telephone for every two persons in Canada. In the United States, more than half of all telephones are now wireless, and the average number of minutes of use per month on wireless telephones is greater than the average number of minutes per person per month logged on residential landline telephones. It is perhaps not surprising, then, that many wireless telephone users in both countries have considered "cutting the cord"—substituting their residential landline telephone with a wireless telephone.

Most major survey research organizations in Canada and the United States do not include wireless telephone numbers when conducting random-digit-dialed (RDD) household telephone surveys. The inability to reach adults who have only wireless telephones (or who have no telephone service) has potential implications for health research conducted using RDD surveys.

This presentation will first highlight the most up-to-date estimates available from Statistics Canada and the U.S. National Center for Health Statistics concerning the size and characteristics of the wireless-

only population. Canadian estimates will be drawn from the Residential Telephone Service Survey, a supplement to the Labour Force Survey. Estimates for the United States will be drawn from the National Health Interview Survey.

Next, to better understand the implications of wireless substitution for RDD household health surveys, this presentation will highlight National Health Interview Survey estimates on the health and health care access of wireless-only adults. Finally, we will discuss whether demographic differences between wireless-only adults and adults with landline telephones can largely explain observed differences in health and health care access.

(E) Exploring the Impact of Mode on Key Health Estimates in the National Health Interview Survey

Catherine Simile, Barbara Stussman, James Dahlhamer, National Center for Health Statistics, Hyattsville, MD, USA

In an effort to increase response rates and decrease costs, many survey operations have begun to use several modes to collect relevant data. While the National Health Interview Survey (NHIS), a multipurpose household health survey conducted annually by the National Center for Health Statistics, U. S. Centers for Disease Control and Prevention is primarily a face-to-face survey, NHIS interviewers sometimes rely on the telephone to complete interviews. This has raised questions about the comparability of resulting data. For example, do questionnaire items that usually require hand cards in face-to-face interviews produce different responses when adapted for use over the telephone? Also, many studies have indicated that using a different mode may produce different results on items considered sensitive. Do estimates on items considered sensitive differ depending on whether they are collected by telephone or face-to-face in the NHIS?

To address these questions, analyses are performed with contact attempt history and core data from the 2005 NHIS. We compare estimates that result from two specific types of questionnaire items: 1) items that may work well in face-to-face interviews, but may be problematic over the phone (e.g., those with hand cards), and 2) items considered sensitive (e.g., income, sexually transmitted diseases). The implications of our findings on current field operations and on the quality of NHIS estimates will be discussed.

(E) Impact of Telephone versus Face to Face Repeat 24 Hour Recall Interviews on Food and Nutrition Surveys

S. Hayward, E. Junkins, M. Vigneault, M. Villeneuve, Food Directorate, Health Canada, Ottawa, ON, Canada

Many population surveys collecting food consumption data use 24 hour recall methodology to capture detailed one day intakes. In order to estimate longer term intakes of foods and nutrients from these data, methods have been developed that required a repeat recall to be collected from at least a subset of responders in order to estimate day to day variability During the Canadian Community Health Survey Cycle 2.2 Nutrition Focus Survey, most first interviews were collected in person and repeat interviews were conducted by telephone. As part of the survey design, the repeat interviews were done in person for a subset of responders to allow assessment of in the impact of mode of interview. This paper looks at the impact of the mode of the repeat interview on the reported foods and nutrients on the repeat day and the estimation of intra individual variability between the first and the second interviews.

(F) How To Account For An Important Change In Methodology When Analysisng Data From The Canadian Community Health Survey?

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

The Canadian Community Health Survey (CCHS) is a general health survey of approximately 130 health regions across Canada. Statistics Canada carried out cycle 1.1 of the CCHS in 2000-2001 and cycle 2.1 in 2003.

The CCHS uses two types of survey frames: area frames and telephone frames. In cycle 1.1, approximately 95% of the Quebec weighted sample came from the area frame. In cycle 2.1, that proportion was less than 50%. Face-to-face collection is mainly used for area frame samples, whereas the collection method associated with the telephone frame is exclusively via telephone surveys. Between these two cycles, the proportion of the sample from the telephone frame increased.

Findings of the St-Pierre and Béland study (ASA conference, 2004) on the impact of collection methods indicate differences affecting several of the survey variables, revealing problems in data comparability between the two cycles. Caution is therefore recommended when using this data. The comparability issue was raised early on by Quebec users: Can the data in the two cycles be compared? Which variables can be compared?

In order to answer users' questions ─ and not to find the reason for the differences between the data in the two cycles ─ the Institut de la statistique du Québec (ISQ) looked into the main methodological change in the most recent survey, that is, the change in sample distribution based on the survey frame. For certain variables, this change in methodology adversely affects comparability between the two cycles. The ISQ has developed guidelines so that Quebec users can apply the same analysis methods, according to whether or not the variables can be compared.

The purpose of this paper is to describe the selected methodology and the results.

Session 11 – Sampling – Theory and Applications

(E) Simulation-based Systematic PPS Sampling With Unit Substitution

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

The International Tobacco Control (ITC) China Survey uses a multi-stage unequal probability sampling design with upper level clusters selected by the systematic PPS sampling method. A difficulty arises in the execution of the survey: several selected upper level clusters refuse to participate in the survey and have to be replaced by substitute units, selected from the remaining units once again using the systematic PPS sampling method. Under such a scenario (assuming random refusal) the first order inclusion probabilities of the final selected units are very difficult to calculate and the second order inclusion probabilities become virtually intractable. We develop a simulation-based approach for computing the first and second order inclusion probabilities and also implement the method using the statistical software R/SPLUS. The efficiency and effectiveness of our proposed approach are demonstrated through numerical examples. The approach can be extended to handle more complex refusal situations one may encounter in practice.

(E) On Reuse of Clusters in Repeated Studies

Stanislav Kolenikov[1] and Gustavo Angeles[2,1] University of Missouri, Columbia, MS,[2] University of North Carolina, Carolina Population Center, Chapel Hill, NC, USA

Suppose data for a survey with multistage design is to be collected in two periods of time. This paper assesses the relative merits of keeping the same clusters in the sample vs. sampling new clusters, under differential statistical (correlation between clusters and overtime) and logistical (costs of survey) scenarios. The design effect of reusing the

same clusters from the master sample over time is of the form $1 - A\rho\pi/n$ where ρ is intertemporal correlation of the cluster totals, n is the number of clusters, π is the proportion of clusters retained from the previous round, and $A > 0$ is a fixed constant. As long as the efficiency gains appear to be minor, the value of the designs that reuse the clusters comes from the logistical (cost of the survey) considerations. Empirical demonstrations that use Demographic and Health Survey (DHS) data for Bangladesh, 1996 and 2000, and a Monte Carlo simulation, are provided.

(F) Sampling Design for the Canadian Health Measures Survey

Suzelle Giroux, Statistics Canada, Ottawa, ON, Canada

The objectives of the brand new Canadian Health Measures Survey (CHMS) are to estimate the prevalence and distribution of directly measured health indicator such as overweight and obesity, physical activity, nutrition, chronic disease risk factors and exposure to environmental pollutants and infectious agents.

To meet these ambitious objectives, a nationally representative sample of roughly 5,000 Canadians, both females and males, aged 6 to 79, will participate in the CHMS. Respondents will be asked to answer a questionnaire in their home, and also attend the CHMS clinic to have physical health measurements taken by professionals such as height, weight, blood pressure, fitness tests and blood and urine specimens for laboratory analysis. The survey is expected to start in February 2007 and will take place over a period of two years.

The presentation will focus on the sample design developed to meet the objectives and the logistics of the survey. Among the challenges is the necessity to have selected respondents within a reasonable commuting distance to the clinics, the difficulty to reach the desired sample size for youths and sub-sampling of measures related to the exposure to environmental pollutants. The sample design will propose solutions to these challenges, among those are the creation of collection sites, the use of multiple survey frames and the person selection strategy.

(E) Challenges in the Design of the National Health and Nutrition Examination Survey

Leyla Mohadjer, Westat Inc. Washington, DC, USA

The National Health and Nutrition Examination Surveys (NHANES) are one of the series of health-related programs sponsored by the United States National Center for Health Statistics. A unique feature of NHANES is the administration of a complete medical examination for

each respondent in the sample. To standardize administration, these examinations are carried out in mobile examination centers (MECs). The examination includes physical measurements, tests such as eye and dental examinations, and the collection of blood and urine specimens for laboratory testing.

NHANES is an ongoing annual health survey of the noninstitutionalized civilian population of the United States. The analytic goals of NHANES include; estimating the number and percentage of persons in the U.S. population and in designated subgroups with selected diseases and risk factors; monitoring trends in the prevalence, awareness, treatment, and control of selected diseases; monitoring trends in risk behaviors and environmental exposures; analyzing risk factors for selected diseases; studying the relationship between diet, nutrition, and health; and exploring emerging public health issues and new technologies.

The sample design for NHANES needs to create a balance between the requirements for efficient annual and multiyear samples and the flexibility that allows changes in key design parameters to make the survey more responsive to the needs of the research and health policy communities. This paper discusses the challenges involved in designing and implementing a sample selection process that satisfies the goals of NHANES.

Session 12 – Health Measurement: An International Perspective

(E) Measuring the Health of Populations: The Conceptual and Analytic Approach of The Global Burden of Disease Study

Colin D. Mathers, Department of Measurement and Health Information, World Health Organization, Geneva, Switzerland

In the last two decades, considerable international effort has been put into the development of summary measures of population health that integrate information of mortality and non-fatal health outcomes and international policy interest in such indicators is increasing. There are two main classes of summary measures of population health: health gaps and health expectancies. The Disability-Adjusted Life Year (DALY) is the best known health gap measure and quantifies the gap between a population's actual health and a normative health goal, defined in terms of a global standard life table specifying the healthy years of life lost due to a death at any given age.

This paper gives an overview of the Global Burden of Disease (GBD) conceptual framework, the relationship of the DALY to other measures of population health, and the GBD analytical approach, with particular

attention to issues in (1) dealing with biased and missing data, (2) dealing with uncertainty and (3) specific technical issues in ensuring cross-population comparability. The latter include dealing with variations in quality and completeness of cause of death information, explicit use of a comprehensive framework and internal consistency checks for improving comparability of estimates of incidence, prevalence and mortality for causes, the assessment of disability weights, and techniques for improving the comparability of the assessment of the disease burden attributable to risk factors.

(E) International Comparisons In Measuring Health States: Experiences From The World Health Surveys

Somnath Chatterji, Department of Measurement and Health Information, World Health Organization, Geneva, Switzerland

During the last three decades, there has been general acceptance of an approach to describing health states of individuals in terms of multiple domains of health, and in developing self-report instruments that seek information on each of these domains. A health state is thus a multi-dimensional attribute of an individual that reflects his or her levels on the various components or domains of health. Thus, a health state differs from pathology, risk factors or etiology, and from health service encounters or interventions.

How to describe health states, is a central challenge in undertaking the measurement of health. The relationship of health states to other aspects of health such as future non-fatal health outcomes or risk of mortality need to be examined. The way people report their own health varies consistently with factors such as education, sex, age, or other cultural factors. Various people use different response category cut-points across cultures or population sub-groups, and this 'response shift' implies that self-report categorical data are not comparable across individuals. The responses cannot be directly used to measure health without adjustment.

In recognition of this the WHO World Health Surveys (WHS), used a set of questions across a core set of domains to measure health states and employed vignettes to detect and correct for biases in self-report in order to adjust for response category cut-point shifts. This paper will describe the instrument used in the WHS and the methods used to provide cross population comparable data. It will present results from the WHS demonstrating the existence of systematic reporting biases, the ability of respondents to rate vignettes and their use to adjust for biases in order to make data more comparable. Future strategies to address these problems will be discussed.

(E) A Research Agenda for Developing Comparable Measures of Functional Health Status for use in Population Surveys
Chris Murray, Harvard University, Boston, MA, USA

Session 13 – Confidentiality and Disclosure Control

(E) Application of Statistical Disclosure Control Methods to the Canadian Hospitals Injury Reporting and Prevention Program Database
Ann Brown[1] and Margaret Herbert[2],[1] Statistics Canada,[2] PHA Canada, Ottawa, ON, Canada

In this paper we describe methods to develop a micro-data file suitable for public release of the Canadian Hospitals Injury Reporting and Prevention Program (CHIRPP) database. CHIRPP is a national injury surveillance database managed by the Public Health Agency of Canada (PHAC) and its purpose is to contribute to reducing the number and severity of injuries in Canada. CHIRPP contracts ten pediatric and four general hospitals located in the major cities of seven provinces and one territory to collect information on injuries sustained by patients reporting to their emergency departments. CHIRPP contains detailed information about patient characteristics and circumstances of injury events. CHIRPP is the first PHAC database to develop a micro-data version using statistical disclosure control methods.

Disclosure control methods were developed with a view to: maximize the amount of data released; rigorously protect personal privacy; and, develop a cost-effective, systematic approach for future updates. Challenges presented by CHIRPP included its skewed nature, level of detail, and sensitivity of injury event information (sometimes events are widely known in the community). The results include measures of suppression at record and global levels.

The paper concludes with a discussion of the implication of this work for the health information field within the context of increasing demand for micro-data, advances in technology, increasing sophistication of users, and the growing number of public and private data files.

(E) Masking for Discrete Variables
Myron J. Katzoff and Jay J. Kim, National Center for Health Statistics, Hyattsville, MD, USA

The use of discrete variables having known statistical distributions in the masking of data on discrete variables has been under study for some time. This paper presents further results from our research on this topic. The consequences of sampling with and without replacement from finite populations are one principal interest. Estimates of first and

second order moments which attenuate or adjust for the additional variation due to masking of known type are developed. The impact of masking of the original data on the correlation structure of concomitantly measured variables and on tests for statistical significance are of special interest. The need for further developments for undertaking analyses of multivariate data is discussed.

(E) Creation of Public Use Micro-Data Files for the National Survey on Drug Use and Health (NSDUH)

Feng Yu, Lanting Dai, Moshe Feder, and James R. Chromy, RTI International, Research Triangle Park, NC, USA

The National Survey on Drug Use and Health is an annual cross-sectional national household survey on the prevalence and correlates of drug use and mental health in the United States. Disclosure risk is a major concern in the creation of public use micro-data files (PUFs) because the survey data contain highly sensitive information about respondents' drug behavior and health issues. Therefore, strong confidentiality assurance is needed to protect respondents' privacy and to reduce respondents' perception of disclosure risk.

We discuss and assess both inside and outside intrusion threats. Our presentation focuses on our process of the creation of PUFs. The process is mainly based on, but not limited to, the MASSC methodology, which helps introduce sufficient uncertainty about the presence and identity of a respondent in the database for protecting against inside intrusion. To further protect against outside intrusion, we collapse categories with rare responses, recode variables with extreme values, suppress variables that have no analytic value but pose high disclosure risk, and treat the stratification variables.

The treatment process is concluded with assessment of the analytic utility of the treated database by comparing the PUF (approximately with 54,000 to 59,000 records) to the original data on a range of univariate and multivariate statistics and their standard errors.

(E) Confidentiality and Replication-based Variance Estimation

Wilson Lu, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada

Protecting against inadvertent data disclosure in publicly released data files from complex surveys is becoming increasingly important. Most of the literature on the topic considers the masking of the released data itself. One simple measure taken to help avoid data disclosure is to suppress the strata and PSU identifiers. This can make it difficult for

direct variance estimation as strata and PSU identifiers are typically needed to obtain asymptotically unbiased variance estimators.

One might believe that releasing a set of replicate weights together with the data which also have the PSU and strata identifiers suppressed would circumvent the problem. This is not the case. We show that by viewing the problem of reconstructing PSU and/or strata identifiers from replicate weights as a high-dimensional clustering problem, one can easily reconstruct PSU identifiers from any of the usual replication methods without knowledge of which replication method was used. Thus even unsophisticated users can quite easily obtain PSU identifiers from the set of replicate weights if great care is not taken.

A common method for directly masking the data in complex surveys is to alter data or swap data values between cases. In our setting, one might swap strata and/or PSU identifiers before creating the set of replicate weights. This would only impact the variance estimation and could be done so as to limit the impact on resulting variance estimates for the characteristics measured in the survey. We introduce two automatic algorithms for swapping PSU identifiers and evaluate the performance.

Session 14 – Issues in Designing and Implementing Health Surveys

(E) Methodological Issues in Measuring the Mental Health of Children and Young People in Great Britain

Howard Meltzer, Amanda Wilmot, Abigail Dewar, Office for National Statistics, London, England

In 1999, the first nationally representative survey of the mental health of children and young people aged 5-15 was carried out in Great Britain. A second survey was carried out in 2004. The aim of these surveys was threefold: to estimate the prevalence of mental disorders among young people, to look at their use of health, social and educational services, and to investigate risk factors associated with mental disorders. The achieved number of interviews was 10,500 and 8,000 respectively.

Some key questions had to be addressed on a large number of methodological issues;

How to obtain a representative sample of children aged 5-15?
Whether to use a one-phase or two-phase sampling design?
Whether to use a standard epidemiological assessment instrument or create a new one?

Information was required from three sources - children, their parents and teachers – what survey procedures were required for each source?
Which parent should be interviewed?
In what order should the interviews take place?
What were the specific issues involved in interviewing children and young people?
How to deal with respondents who could not speak English?
Both quantitative and qualitative data were required – what was the best way to organise the questionnaire for this purpose?
How to collate both quantitative and qualitative data from three sources?
How to analyse the data when there was only one or two sources?
What should be the level of clinical input and in what form?
Could the data be used to produce small area estimates?

The factors taken into account to reach decisions on all these issues are discussed.

(E) Sampling the Maori Population in the New Zealand Health Survey

Robert Clark[1] and Sarah Gerritsen[2,1] University of Wollongong, Wollongong, NSW, Australia,[2] Public Health Intelligence, Ministry of Health, Wellington, New Zealand

Achieving sufficient sample size for ethnic subpopulations is a common challenge in population survey sample design. About 11% of New Zealand's population belongs to the indigenous Maori population. This ethnic group is geographically clustered to some extent, but even so most Maori live in areas which have relatively low proportions of Maori, making it difficult to sample this group efficiently. In the 2006/2007 New Zealand Health Survey design, two strategies for sampling Maori, Pacific Peoples and Asians have been utilised to achieve sufficient sample size while maintaining low design effects. The first strategy is light targeting, where dwellings in regions with higher proportions of Maori are given higher probabilities of selections. The second strategy is to select (a) a core sample where people are selected without reference to their ethnicity, and (b) a large screening sample where ethnicity of each household member is collected and only those in ethnic groups of interest are eligible for the full survey. The trade-off between screening and targeting will be described for the New Zealand Health Survey, and theory will be presented for optimising this trade-off in general.

(E) Development And Testing Of A Caregiver-Proxy Child Health Questionnaire For The New Zealand Health Survey
Sarah Gerritsen, Public Health Intelligence, Ministry of Health, New Zealand

The New Zealand Ministry of Health has revised its population health survey, the New Zealand Health Survey (NZHS) to include a questionnaire specifically on child health. In addition to collecting information face-to-face from an adult in sample households, the 2006/07 NZHS will randomly select a person under the age of 15 years, if there are any, in each selected household and collect additional information from the primary caregiver about the child.
The principal aim of the NZHS child questionnaire is to collect new health data which can be used for monitoring population-level child health status, health service utilisation, and the health risk and protective behaviours that have their origins in childhood. Previously, only data collected through child contact with the health system, for example hospital administration records and disease/injury databases, have been available for monitoring child health in New Zealand.

This paper reviews the questionnaire development for the child health component of the 2006/07 New Zealand Health Survey, including topic selection, question development, cognitive-testing, preliminary sample design, final questionnaire drafting, and dress rehearsal testing. The resulting questionnaire will be of interest to government health researchers who currently collect caregiver-proxy child health data in surveys, and those who are considering adding this important element to their population health monitoring programme.

(E) Designing Questions to Identify People with Disabilities in Labor Force Surveys: The Effort to Measure the Employment Level of Adults with Disabilities in the CPS
Terence M. McMenamin, Division of Labour Force Statistics, Bureau of Labour Statistics, Washington, DC, USA

In 1998, Executive Order 13078 mandated the development of an accurate and reliable measure of the employment rate of people with disabilities, to be published as frequently as possible. The Bureau of Labor Statistics, in cooperation with the Employment Rate Measurement Methodology interagency workgroup, identified the goal of placing a small set of questions within the Current Population Survey. A set of potential questions was drawn from existing surveys, cognitively tested, and placed in the National Comorbidity Survey (NCS) for a field test. The BLS analyzed the test data to determine the disability status of respondents, using a Dephi process for difficult cases. A variety of tests were used to identify which small set of questions best approximated the results of the initial analysis. The BLS

then cognitively tested the set of questions to identify any comprehension problems respondents might have in a CPS context. In February 2006, the disability question set was added to a split panel of CPS respondents. An analysis of these data will determine whether the questions work in a labor force survey, as well as whether the addition of this set of questions will adversely affect the CPS.

Poster Session A

(E) Simulation Research on Hot Deck and Multiple Imputation Methods Using HCSDB Data

Donsig Jang, Amang Sukasih, and Xiaojing Lin, Mathematica Policy Research. Inc., Washington, DC, USA

Currently, the Health Care Survey of DoD Beneficiaries (HCSDB) does not impute any item missing values for its data analysis. Rather, it excludes all cases with any item missing data from analysis and thus results in a reduction of effective sample sizes for analysis. While geographic areas defined as catchment areas have long been important domains of analysis by researchers and policy makers who are using the HCSDB data, small sample sizes, low response rates, and item missing data might keep them from producing statistically reliable estimates at the catchment area-level. Imputation of item missing data can help not only reduce the bias of the estimate but also enhance estimation capacity in terms of domains covered.

Choosing an imputation method among the methods available, however, may involve non-technical considerations. We consider two imputation methods: a single imputation method--the within-class nearest neighbor hot-deck imputation, and a multiple imputation method--the sequential regression multivariate imputation (SRMI). We plan to perform a simulation study that would allow us to quantify the extent of bias resulting from nonresponse in a few selected items, and then compare results from the two imputation methods above. The reported data values for those items will be treated as the “true” values, and we will simulate the item nonresponses by generating approximately 10-15 percent missing values among these “true” values using different nonresponse mechanisms. Results from this simulation study may be used to make a decision about which method should be implemented for the HCSDB imputation.

(E) Estimating Sample Size for Complex Surveys: Building Consensus in an Environment of Multiple Hypotheses, Multiple Stakeholders, and Limited Budgets

JC Victor[1,2], L Diemert[1,2], S Bondy[1,2], KS Brown[1,3], J Cohen[1,2], R Ferrence[1,2], PW McDonald[1,3], P Selby[1,4], T Stephens[1]
[1]Ontario Tobacco Research Unit; [2]University of Toronto; [3]University of Waterloo; [4]Centre for Addiction and Mental Health

Rationale: Sample size estimation is critical to study design. For studies involving small numbers of dependent and independent variables of interest, this is readily performed using commercial software. However, how does one estimate sample size while building consensus among multiple stakeholders for multi-purpose projects such as complex population-based surveys? We present a demonstration template using common spreadsheet software to provide real-time estimates of power, precision, and financial costs under varying sample size scenarios, as was used in development of the Ontario Tobacco Survey.

Methods: The Ontario Tobacco Survey is a longitudinal survey of smokers and cross-sectional survey of smokers and non-smokers. Survey design, hypotheses and questions were developed by consensus of investigators, whose needs ranged from cross-sectional trend estimates to repeated-measures analyses.

Results: Regionally-stratified estimates of smoking prevalence and incident smoking-related outcomes from cross-sectional surveys were used to calculate resulting numbers of events for study under different sampling strategies. Appropriate complex sample size formulae, incorporating loss-to-follow-up and design effects, were nested within an Excel spreadsheet. The utility displayed power for several cross-sectional and longitudinal analyses, and total survey cost based on user-input per-interview costs.

Conclusions: Developing a complex survey, reporting multiple population estimates and testing multiple hypotheses while being cost-effective, requires a multi-faceted approach to sample size estimation. Common spreadsheet software, and graphical displays were an important adjunct to complex formulae to facilitate the dialogue between methodologists and stakeholders. The utility helped demystify the 'sample size black box' and facilitated decision-making tradeoffs in design elements.

(E) The Effect of Model Specification on Multiply Imputed Data: Lessons Learned from the NIH DC-HOPE Study

Marie G. Gantz, Ph.D. and M. Nabil El-Khorazaty, Ph.D. for the NIH-DC Initiative, RTI International, Research Triangle Park, North Carolina, USA

Background: Multiple imputation is often used to fill-in missing data in longitudinal studies; however, assumptions made in the imputation model can greatly influence imputed values. We compare two versions of a multiply-imputed dataset, generated by models that differ in a key aspect. Study results based on the two imputed datasets are compared with results from the data prior to imputation.

Methods: The dataset comes from a study in which pregnant African American women in Washington, DC were recruited and randomized to either usual care or an educational and counseling integrated intervention to reduce risks related to poor pregnancy outcomes. Women were interviewed at four time points. Multiple imputation was performed twice: (1) for all participants together; (2) separately for women in each care group, thereby explicitly accounting for the effect of treatment assignment and interactions involving care group. In both imputations, number of intervention sessions attended was among the predictive variables, to control for intervention delivery.

Results: Analyses of both multiply-imputed datasets revealed fewer significant treatment effects than analysis of the pre-imputation data. Compared to imputing for all participants together, imputing for each care group separately resulted in more significant treatment effects. However, it potentially over-estimated the effect of treatment assignment by assuming that women in the intervention group with missing data behaved like other women in the intervention group and not like those in the usual care group.

Conclusions: The choice of imputation model can greatly impact the values of imputed data, and thus conclusions drawn from the study.

(E) Validation of Qualitative Methodology Applied to a Multidimensional Instrument with an Open Question

José Eduardo Corrente, Cinthia Esbrile Moraes and Tania Ruiz, UNESP, Botucatu, São Paulo, Brazil.

The Brazilian population has experienced an ageing process, thus characterizing an increase in the number of elderly people. Such population, nevertheless, should be given the opportunity to age with quality, since living for long periods implies the loss of autonomy and the onset of limitations. Therefore, instruments have been developed in order to measure the quality of life of elderly individuals. Hence, a

questionnaire consisting of various validated instruments and an open question was applied to a group of elderly citizens in the city of Botucatu, SP, Brazil. The analysis of the open question, assessed by qualitative methods, generated eleven categories concerning the elderly people's opinions as regards quality of life. A cluster analysis of such answers was carried out, and the answers were transformed into binary variables, thus producing three groups of elderly individuals. Proceeding with cluster analysis, other variables of the instrument were subjectively selected as those which most related to quality of life. Therefore, this work aimed at validating the categories obtained by the open question with the closed questions of the instrument by means of associations and application of chi-square tests at a level of significance of 5%. It was observed that qualitative analysis identifies phenomena regardless of category saturation. The quantitative method, on the other hand, shows the power of each category in a set, that is, as a whole. It was also observed that variables such as spirituality, rectitude, charity, knowledge and work are very subjective issues and difficult to evaluate by a multidimensional instrument. For such cases, more specific instruments must be designed.

Poster Session B

(E) Development and Design of the Ontario Tobacco Survey

S. Bondy[1,2], L Diemert[1,2], JC Victor[1,2], KS Brown[1,3], J Cohen[1,2], R Ferrence[1,2], PW McDonald[1,3], P Selby[1,4], T Stephens[1]
[1]Ontario Tobacco Research Unit; [2]University of Toronto; [3]University of Waterloo; [4]Centre for Addiction and Mental Health

Rationale:
Over the past 15 years, adult smoking prevalence has dropped to approximately 20% in Ontario. Subgroup analysis of smokers using population-based cross-sectional surveys often results in statistically unstable, or non-reportable estimates. This has been a barrier to studying the effect of anti-smoking campaigns and public policy. Longitudinal surveys, while providing increased power, often cannot provide estimates generalizable to the whole population. We discuss a methodology that addresses these barriers by combining both cross-sectional and longitudinal components in the Ontario Tobacco Survey (OTS).

Methods:
To address the needs of Ontario-based tobacco use researchers, a new study was needed to over-sample smokers, yet derive timely results on budget. The study design borrowed from NPHS methodology, a mixed approach of a rolling longitudinal component, coupled with a repeated cross-sectional component.

Results:
The OTS is a random digit dial cross-sectional survey of smokers and non-smokers with three follow-up interviews of smokers that began in July 2005. Smokers and non-smokers are asked identical questions on public policy and opinions; smokers' behaviour and attitudes are evaluated every six months. This design allows general population estimates related to public policy, while providing an abundance of data suitable for hypothesis testing regarding smoking cessation.

Conclusion:
A mixed panel and cross-sectional study results in analytic and weighting complications, and may still result in outcome specific estimates that are unreliable or not reportable. However, by restricting membership in the longitudinal component to a subgroup of interest, in our case smokers, we demonstrate that this design provides both statistical and financial efficiency.

(F) A model for the estimation of life expectancy in small Canadian cities
Philippe Finès, Statistics Canada, Ottawa, ON, Canada

Life expectancy at birth (LE) is frequently used as a health indicator. For large cities, computation of LE using deaths over three consecutive years produces a 95% confidence interval (CI) with a range of less than half a year. This estimate is precise enough to allow comparisons between cities and to evaluate the income-mortality gradient within a city. However, when a city's population is less than 100,000, the LE estimate becomes imprecise and unreliable. The objective of this work is to obtain simultaneously an accurate estimate of LE and a CI that is not too wide, even for these small cities. We can reach this objective if we take into account information available outside these small cities. For example, the general shape of the mortality curve – which is similar for all Canadian cities – may be used to help reduce the CI's width. The utilisation of such external information is the basis of "small area estimation" theory. In this presentation, we will show specific techniques developed to attain our objective of improving the precision of LE in small Canadian cities. In particular, we will present the key equations used as well as some interesting results.

(F) Myocardial infarction (MI) in Quebec: Are there differences between immigrants and native-born Canadians?
Maria Gabriela Orzanco, Alain Vanasse, Théophile Niyonsenga and Josiane Courteau, Université de Sherbrooke, Sherbrooke QC Canada

The purpose of this research project, which forms part of the "Système d'information spatio-temporel sur les maladies chroniques (SIST-MC)" program funded by the Geoïde network, is to develop an "immigration"

filter for the SIST-MC program, because immigrants have specific health characteristics. The research project seeks to determine whether there is an association between characteristics of Montréal's immigrant population and MI incidence and health outcomes. The research questions are as follows: (1) Do the rates of MI incidence, revascularization and survival in census tract (CT) populations with numerous immigrants differ from the rates in other CT populations? (2) Do these differences persist over time, and what factors are associated with changes? (3) What characteristics are most strongly associated with these health outcomes?

Units of analysis used in the research project: 1991, 1996 and 2001 CT populations. Variables studied: (1) rate of MI incidence; (2) rate of revascularization; and (3) rate of survival. Covariables: age; gender; co-morbidity; use of heart disease prevention drugs. Sources of data: (1) population censuses; and (2) Quebec's medical and administrative database, MED-ECHO. CT populations are described in terms of immigration, demographics and socio-economic status, and are classified into "immigration-disadvantage categories" (IDCs). IDCs are compared in terms of health outcomes over the three periods. The most important variables in these health outcomes will be identified using regression models.

(E) Geovisualization of Health and Social Capital Data Derived from Statistics Canada Surveys

Daniel Rainham, Daniel Krewski, Ian McDowell and Mike Sawada, University of Ottawa, Ottawa, ON, Canada

The classification and identification of locations where persons report to be more or less healthy or have more or less social capital, within a specific area such as a health region, is tremendously helpful for understanding place and health associations. The objective of the proposed study is to classify and map areas within the Zone 6 Health Region (Figure 1) of Nova Scotia (Halifax Regional Municipality and Annapolis Valley regions) according to health status (Dimension 1) and social capital (Dimension 2). We abstracted responses to questions about self-reported health status, mental health, and social capital from the master files of the Canadian Community Health Survey (Cycles 1.1, 1.2 and 2.1), National Population Health Survey (Cycle 5), and the General Social Survey (Cycles 13, 14, 17, and 18). Responses were geocoded using the Statistics Canada Postal Code Conversion File (PCCF+) and imported into a geographical information system (GIS) so that the postal code associated with the response will be assigned to a latitude and longitude within the Nova Scotia Zone 6 health region. Kernel density estimators and additional spatial interpolators were used to develop statistically-smoothed surfaces of the distribution of respondent values for each question. The smoothing process

eliminates the possibility of revealing individual respondent location and confidential Statistics Canada sampling frame information. Using responses from similar questions across multiple surveys improves the likelihood of detecting heterogeneity among the responses within the health region area, as well as the accuracy of the smoothed map classification.

Poster Session C

(E) The effects of different approaches to selecting hospital separation records for use in reporting on injuries

Susan G. Mackenzie, Public Health Agency of Canada, Ottawa, ON, Canada

Introduction:
Significant differences exist between the Public Health Agency of Canada (PHAC) and the Canadian Institute for Health Information (CIHI) in their approaches to selecting hospital separation records for reporting on injuries. Three other countries use approaches that differ from each other and from both Canadian approaches.

Objective:
To compare the effects of the five approaches.

Methods:
Hospital separation records from CIHI were used to prepare a set of 14,772 records with either an external cause of injury on the record or a tabulating diagnosis of injury from separations reported by one Canadian province in 2000-01. The data were subjected to all five approaches.

Results:
Of the selected records 87% included an external cause, 70% had an injury diagnosis, and 56% had both. The percentages of the14,772 records retained by the five approaches were: The Netherlands 87%, Canada (PHAC criteria) 73%, Canada (CIHI criteria) 64%, Australia 56%, the US 53%. The five approaches also led to important differences in the frequencies of some types of injuries. These will be described in the presentation.

Conclusions:
Standard approaches to selection of records for reports on external causes of injury would be useful.

(E) Going from ICD 9 to ICD 10-CA in hospital health care studies

Christie Sambell, Hude Quan and Helen Johansen, Statistics Canada, Ottawa, ON, Canada

The use of health administrative data for health research in Canada has been facilitated by their availability, their wide geographic coverage and their relatively complete capture of contacts with the health system for a defined population. However, definitions have to be unified in order to allow comparison between provinces and years. A key methodological tool is the International Classification of Diseases (ICD) by the World Health Organization (WHO) which gives a standardized format to code diagnoses, thereby enabling longitudinal and comparative studies. In 1992, the 10th Revision of ICD (ICD-10) was introduced and beginning in 2001, the Canadian provinces started to phase in a Canadian version of ICD-10 (ICD-10-CA) for coding hospital diagnoses. However, its implementation means that a number of established methodological tools applicable to ICD-9-CM, in particular the Charlson index and the Elixhauser comorbidities which give an estimation of disease severity need to be redesigned for application in ICD-10-CA. Our study has looked at rates for the above co-morbidities by province before and after the change over and by looking at different combinations of corrections suggests the most consistent way to identify the rates with new coding algorisms.

(E) Exploring the Impact of Participant Reluctance on Data Quality in the National Health Interview Survey

James M. Dahlhamer, Catherine M. Simile and Net Taylor, National Center for Health Statistics, U.S. Centers for Disease Control and Prevention

Government surveys in the United States have witnessed a gradual but steady decline in response rates over the past two decades, prompting data collectors to expend considerable resources to maximize completions through refusal conversions. While it is often presumed that such efforts will reduce nonresponse bias and other forms of survey error, refusal conversions may actually produce unintended, deleterious effects on data quality. For example, reluctant respondents may exhibit satisficing behaviors, i.e., they may have a tendency to provide the minimum response necessary to continue with the interview, which could result in elevated levels of item missingness and partially completed interviews. Do reluctant respondents contribute as much data as willing respondents? And are there differences by the type of reluctance communicated (e.g., “too busy,” “not interested,” “privacy concerns”)?

To address these questions, analyses are performed with contact attempt history, frame, and core data from the 2005 National Health Interview Survey, an in-person household health survey conducted annually by the National Center for Health Statistics, U.S. Centers for Disease Control and Prevention. More specifically, we perform comparisons of willing and reluctant participants on partial interview rates, and item refusal and don't know responses to a set of socioeconomic and critical health indicators. We then explore the impact of reluctance on these quality indicators in a multivariate context, controlling for sociodemographic and social environmental measures. The implications of our findings for current survey operations, indicators of data quality, and the mechanisms by which we communicate data quality to public users are discussed.

(E) Impact of unconditional incentives on a multistage survey: The Health Survey for England

C. Deverill, H. Wardle, R. Craig, National Centre for Social Research, London, UK

The Health Survey for England (HSE) is a continuous household survey, with interviewer and nurse stages, providing health and lifestyles data about the population. As with other surveys in Britain, household response rates have gradually fallen over the past 10 years. HSE response has decreased by 6 percentage points (p.p) and individual response to various stages of the survey have declined.

HSE uses a high quality pre-selected address sample, with advance letters sent to all selected addresses. A previous study found that co-operation rates were 3-5 p.p higher if people received a book of postage stamps with the advance letter. This was conducted in a single stage survey. Our experiment was devised to see if unconditional incentives of stamps or bookmarks sent with the advance letter would have an effect on both household response and response to other survey stages (i.e. nurse visit, anthropometric measurements). From July 2005, addresses were randomly allocated to one of three conditions: book of stamps, bookmark with relevant findings from HSE or advance letter only (control group).

Preliminary results show household response is 2-3 p.p higher among those who received the stamps, as opposed to the bookmark or the letter only. However, co-operation rates to subsequent stages of the survey (anthropometric measurements and nurse visit) were lowest among individuals in households who received the stamps.

These results suggest that unconditional incentives can improve household response rates, but co-operation is not carried through to

follow-up stages of the survey. Full findings from this experiment will be available in summer 2006.

(E) A comparison of self-reported primary mental health care utilization in the Canadian Community Health Survey with respondents' provincial health insurance records

JoAnne Palin, University of British Columbia, Vancouver, BC, Canada

Accurate information about the timing of access to primary mental health care is critically important in order to identify potentially modifiable factors which could facilitate timely and on-going management of care. No "gold standard" measure of mental health care utilization exists, so it useful to know how strengths, gaps, and limitations in different data sources influence study results. This study compares two population-wide measures of primary mental health care utilization data: the Canadian Community Health Survey of Mental Health and Well-being (CCHS, cycle 1.2) and provincial health insurance records in the province of British Columbia. It explores four questions: (1) Is 12-month prevalence of contacts with general practitioners for mental heath issues the same regardless of whether survey data or administrative data are used? (2) What is the level of agreement between the survey data and administrative data for having had any contact with a general practitioner for mental heath issues during the 12 month period before the survey interview? (3) Is the level of agreement constant throughout the 12-month period or does it decline over more distant sub-timeframes within the 12-month period? (4) What kinds of respondent characteristics, including mental disorders, are associated with agreement or lack of agreement?
The results of this study will provide useful information about how to use and interpret each measure of health care utilization. In addition, it will contribute to survey design research, and to research which aims to improve the methods for using administrative data for mental health services research.

Final Thoughts

Thank you for attending Statistics Canada's 23rd International Symposium on Methodological Issues, **"Methodological Issues in Measuring Population Health"**. We hope that you have found it valuable and beneficial.

Please complete the symposium evaluation form before leaving, and drop it into one of the boxes in the foyer near the exit.

Proceedings from this symposium will be available in CD format around October, 2007. All registered participants will receive a complimentary copy.

Next year, instead of organizing the Symposium, Statistics Canada is co-organizing the Third International Conference on Establishment Surveys (ICES III), June 18 - 21, 2007, in Montréal, Québec, Canada. We look forward to seeing you there. We will resume our Symposium series in 2008.

The 2006 Symposium Organizing Committee

Statistics Canada Statistique Canada

Canada