Survey Methodology

June 2005

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

June 2005

In This Issue

This issue of *Survey Methodology* is dedicated to Gordon J. Brackstone, who recently retired from Statistics Canada. He was Assistant Chief Statistician for the Informatics and Methodology field and had been chairman of the *Survey Methodology* management board since 1987. His continuous support to the journal has been marked by great insight and motivated by a constant desire to foster high standards of methodology practices. Further, he also authored several articles that appeared in the journal. We wish to express our extreme gratitude to Gordon J. Brackstone.

The current issue contains eight regular papers on a variety of topics, and three short communications. As mentioned in the previous issue of the journal, we are introducing a new Short Communications section in *Survey Methodology*. This section will contain shorter papers, typically around four pages. Possible topics of short communications include presentation of new ideas without the full development of a regular paper, brief reports of empirical work, and discussions or supplements related to other papers published in the journal.

For the past four years the June issue of *Survey Methodology* has included an invited paper in honour of Joseph Waksberg. Starting this year, this annual invited paper will be published in the December issue of the journal, bringing it more in line with the associated Waksberg address delivered at Statistics Canada's annual methodology symposium in the autumn. The author of this year's Waksberg paper is J.N.K. Rao and his paper will be on the "Interplay Between Sample Survey Theory and Methods: an Appraisal".

In the opening paper of this issue, Winglee, Valliant and Scheuren present a new simulation approach to estimation of error rates for threshold selection in record linkage. For each potential matched pair there is a vector of comparison outcomes that determines the linkage weight. A multinomial model is assumed for each comparison outcome, with different multinomial distributions for true matches and true non-matches. The distributions are estimated from a sample, and then used to simulate the distributions of the linkage weights for true matches and true non-matches. The method is illustrated in a case study using data from the U.S. Medical Expenditure Panel Survey (MEPS).

Krewski, Dewanji, Wang, Bartlett, Zielinski and Mallick investigate the effects of record linkage errors, both false positives and false negatives, on risk estimates in cohort studies. They show analytically how linkage errors introduce both bias and additional variability into observed and expected numbers of deaths, as well as into estimates of standardized mortality ratios and relative risk regression coefficients. They discuss their results in their conclusions, and point to further work that needs to be done in this area.

The paper by van den Brakel and Renssen addresses the problem of testing hypotheses between different survey implementations, such as different questionnaire designs, when a complex sampling design is used. A design-based theory is developed for cases where the survey implementations are assigned to subsamples through completely randomized experimental designs or randomized block experimental designs. The theory also makes use of measurement error models. Design-based Wald statistics are used to compare the different survey implementations.

Tsuchiya approaches the long-standing problem of asking respondents sensitive questions in an interesting fashion. Instead of using the randomized response approach that allows little control for the researcher, he proposes that the item count technique be adapted for sensitive questions. The item count technique presents the respondent with a list of several phrases, from which the respondent selects all that apply to him. The researcher constructs the list in two ways: the first list contains the sensitive phrase while the second list does not. Tsuchiya presents various estimators for this technique and gives an interesting example related to the Japanese national character.

In the paper by DiZio, Guarnera and Luzi, finite mixture models are used to detect errors that are due to an incorrect unit of measurement at the collection stage of the survey. In a multivariate context and assuming that the data are multivariate normal, the procedure can identify which variables are in error for a given sampled unit. The authors also provide diagnostics for prioritizing cases to be investigated more deeply through clerical review. The proposed methodology is illustrated through an example with simulated data and an example with real data.

Chiu, Yucel, Zanutto and Zaslavsky present a method for multiple imputation of missing contextual variables for use in regression analysis. For each record missing the variable, and for a sample of complete records, matched cases are selected based on a set of matching variables. The sample of complete records is then used to estimate a regression adjustment for other variables not included among the matching variables. The contextual variables for the incomplete records are then multiply imputed. The authors then show an application to a colorectal cancer study, and use simulations to compare their approach to three other nonresponse adjustment methods.

Nandram and Choi examine the important problem of nonignorable nonresponse in small-area estimation of a health status variable. When confronted with an example where the usual estimators are biased because of the excessive number of nonrespondents, they attempt to account for the differences through modeling. Nandram and Choi use two nonignorable nonresponse hierarchical Bayes models, a selection model and a pattern model, to analyze the health data. An important consideration to their modeling is the incorporation of the input from doctors concerning the nonresponse pattern and the outcome variable. The results give an accurate non-response adjustment and a better measure of precision.

Park and Fuller propose a method to reduce the probability of obtaining negative estimation weights when using a regression estimator. Their method consists of first approximating inclusion probabilities, conditional on Horvitz-Thompson estimates for a vector of auxiliary variables, and then using these approximate conditional inclusion probabilities as initial weights in a regression estimator. Their method is shown to work well in a simulation study. The weights obtained from this method are also compared to weights from quadratic programming, the raking ratio, the logit procedure and maximum likelihood.

In the first of three short communications included in this issue, Andersson and Thorburn show that the optimal regression estimator can be expressed as a calibration estimator with an appropriately chosen distance function. The resulting optimal estimator is asymptotically more efficient than the usual Generalized Regression (GREG) estimator. A small simulation study illustrates several situations where the optimal estimator if significantly more efficient than the GREG estimator.

Lynn and Gabler extend the results of Gabler, Hader and Lahiri (volume 25, 1999) on Kish's expression for the design effect due to clustering. They give a practical approach to estimating Kish's quantity at the sample design stage when only the total numbers of observations and of clusters are needed.

Meza and Lahiri examine the limitations of a standard regression model selection criterion, Mallows' statistic, for nested error regression models. They show, that while a straightforward application of Mallows' statistic may result inefficient model selection methods, a suitable transformation of the data may be the answer.

Finally, we would like to inform you that Harold Mantel will now hold the new position of Deputy Editor. Harold has been part of the Editorial Board for the last 15 years. His dedication to the journal has been notable and his continuous involvement in the editorial process has been instrumental in ensuring that *Survey Methodology* remains a high quality publication.

M.P. Singh