

the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 71

January 2015



International Statistical Institute



Institut International de Statistique



Editors

Natalie Shlomo and Eric Rancourt

Section Editors

- Pierre Lavallée** - Country Reports
Kennon Copeland - Ask the Experts
Mick Couper - New and Emerging Methods
Natalie Shlomo - Book and Software Review

Circulation/Production

Carole Jean-Marie
Henry Chiem
Courtney Williamson
Olivier Dupriez

The Survey Statistician is published twice a year by the International Association of Survey Statisticians and distributed to all its members. *The Survey Statistician* is also available on the IASS website at <http://isi.cbs.nl/iass/alluk.htm>.

Enquiries for membership in the Association or change of address for current members should be addressed to:

IASS Secretariat Membership Officer
 Margaret de Ruiter-Molloy
 International Statistical Institute
 P.O. Box 24070, 2490 AB The Hague,
 The Netherlands

Comments on the contents or suggestions for articles in *The Survey Statistician* should be sent via e-mail to the editors, Natalie Shlomo (natalie.shlomo@manchester.ac.uk) or Eric Rancourt (eric.rancourt@statcan.gc.ca).

In This Issue

- 5 Letter from the Editors**
- 7 Letter from the President**
- 14 Report from the Scientific Secretary**
- 16 News and Announcements**
- 21 Ask the Experts**
How Should I Run a Regression Based on a Complex Survey that Has Missing Item Values?
 By Phillip S. Kott
- 27 New and Emerging Methods:**
Paradata – Big Data of Survey Research
 By Frauke Kreuter
- 34 Book & Software Review:**
Register-based Statistics: Statistical Methods for Administrative Data, 2nd Edition (Chichester:Wiley)
 Book Review by Dan Hedlin, Stockholm University
- 42 Country Reports**
- Australia
 - Bosnia and Herzegovina
 - Brazil
 - Canada
 - Estonia
 - India
 - New Zealand
 - Spain
 - Switzerland
 - United Kingdom
- 52 Upcoming Conferences and Workshops**
- 67 In Other Journals**
- 80 Welcome New Members**
- 81 IASS Officers and Council Members**
- 82 Institutional Members**
- 83 Change of Address Form**



Letter from the Editors

Letter from the Editors January 2015 Issue:

The January 2015 issue of the IASS *The Survey Statistician* is the first issue for the new co-editor Eric Rancourt and assisted by Carole Jean-Marie from Statistics Canada and we welcome them both to the Editorial team. Natalie Shlomo is continuing as co-editor and is assisted by Henry Chiem and Courtney Williamson from the Australian Bureau of Statistics. We also welcome our new editor for the Ask the Expert Section, Kennon Copeland. This issue contains articles of interest and important information regarding upcoming conferences, journal contents, updates from the IASS Executive and more. We hope you enjoy this issue, and we would be happy to receive your feedback and comments on how we can make improvements.

This issue of *The Survey Statistician* includes the letter from our IASS President, Danny Pfeffermann containing important updates and in particular a draft strategic plan for the IASS which will be further developed in a working group composed of Geoff Lee (chair), Steve Heeringa, Daniela Cocchi and Natalie Shlomo. Please read the draft carefully and send comments to Geoff Lee (geoff.lee99@bigpond.com). We also have included the letter from our IASS Scientific Secretary, Mick Couper, on supported IASS conferences and recent travel awards funded by the ISI and the World Bank to the Statistics Canada Methodology Symposium which was held in October 2014.

In the New and Emerging Methods Section (edited by the Scientific Secretary Mick Couper), Frauke Kreuter from the University of Maryland, the Ludwig-Maximilians-Universität in Munich, Germany and head of the Statistical Methods Research Department at the Institute for Employment Research (IAB) in Nürnberg, Germany, has contributed an article titled 'Paradata – Big Data of Survey Research'. In the article, she addresses paradata as a by-product of the survey production process, what are the uses of paradata and why it has the characteristics of Big Data. She concludes with some of the international attention that this important area of research has generated.

In the Ask the Experts Section (edited by Kennon Copeland), Phil Kott from RTI International in the US, answers the question: How Should I Run a Regression Based on a Complex Survey that Has Missing Item Values? Phil addresses methods for dealing with missing data: listwise deletion, imputation, multiple imputation, maximum likelihood and adding a dummy variable to denote a missing covariate.

For the Book and Software Review Section, Dan Hedlin from [Stockholms Universitet](#) in Sweden has contributed a review of the book: Register-based Statistics: Statistical Methods for Administrative Data (Wiley Series in Survey Methodology), 2nd edition (2014) authored by Anders Wallgren and Britt Wallgren.

Please let Mick Couper (mcouper@umich.edu) know if you would like to contribute to the New and Emerging Methods Section in the future. If you have any questions

which you would like to be answered by an expert, please send them to Kennon Copeland (copeland-kennon@norc.org). If you are interested in writing a book or software review, please get in touch with Natalie Shlomo (Natalie.Shlomo@manchester.ac.uk).

The Country Report Section has always been a central feature of the IASS *The Survey Statistician* and we thank all country representatives for their contribution and coordination of the reports. We also thank the editor of the section, Pierre Lavallee (Pierre.Lavallee@statcan.gc.ca) for his continuing efforts to obtain timely reports from the different countries. We ask all country representatives to please share information on your country's current activities, applications, research and developments in survey methods. Please contact Geoff Lee (geoff.lee99@bigpond.com) if there is any change or addition to country representatives.

In the News and Announcement section we have an announcement from Susie Fortier, Production Manager of Survey Methodology journal on the 2014 Waksberg award winner Constance F. Citro who presented her paper at Statistics Canada's Methodology Symposium. We also have a letter from one of those receiving the travel award, Mr. Diego Andres Perez Ruiz from Mexico, currently a Ph.D. student in Statistics at the University of Manchester, School of Mathematics, who shares his experiences in being able to attend the Statistics Canada Methodology Symposium. Another News and Announcement item is on the Partnership in Statistics for Development in the 21st Century ([PARIS21](#)). This is a global partnership of national, regional and international statistics experts and decision makers guided by a board of international stakeholders with representatives from developing countries, bilateral donors, multilateral institutions and key partners.

We thank Marcel Vieira for putting together the list of conferences for inclusion in the newsletter. Please send to Marcel (Marcel.Vieira@ice.ufjf.br) any conference announcements that you would like advertised in the next *Survey Statistician* to be issued in July 2015. We also thank Carole Jean-Marie from Statistics Canada for collating the advertisements of upcoming conferences and for preparing the tables of contents in the In Other Journals section.

As always, we have many thanks for everyone working hard to put *The Survey Statistician* together, and in particular to Carole Jean-Marie from Statistics Canada, and Henry Chiem and Courtney Williamson from the Australian Bureau of Statistics for their invaluable assistance.

Please take an active role in supporting the IASS newsletter by volunteering to contribute articles, book/software reviews and country reports. We also ask IASS members to send in notifications about conferences and other important news items about their organizations or individual members.

The Survey Statistician is available for downloading from the IASS website at <http://isi.cbs.nl/iass/allUK.htm>.

Eric Rancourt Eric.Rancourt@statcan.gc.ca

Natalie Shlomo Natalie.Shlomo@manchester.ac.uk

Letter from the President

Dear Colleagues,

Here comes my third "letter from the President". I thought that it would be a short one, but somehow it is long again. As in my previous letters, I shall mostly report on what we did during the last 6 months, and what is planned, until and during the ISI meeting in Rio.

Let me start with congratulations: I am very happy to congratulate Pedro Silva on your behalf upon his election as the new ISI president. I have known Pedro for many years and he is definitely the right man at the right place. I would like to remind you that he was our president not long ago, so we should all be very proud that one of us is now presiding over the entire ISI. Good luck Pedro.

These are not the only congratulations: In my previous letter I told you of the special "Journal Papers Session" that we shall have at the World Statistics Congress (WSC) in Rio. Now I can tell you who the lucky winners are. The selected paper by *Survey Methodology* is: "Design-based analysis of factorial designs embedded in probability samples", authored by Jan Van Den Brakel. The selected paper by the *Journal of Survey Statistics and Methodology* is: "AIC and BIC for modelling with complex survey data", authored by Alastair Scott and Thomas Lumley. Congratulations to Jan, Alastair and Thomas. Very well deserved. Steve Fienberg agreed to discuss the two papers during the session. It is fantastic to have you all in the same session, which promises to be a real hit.

Now that we are all in the right mood, here is some great news that applies to each one of you. The *American Association of Public Opinion Research (AAPOR)* and the *Survey Research Methods Section of the American Statistical Association (ASA-SRMS)*, kindly agreed to offer a one-year **free on-line subscription** to the *Journal of Survey Statistics and Methodology (JSSAM)* to **all** IASS members. The free subscription will start in January 2015. I am very grateful to Michael Link, President of AAPOR and to Joe Sedransk and Roger Tourangeau, for all their input in securing this wonderful offer, of which I am sure you will take full advantage. Special thanks are also due to Oxford University Press. See **Appendix 1** for a letter from Professor Link, which also describes how to get access to the journal.

Returning to the WSC in Rio, I would like to remind you of the Cochran-Hansen prize competition for the best paper on survey research methods submitted by a young statistician from a developing or transition country. The deadline for application is reaching soon (**15 Feb, 2015**). Full details can be found on our website: <http://isi-iass.org/home/cochran-hansen-prize/>. This is a wonderful opportunity for one of our rising young stars to attend the meeting (all travel and accommodation expenses covered by the IASS), so please encourage appropriate candidates to apply. The Prize consists of books and journal subscriptions in the value of €500.

As in previous WSCs, The World Bank is funding participants from developing countries to attend the WSC-2015 and associated events. Funding will consist of return airfare, registration and visa fees, and accommodation. Successful candidates will also be able to apply for one of the short courses before the WSC, or one of the Satellite meetings right after. We received a short notice of this competition and I sent you an email about it, which I hope I reached you early

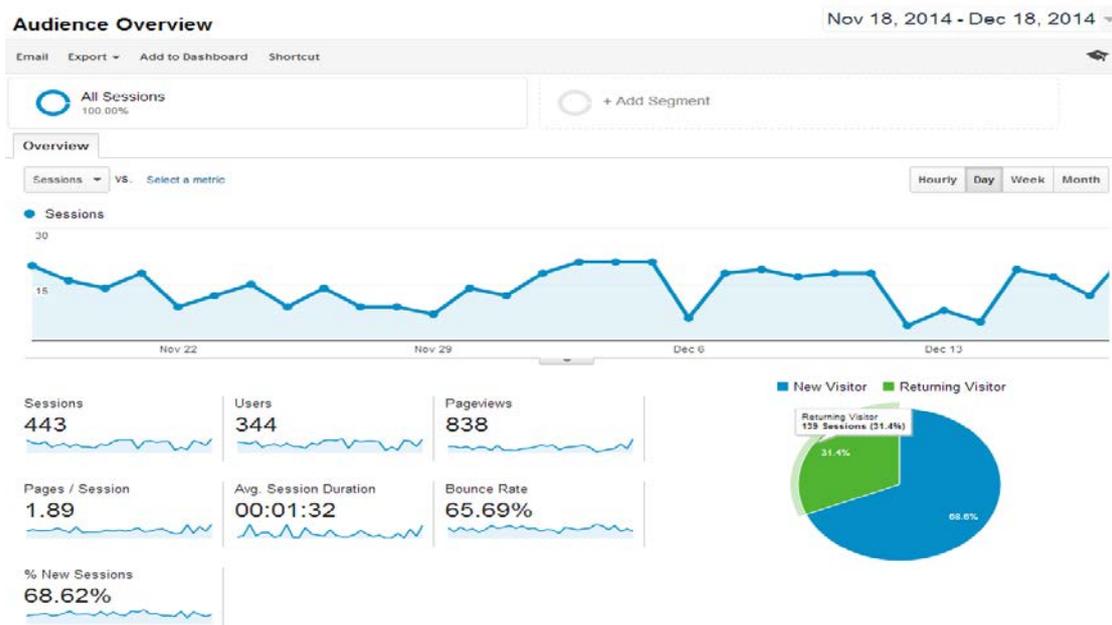
enough to allow you to encourage colleagues and students to apply. We are very grateful to the World Bank for this funding.

We are now in the final stages of preparations for election of new officers for our Association. We are about to elect a new president, two new vice presidents, a new scientific secretary and six new council members. Jean Opsomer kindly agreed to head the nomination committee and I presume that you will soon hear from him about the election process. I have no idea who will be nominated, but I am sure that the new elected officers will do a great job in moving our association forward.

Speaking of the future, we have been recently asked by the ISI to prepare a strategic plan for the IASS. The request was sent to us with a very short notice, not giving us sufficient time to prepare a decent, well thought plan. We decided therefore to move in two directions. With the consent and approval of our council members, I drafted with the help of Tom Kaplan a strategic plan and sent it to the ISI. More important, we established a working group composed of Geoff Lee (chair), Steve Heeringa, Daniela Cocchi and Natalie Shlomo, with the task of preparing a strategic plan for the future of our association. We plan to discuss the plan at the general assembly during the WSC in Rio. I hope that you will all attend that meeting because it is time for us to re-define the goals of the association and how to achieve them. In **Appendix 2** of this letter you will find the draft strategic plan that I sent to the ISI. Please read it thoroughly and if you have any comments/suggestions, write to Geoff, geoff.lee99@bigpond.com, so that the working group can consider them when preparing the new plan.

We continued sponsoring conferences and workshops that are appropriate for our members. The full list of conferences appears in the detailed report of Mick Couper, our scientific secretary, elsewhere in this issue.

We seem to do well with our new website. As the following picture shows, it has been visited by about 344 users in the last few months. Well done Olivier!



Last, but definitely not least, this is the time that we ask you to renew your membership. I won't bother you again with details on how bad we are size-wise, but

this is my last opportunity to encourage each one of you to make every effort to bring in new members. Talking of our new strategic plan, this undoubtedly should be our main strategic mission, **double** our membership!

By the time you receive this issue of *the Survey Statistician*, the big holidays will be over, but it is not too late to wish you and your families a wonderful new year. There are many events ahead of us and I look forward to meeting you all.

Danny Pfeffermann,
President, IASS

Appendix 1. Letter from Professor Michael Link, President of AAPOR

Dear Members of the International Association of Survey Statisticians:

On behalf of the American Association of Public Opinion Research (AAPOR) and the Survey Research Methods Section of the American Statistical Association (ASA-SRMS), I am pleased to offer you a complimentary, one-year, online only subscription to our *Journal of Survey Statistics and Methodology* (JSSAM). Members of AAPOR and ASA-SRMS enjoy receiving hard-copy and online access to JSSAM as part of their membership benefits.

TO SET UP YOUR FREE ONE YEAR ONLINE SUBSCRIPTION TO JSSAM

1. Go to: www.oxfordjournals.org/register
2. Create an account (or log in)
3. Click on 'manage my subscriptions'
4. Enter access code (**IASS2015**) into the 'subscriber number' box
5. Click 'add subscription'

**Trial period expires December 31, 2015*

The *Journal of Survey Statistics and Methodology* is sponsored by AAPOR and the American Statistical Association. Its objective is to publish cutting edge scholarly articles on statistical and methodological issues for sample surveys, censuses, administrative record systems, and other related data. It aims to be the flagship journal for research on survey statistics and methodology.

Topics of interest include survey sample design, statistical inference, nonresponse, measurement error, the effects of modes of data collection, paradata and responsive survey design, combining data from multiple sources, record linkage, disclosure limitation, and other issues in survey statistics and methodology. The journal will publish both theoretical and applied papers, provided the theory is motivated by an important applied problem and the applied papers report on research that contributes generalizable knowledge to the field. Review papers are also welcomed. Papers on a broad range of surveys are encouraged, including (but not limited to) surveys concerning business, economics, marketing research, social science, environment, epidemiology, biostatistics and official statistics.

We hope that you enjoy the journal. We encourage you to consider submitting your own work for publication. Best wishes for the holiday season and for good health and prosperity in the New Year.

Sincerely,

Michael W. Link
President, American Association for Public Opinion Research

DRAFT PROPOSAL

1. Introduction/ Vision

The International Association of Survey Statisticians (IASS) aims to promote the study and development of theory and practice of sample surveys and censuses. It also aims to increase the interest in surveys and censuses and raise awareness of their importance among statisticians, governments and the public in the different countries of the world.

The IASS represents survey statisticians from across the globe. The survey statistician profession has applications and implications for a wide range of fields, and the professional work which is carried out, and the collaboration by means of the association contributes very significantly to the society that we belong to, in almost all aspects of life. Statistics has changed quite dramatically in the last decade with more and more emphasis on analysing big data sets, employing advanced computing algorithms, what is referred to these days as "data science". There are major challenges facing the statistics world, and the IASS which represents professional survey statisticians from around the world is in a position to contribute ways of dealing with them. Examples of challenges facing survey statisticians include the proper use of big data and data obtained from web-panels, integration of computer science in the production of big data estimates, addressing problems of response burden and nonresponse, development of indices that measure quality of life, standardization and globalization of official statistics, mode effects and proxy surveys, adaptive survey designs, data accessibility, privacy and confidentiality, integration of statistical and geospatial information, and future censuses and small area estimation.

The vision of the IASS is that by using modern approaches to sample survey statistics, including technical developments that have occurred and will occur, the IASS professionals can collaborate in developing new methods that will address new challenges.

2. Objectives and Strategies

Objective A: Foster the growth, development, application and adaptation of survey statistics in all areas of activity that can benefit from it. Promote research into new applications of survey statistics and in the applications of modern technology in survey statistics.

Strategy 1. Encourage research in all areas and use the World Statistics Congress and other conferences and forums to allow for discussion and collaboration on these topics. The IASS program for the WSC in Rio contributes greatly to this effort. The program includes sessions on methodologies relating to big data applications, statistics disclosure control, sampling frame and nonsampling error in internet surveys, adaptive survey designs, new developments in use of model-based methods for the production of official statistics, census during times of changing

methodologies and technologies and the use of remote sensing for agricultural statistics.

Strategy 2. Ensure that the IASS newsletter 'The Survey Statistician' provides current articles on new and emerging methods in survey statistics, provides knowledge exchange, and a platform for providing guidance and advice from experts.

Objective B: Foster Statistical Capacity Building in developing countries.

Strategy Develop a multi-year program of teaching sample survey statistics and consulting in developing countries.

Objective C: Expand greatly the membership of the IASS so as to be inclusive of the vast majority of survey statisticians, and thereby providing a sense of community among survey statisticians. Aim at transforming the organization into a world centre for collaboration on survey statistics, possibly with continental branches.

Strategy 1. Carry out an extensive membership campaign based on "a member brings a member" approach.

Strategy 2. Advertise the IASS newsletter 'The Survey Statistician' more broadly and include a call for membership.

Strategy 3. Carry out a concerted effort to bring in new institutional members.

Objective D: Provide a forum for the exchange of ideas and collaboration between theoreticians and practitioners.

Strategy 1. Expand the membership to include more academic statisticians as well as more applied statisticians in national statistics institutes and other government ministries and the private sector.

Strategy 2. Encourage the development of collaborative projects by teams composed of both academic and applied statisticians. These projects could be the basis for future sessions at the WSC and other forums.

Strategy 3. Encourage statistics students to work with applied survey statisticians. The IASS could develop a program of internship to that end.

Objective E: Promote survey statistics education; boost teaching of sample statistics at universities and colleges around the world.

Strategy 1. Modular courses in survey statistics for adaptation by universities and colleges.

Strategy 2. Encourage senior IASS statisticians to take on assignments which would involve teaching these courses and mentoring local statisticians who could then further develop and teach these courses.

Strategy 3. Develop online courses in survey statistics, encourage universities and colleges to adapt them and recruit senior IASS statisticians that will be responsible for these online courses.

Objective F: Encourage the entrance of young people (the future of our discipline) into the fields of survey statistics and thereby increase their participation in the IASS activities.

Strategy 1. Foster a wide competition for the Cochran - Hansen and possibly other, new established prize competitions.

Strategy 2. Encourage current academic members of the IASS to bring their students into the IASS and perhaps use some of their resources to bring students to the WSC and other conferences.

Strategy 3. Hold regional conferences, workshops and seminars aimed at university students on the subject of survey statistics.

Strategy 4. Provide a forum for joint survey statistics research projects by partnerships of IASS members and students.

Objective G: Exploit the use of the internet for the benefit of the IASS and its members.

Strategy. The IASS website was recently revised considerably. Translate the content of the website into French and other languages. Encourage members to take an active role in the development of the website by posting questions, proposing new ideas and participate in open discussions. Advertise job openings.



Report from the Scientific Secretary

A key activity of the Association is to support regional conferences for survey statisticians. In the past year, IASS has provided – or has agreed to provide – financial support to the following conferences and workshops:

- III International Workshop on Surveys for Policy Evaluation, at the Federal University of Brasília (Brazil), in November 2013.
- 8th French Colloquium in Survey Sampling, held at the University of Burgundy/Université de Bourgogne, Dijon, in November 2014.
- The 2014 Baltic-Nordic-Ukrainian Network on Survey Statistics Workshop, held in Tallinn, Estonia on August 25-28, 2014. IASS has also agreed to support the 2015 Baltic-Nordic Conference on Survey Statistics.
- ITACOSM 2105, the biannual meeting of the Italian section of the Italian Society of Statistics, to be held in Rome, June 24-26, 2015.
- Satellite meeting on small area estimation to be hosted by the Catholic University (UC) of Chile, from August 2-4, 2015, following the 2015 ISI World Congress in Rio de Janeiro
- The European Establishment Statistics workshop to be held in Poznan, Poland, in September 2015.

All meetings supported by the IASS are encouraged to provide reduced fees for IASS members, and to promote IASS membership at their meetings. As many survey statisticians – especially those from developing countries or in the early stages of their careers – are unable to attend the biennial World Statistics Congresses, these regional meetings are an important way to provide value for IASS members and to encourage new members.

With generous support from the ISI and World Bank, IASS was able to provide travel funds for four early-career survey statisticians to attend the Statistics Canada Symposium in Gatineau, Quebec, in October 2014. Statistics Canada also generously waived registration and workshop fees for the four scholars. The four awardees are:

- Omotola Dawodu, an Assistant Lecturer in the Department of Physical and Computer Sciences (Statistics Unit) at McPherson University in Nigeria. She is currently pursuing a Ph.D. in Statistics (Survey Sampling).
- Diego Andres Perez Ruiz, from Mexico, currently a Ph.D. student in Statistics at the University of Manchester, School of Mathematics. He has an M.Sc. in Probability and Statistics from the Center for Mathematical Research (CIMAT) in Guanajuato, Mexico.
- Andres Gutierrez, is the Dean of The Faculty of Statistics, Universidad Santo Tomas, Colombia. He completed his Ph.D. in Statistics in 2014.

- Sâmela Batista Arantes, is a researcher at IBGE in Brazil. She obtained an M.Sc. in Population Studies and Social Research at the Escola Nacional de Ciências Estatísticas (ENCE) in 2012.

The awardees seemed to be networking well at the conference, and benefitting from their attendance. Upon her return to Nigeria, Omotola Dawodu sent this note: “I write to express my sincere appreciation to you all for giving me the golden opportunity to attend the 2014 International Methodology Symposium organized by Statistics Canada. I thank the International Association of Survey Statisticians (IASS) for selecting me for the award. I also thank the ISI and World Bank for giving me the travel grant. Finally, I thank Statistics Canada for helping to waive the registration fees for the workshop and symposium. I really learnt a lot in the workshop and symposium.”

As part of the effort to promote the activities of the IASS, the brochure that was originally developed by the Australian Bureau of Statistics (ABS) for the World Statistics Congress in Hong Kong was updated (with ABS’ permission) and is now available on the IASS website (see <http://isi-iass.org/home/about/join-iass/>). All members of IASS are encouraged to print copies of the brochure and distribute at local and regional meetings to promote membership in the IASS. All meetings that receive IASS support (see above) are asked to make copies of the brochure available at their meetings.

An important role of the IASS scientific secretary in the past has been the organization of short courses at the WSC. Since Hong Kong, this function has been taken over by the ISI. We have representation on the ISI short course committee, and IASS submitted a strong slate of short courses. Final decisions about the short courses to be offered at the WSC in Rio de Janeiro are still being made, and members will be kept informed through the IASS (<http://isi-iass.org/home/>) and WSC (<http://www.isi2015.org/>) websites.

I look forward to meeting many of you at the 60th WSC in Rio de Janeiro in July. If you have any suggestions on how to increase the membership (both individual and organizational) of the IASS and to provide enhanced services for members, please contact me or the IASS President.

Mick P. Couper
mcouper@umich.edu

News and Announcements

2014 Waksberg Award to Constance F. Citro

The journal [Survey Methodology](#) has established an annual invited paper series in honour of Joseph Waksberg, who has made many important contributions to survey methodology. Each year a prominent survey researcher is chosen to author an article as part of the Waksberg Invited Paper Series. The paper reviews the development and current state of a significant topic within the field of survey methodology, and reflects the mixture of theory and practice that characterized Waksberg's work.

The 2014 recipient of the Waksberg award is Dr. Constance F. Citro. Dr. Citro is the Director the Committee on National Statistics (CNSTAT) of the U.S. National Research Council/National Academy of Sciences.



Picture: Paddison Wong, Statistics Canada

Constance F. Citro receiving the award from Graham Kalton from Westat

In her paper titled "[From multiple modes for surveys to multiple data sources for estimates](#)", Dr. Citro argues that we can and must move from a paradigm of producing the best estimates possible from a survey to that of producing the best

possible estimates to meet user needs from multiple data sources. To do so, she first reviews the rise and benefits of probability sampling for official statistics in the United States as well as the various growing threats to the quality of those survey-based estimates. She then considers the strengths and weaknesses of administrative records and other non-probability-survey data sources for official statistics. She presents illustrative examples and opportunities in the United States to transform ongoing household survey programs. She concludes by presenting barriers to moving to a multiple data sources paradigm and by suggesting ways to lower those barriers.

Dr. Citro presented her paper in an inspiring talk during the 2014 International Methodology Symposium, held from October 29 to 31, 2014 in Gatineau, Canada, under the theme “Beyond traditional survey taking: adapting to a changing world”.

My experience attending the 2014 International Methodology Symposium

**Beyond traditional survey taking: adapting to a changing world.
October 29 to 31, 2014
Palais des congrès de Gatineau,
Gatineau, Quebec, Canada.**

By Diego Andres Perez Ruiz
PhD student, The University of Manchester, United Kingdom
diego.perezruiz@manchester.ac.uk

To attend the Symposium, I was awarded with a travel scholarship from the ISI, IASS and the World Bank. These organizations recognize that the opportunity to attend international meetings is especially important in the early career development stages of young statisticians. I first read about the travel award on the International Statistical Institute web page (in the award section). I applied and submitted the required documents including a CV, letter of recommendation, cover letter and personal statement. Some weeks after, I received an email notifying me that my application for the travel award was successful and I would attend the International Methodology Symposium. In addition to myself, three other young statisticians from developing countries were awarded this travel grant, namely, Omotola Dawodu from Nigeria, Andres Gutierrez from Colombia, and Samela Batista from Brazil.

Once I had made the necessary arrangements to attend the Symposium, I booked my tickets to Ottawa and awaited anxiously for the date.

When I arrived in Ottawa, it was windy. I could feel the winter starting in Canada. I took the bus and I went straight to the conference hotel to rest from the long trip. On my way to the hotel, I was feeling excited to be in Canada and to attend the Symposium. I was sure that the next three days were going to be inspirational and informative.

At 8:45 on Wednesday 29th October, Rosemary Bender, Assistant Chief Statistician, gave the welcome talk to the 2014 International Methodology Symposium. The

session was followed by Ray Cambers from University of Wollongong, the keynote speaker. He spoke about survey sampling in official statistics which involved some thoughts on future directions. In particular, he focused on the main survey-related methodological issues that National Statistical Institutes (NSI) will face in coming years, and how methodologists will need to change the way that they view “official statistical inference” in order to deal with them in an effective way.

The three days had a packed schedule. The sessions were divided in two main rooms, Chapleau and Gatineau Rooms. In these rooms, we discussed topics including Data Collection, Web Surveys, Record Linkage, Non-sampling Errors, Big Data and Microsimulations. Along with the talks, there was a poster session where the topics concerned interviewer's influence on bias, secure record linkage and automatic coding of occupations.

When the sessions finished, we had the conference dinner. I met people from Statistics Canada and they explained to me their job in Canada's national statistical agency, and the importance of official statistics. After the dinner finished, I returned to my room to rest from a very long day.

All the invited speakers were phenomenal. I met people working in Data Collection and Mode Effects, in Record Linkage, Web-Surveys and Microsimulations, as well.

On Thursday, the first talk was the Waksberg Award talk. The award winner was Constance F. Citro, from the Committee on National Statistics of the U.S. National Research Council/National Academy of Sciences. Her talk was titled “Multiple modes of surveys to multiple data sources for estimates”. In her talk, she argued that we can, and must, move from a paradigm of producing the best estimates possible from a survey to that of producing the best possible estimators to meet user needs from multiple data sources. Such sources include administrative records, transactions and internet based data. Her talk concluded by suggesting different ways to inculcate a culture of official statistics that target the end result being relevant, timely, accurate and cost effective statistics, and to treat surveys as a means to an end.

One session I was fascinated with was the session about Big Data, where Mick Couper spoke. He started his talk with a quote;

“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”

His point of view was that Big Data is unlikely to replace high quality surveys, but the two methods can serve complementary functions. He suggested how big data can be used for reliable population-based inference.

During the three days, I listened to many talks, some of which included: Jelke Bethlehem, Statistics Netherlands; Mike Brick, Westat; Piet Daas, Statistics Netherlands; Abel DaSylva, Statistics Canada; Gaël De Peretti, Institut national de la statistique et des études économiques; Jill Dever, Research Triangle Institute; John Eltinge, U.S. Bureau of Labour Statistics; Piero Demetrio Falorsi, Italian National Statistical Institute; Wayne Fuller, Iowa State University; Lilli Japac, Statistics Sweden; Christian Nadeau, Statistics Canada; Donald B. Rubin, Harvard University; Joe Sakshaug, University of Michigan, among others.

Most importantly, I was able to experience poster sessions and talks as well as have conversations with colleagues. I feel lucky to have attended the Methodology Symposium as a PhD student. I now know what to expect from future conferences and meetings, and have begun to create a professional network. For those interested, the proceeding of the Symposium and the slides of the talks will be available at the web page of the Symposium (<http://www.statcan.gc.ca/eng/conferences/symposium2014/index>) next August, and will be available in English and French.

I also had time to explore Ottawa, a city full of museums divided by the Ottawa River and connected by the Alexandra bridge amongst others. I feel Ottawa is a multicultural city with diverse population. I had time to visit the Parliament, the political and cultural heart of the city using a visit guide. I also visited the Peace Tower where you can experience an incredible view of the city.



Photo taken from Ottawa's Parliament Hill.

I would like to extend my appreciation to the ISI, IASS and the World Bank for giving me this honour, and for making this trip possible. Thanks to the logistic committee, scientific committee and the registrar. Special thanks to Mick Couper, from Michigan University, who looked after me so kindly, Ray Chambers and Pedro Luis do Nascimento, president-elect of the ISI, with whom I enjoyed delicious meals.

The experience of attending the International Methodology Symposium will become an invaluable treasured memory, and will be very important in my career development. The knowledge, connections and perspectives I found at the Symposium will continue to inform my thoughts and work and I look forward to future International Methodology Symposiums.

PARIS21, 15 years improving lives through better statistics

Guadalupe De Las Casas, PARIS21 Communications Officer

The Partnership in Statistics for Development in the 21st Century ([PARIS21](#))¹ has supported the strengthening of statistical capacity at country and regional level since its creation in 1999. In the lead up to the UN Sustainable Development Goals, PARIS21's extensive experience in advocacy work remains indispensable to help developing countries to promote the better use and production of statistics.

PARIS21's goal is to foster more effective dialogue among those who produce development statistics and those who use them, through facilitating international events, supporting country-based activities and regional workshops and producing promotional materials and publications such as the Partner Report on Support to Statistics ([PRESS](#)) which annually presents data on technical and financial support to statistical development worldwide.

One of PARIS21's main activities has been the support to the development of National Strategies for the Development of Statistics ([NSDSs](#)) which present a set of milestones to strengthen National Statistical Systems. PARIS21 contributes to the NSDSs, providing [Guidelines](#), supporting mobilisation and leveraging of national and international resources for the implementation of the NSDSs. Additionally, NSDS evaluations and National Statistical Systems [peer reviews](#) are regularly carried out to improve the strategic planning and National Statistical System governance of countries.

Alongside its core activities, PARIS21 is implementing the [Informing a data revolution](#) project. This project focuses on how to improve the understanding of data systems in developing countries, including how they should be better designed, managed and supported to generate the data needed to reduce poverty. As part of this project PARIS21 compiled an [Innovations Inventory](#) of cutting edge uses of new technologies and a [Metabase](#), a data directory that displays statistical capacity indicators across developing and developed countries. A [Road Map](#), the foundational document of the project, will set out the goals, activities and resources needed for developing countries to use data to achieve national priorities and the sustainable development goals.

As the post-2015 agenda will shape development thinking over the coming years, PARIS21 will continue to bring statistics to the core of the development debate to improve the provision and availability of data to users and increase transparency and improve evidence-based decision making in developing countries.

¹ PARIS21 is a global partnership of national, regional and international statistics experts and decision makers guided by a board of international stakeholders with representatives from developing countries, bilateral donors, multilateral institutions and key partners.



Ask the Experts

How Should I Run a Regression Based on a Complex Survey that Has Missing Item Values?

Phillip S. Kott, RTI

1. INTRODUCTION

There have been many books and articles discussing various methods for handling regression modeling with missing item data. Sadly, most have ignored the impact of the complex designs often used in survey samples, sometimes by concentrating wholly on clinical trials. A report by the National Research Council (2010) on this topic is an excellent resource and contains many useful references.

That is not to say that the subject of conducting linear (or logistic) regressions with complex survey data having missing item values has been completely ignored. A new text by Kim and Shao (2013), although about the handling of missing item values in general, provides a penetrating chapter on applications to survey samples. A recent National Institute of Statistical Sciences (NISS) workshop focused on complex survey data analysis in the presence of missing item values². My own contribution to the workshop was on listwise deletion, so I will begin my discussion there.

2. LISTWISE DELETION

Listwise deletion, which is also known as “complete-case analysis,” is a popular, and unjustly vilified, method for handling missing item values in a regression analysis. It simply removes any record with a missing variable of interest from the analysis. This method produces estimates that are often not as accurate (i.e., have larger root mean squared errors) as those produced by more sophisticated competitors like multiple imputation and maximum likelihood. Moreover, using list deletion can produce biased estimates in situations where competitor approaches do not. This can happen when item nonresponse is a function of the dependent variable of the model being fit, that is, y in the linear model $E(y|\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$ or the logistic model $E(y|\mathbf{x}) = 1/[\exp(1 + \mathbf{x}^T\boldsymbol{\beta})]$.

Wilkinson and the Task Force on Statistical Inference (Board of Scientific Affairs, American Psychological Association 1999, p. 598) went so far as calling listwise deletion “among the worst methods available for practical applications.” Despite this widespread belief, *regression estimates using listwise deletion can sometimes be*

² Presentation slides are at: <https://www.niss.org/events/niss-workshop-analyzing-complex-survey-data-missing-item-values>.

nearly (i.e., asymptotically) unbiased when estimates produced by its more sophisticated competitors are not. That can happen when item nonresponse in a covariate of the model – that is, a component of \mathbf{x} in $E(y|\mathbf{x}) = f(\mathbf{x}^T\boldsymbol{\beta})$ (where $f(t) = t$ or $1/[\exp(1 + t)]$) – is a function of the missing value itself. Furthermore, many competing methods cannot as easily incorporate the impact of the sampling design.

The biggest problem with listwise deletion is that it discards incomplete records that may contain useful information for the analyst. When analyzing cross-sectional data especially, this problem does not always loom very large.

It is important to realize that we are assuming here, and we will for the remainder of this article, that the standard model being estimated is correct. By that I mean: that $E(y|\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$ or $E(y|\mathbf{x}) = 1/[\exp(1 + \mathbf{x}^T\boldsymbol{\beta})]$. This assumption doesn't imply that the *sampling weights* are ignorable. They may not be; in particular, a sampling weight may be correlated with the model error: $\varepsilon = y - E(y|\mathbf{x})$.

The oft-repeated assertion that listwise deletion is nearly unbiased only if item nonresponse is missing completely at random is correct when the standard model is allowed to fail. In that situation, the fitted regression does not correspond to what most statisticians would regard as a model at all. In its place is what I have called an "extended model," one that assumes only that the model errors are uncorrelated with the model's covariates, that is, $E(\mathbf{x}\varepsilon) = \mathbf{0}$, which is a much weaker assumption than the standard model's $E(\varepsilon|\mathbf{x}) = 0$. I discuss the impact of sampling weights on the standard and extended regression model in Kott (2007). It is not hard to show that under the standard model, one can scale the weights by any function of the covariate vector \mathbf{x} and retain the near unbiasedness of weighted regression. This explains why listwise deletion remains nearly unbiased when the probability a record is deleted is a function of the covariate vector.

Even when nonresponse *is* a function of the dependent variable of a regression model, it may be possible to adjust the sampling weights (which may already be adjusted to account for undercoverage and/or *unit* nonresponse) in a listwise-deleted data set to remove the bias from the estimated regression coefficients. This can be done by first estimating the probability that a given record will be deleted in the listwise-deletion process as a function (wholly or in part) of the value of the dependent variable. The inverse of that probability is then included as an additional factor in the weight for each record in the listwise-deleted data set. A replication method could be used to measure correctly the resulting standard errors of the regression coefficients.

3. IMPUTATION

Imputation (i.e., replacing a missing value with a particular value) is a very common way to handle missing items values in a survey. Although not always explicitly stated, imputation assumes some model for variable behavior in addition to the model being estimated in the regression analysis itself. The dependent variable in model being analyzed can – and often will be – a covariate in a probabilistic imputation model (i.e., the expected value under the imputation model of a component of \mathbf{x} will be partly a function of y) as can *auxiliary variables* not in the model being estimated

When the imputation model is probabilistic and correctly specified, imputing a missing item value by its expectation under the model will remove the bias from a resulting estimated mean or total of a particular variable. The same, however, cannot be said about an estimated regression model when covariate values are imputed. Adding appropriate random noise to the imputation will fix that problem while not biasing estimated univariate means and totals.

Unfortunately, the standard errors of estimated regression coefficients cannot be easily estimated when each missing value is imputed with its expected value plus a random error. Treating imputed values as real in standard-error estimation almost always underestimates standard errors even assuming the model used in the imputation is correct.

4. MULTIPLE IMPUTATION

Multiple imputation (MI) is one way to get around the standard-error problem of single imputation, while also increasing the accuracy of the estimated regression coefficients themselves. MI does this by estimating each missing value several times (usually 5 times). Taking the average of the multiply-imputed values removes much of the added variance due to the random noise added to each single imputation. In addition, computing the variability of the estimate across the (5) sets of imputed values provides the means for measuring the increase to standard error due to fitting models for the missing variables in the first place. See, for example, Schafer (1999).

Many (e.g., Kott 1998 and Kim *et al.* 2012) have shown the limitations of MI with complex survey data. Appropriately incorporating weighting, strata, and clustering into the imputation models is not trivial. Even when done reasonably, the MI variance formula may still not provide asymptotically unbiased variance estimates. Nevertheless, computing MI estimates of standard error is often more reasonable than ignoring the impact of imputation on standard errors entirely.

5. MAXIMUM-LIKELIHOOD

An alternative to MI and to imputation in general is the direct use of maximum-likelihood (ML) estimation. This approach has become increasingly popular in statistical analyses not involving complex survey data. When fitting a particular model, ML requires that all the variables in a model missing values be treated as random variables in a larger and correctly specified model. Its greatest strength is that it can, in principle, use all available information efficiently; that is, no other way results in estimates with smaller standard errors.

Both MI and so-called full-information ML (or FIML) usually require specification not only of the models for the expected values of the dependent variable and for covariates with missing values but also of their variance/covariance structure and often their probability distributions. As a consequence, the resulting estimates and their standard errors are not as robust to model misspecification as is commonly the case when analyzing survey data, where the errors in a linear regression model are rarely assumed to be normal and their variances need not be specified up to a constant (see Skinner 1989). That robustness can often be regained by treating the score function (the derivative of the log likelihood with respect to the model

parameters, which equals zero when the likelihood is maximized) as an estimating equation and computing robust standard errors for its solution, most easily by replication. Incorporating the weights is trivial in this context because there is no longer a claim that the solution is ML, while clustering and stratification are treated as nuisances in robust variance estimation. The method is more properly called “pseudo-maximum likelihood” (Gourieroux et al. 1984).

Using ML methods requires the analyst to specify a model structure (which can occur indirectly depending on the software employed). The same can usually be said about MI except that the specification is most often the responsibility of the statisticians releasing the survey results, which need not include the survey analyst. Indeed, the same multiple imputation can be used regardless of the analysis, while ML is analysis specific, as is listwise deletion.

Kim and Shao (2013) discuss at length an alternative to MI based on ML principles called “fractional imputation,” in which a missing value is imputed many times and its weight distributed among those imputed values. Standard errors can be estimated using replication methods. Slides discussing fractional imputation can be found using the NISS-workshop link. Unfortunately, there is as of yet no publicly-available software to implement this methodology.

6. ADDING A DUMMY (OR CATEGORY) TO DENOTE A MISSING COVARIATE

A popular method for handling item missingness in a regression analysis that should never be used is adding a dummy variable to indicate when a continuous variable has a missing value or adding an additional category (“missing”) when a categorical variable has a missing value (the two are mathematically equivalent). Jones (1996) demonstrated that this practice will lead to biased estimates of regression coefficients even if the missingness is unrelated to the dependent variable when a component of \mathbf{x} with missing values is correlated with another component of \mathbf{x} . This is almost always the case. The size of the estimated coefficient of the covariate with missing values will tend to be underestimated while the estimated coefficients of correlated covariates incorrectly try to “compensate” for the missing values.

7. SOME CONCLUDING REMARKS

One problem often faced when there are missing item values in a regression analysis is that the pattern of missingness can be very complicated, rendering difficult the development and fitting of a single multivariate model treating the dependent variable and all covariates with missing variables. Raghunathan *et al.* (2001) introduced a chaining approach for MI that greatly simplified the problem. The idea of chaining is to develop imputation models one variable at a time using the previously imputed values for the other covariates. Once the process is completed, it begins again, dropping the imputed values only for the variable whose imputation model is being fit. The use of recursive partitioning (classification and regression trees) can remove the need for developing a parametric imputation model for each variable with missing values. Toth and Eltinge (2011) provide theoretical justification for recursive partitioning with complex survey data.

Despite theoretical questions about whether chaining converges or even whether the links in the chain are logically consistent, the method appears to work well in

practice. See van Buuren et al. (2006). Alternatively, Jerry Reiter discussed a Bayesian mixture-model approach at the NISS workshop mentioned in the Introduction.

Complex nonresponse patterns are not a problem with listwise deletion *per se* except that the amount of data that is lost can be substantial. Even if no single variable is ever missing more than 2% of the time, say, the fraction of complete records in a regression analysis with 10 covariates can be as low as 80%.

It used to be widely believed that the proportion of missing data is directly related to the quality of statistical inferences. We now know that the amount of missing data should not be the sole criterion for assessing its impact on statistical analyses. As demonstrated convincingly by Groves and Peytcheva (2008), the missing data mechanisms and the missing data patterns have greater impact on bias than does the proportion of missing data.

Prudent advice when conducting a regression analysis with missing item values is to assess for oneself with one's own understanding of the data whether missingness is a cause for concern and whether, for example, treating imputed values as real ones undercuts the validity of the inference. Unfortunately, one must try a method like MI or ML to assess the impact on inference of *not* doing MI or ML.

The literature stresses the need to do sensitivity analyses on the models underpinning MI and ML methods especially with regards to the usual assumption that nonresponse is not a function of the missing item values. The same could be said of the response model used in weighting methods.

References

Gourieroux C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681-700.

Groves, R.M. and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly* 72, 167-189.

Kim, J. K., Brick, J.M. Fuller, W.A., and Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society B* 68, 509-521.

Kim, J. K. Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*. London: Chapman & Hall/CRC.

Kott, P. (2007). Clarifying some issues in the regression analysis of survey data. *Survey Research Methods*, 1, 11–18

Kott, P. S. (1995). A paradox of multiple imputation. *ASA Proceedings of the Survey Research Methods Section*, 380-383.

National Research Council (2012). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: National Academies Press.

Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 27, 85-95.

Schafer, J.L. (1999). Multiple imputation: A primer. (1999). *Statistical Methods in Medical Research* 8, 3-15.

Skinner, C. J. (1989). Domain means, regression and multivariate analysis. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester: Wiley.

Toth, D. and Eltinge, J. (2011). Building consistent regression trees from complex sample data, *Journal of the American Statistical Association*, 106, 1626-1636

van Buuren, S., Brand J. P. L., Groothuis-Oudshoorn K., and Rubin D.B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation* 76, 1049–1064.

Wilkinson, L. and the Task Force on Statistical Inference, Board of Scientific Affairs, American Psychological Association (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* 8, 594–604.

Ask the Experts - Call for Questions

If you'd like to ask the experts a question, please contact Kennon Copeland at copeland-kennon@norc.org.



New and Emerging Methods

Paradata – Big Data of Survey Research

Frauke Kreuter, University of Maryland, Ludwig-Maximilians-Universität

Our current digital age is not short on buzz words that describe the changes taking place in society, industrial production processes, and operations of all kinds. The German government is using the term 'Industry 4.0' to describe the fourth industrial revolution, characterized by the computerization of the manufacturing industry. Smart factories use data from production processes to adapt operations and use resources more efficiently. In the US, this trend is described as smart manufacturing, but other buzz words like "the Internet of Things" are also used to refer to the connectivity of devices, systems, and services. Whether it is in industry, advertising, education, or health, there is an increasing availability of streaming data that come in different shapes and forms. Data with these characteristics are often referred to as Big Data; see the discussion by Piet Daas and Marco Puts in the January issue of *The Survey Statistician* (Daas and Puts 2014). Despite the fact that these information streams can be error prone and spotty, great hope is placed on insights that can be gained from these sources.

Many of us are aware of the Big Data discussion, but what does this have to do with surveys? Well, we produce similar data, and it might therefore be worthwhile to pay attention to issues being raised in this emerging field (Lane et al. 2014), and technology that is being developed around it. Back in 1998, Mick Couper pointed out that the computerization of data collection creates digital by-products, paradata, that allow us to adapt our survey operations, to make processes more efficient and to weed out errors that might occur in the process. Couper's examples focused on keystroke data collected in Computer Assisted Interviewing (CAI) operations (Couper 1998), but the term was quickly adopted to capture many more by-products generated throughout the data collection process. Since then, the range of types of paradata, and the interest in paradata, has grown (see Kreuter, 2013).

Paradata as By-Product of the Survey Production Process

If we look closely along the survey statistics production process, depicted by Groves et al. (2004), we see that each step is producing data that can be captured (and digitized). The left hand side in Figure 1 represents the measurement process. Here we find paradata generated as part of pretesting efforts: for example, focus group discussions or field tests can be digitally recorded and data can be derived from

vocal features, speech patterns, etc. Similar recordings can be done during the actual interview process, and keystrokes can of course be captured in computer aided interviewing. Face-to-face interviews allow the capture of additional data through interviewer observations. Logs of actions taken in the editing process are one form of paradata that appears there. In web surveys, a series of additional paradata are available, such as user-agent-strings capture which inform about the device used for accessing a survey. There is not enough space here to detail all types of web paradata, instead the overview chapter by Mario Callegaro (2013) is a highly recommended read.

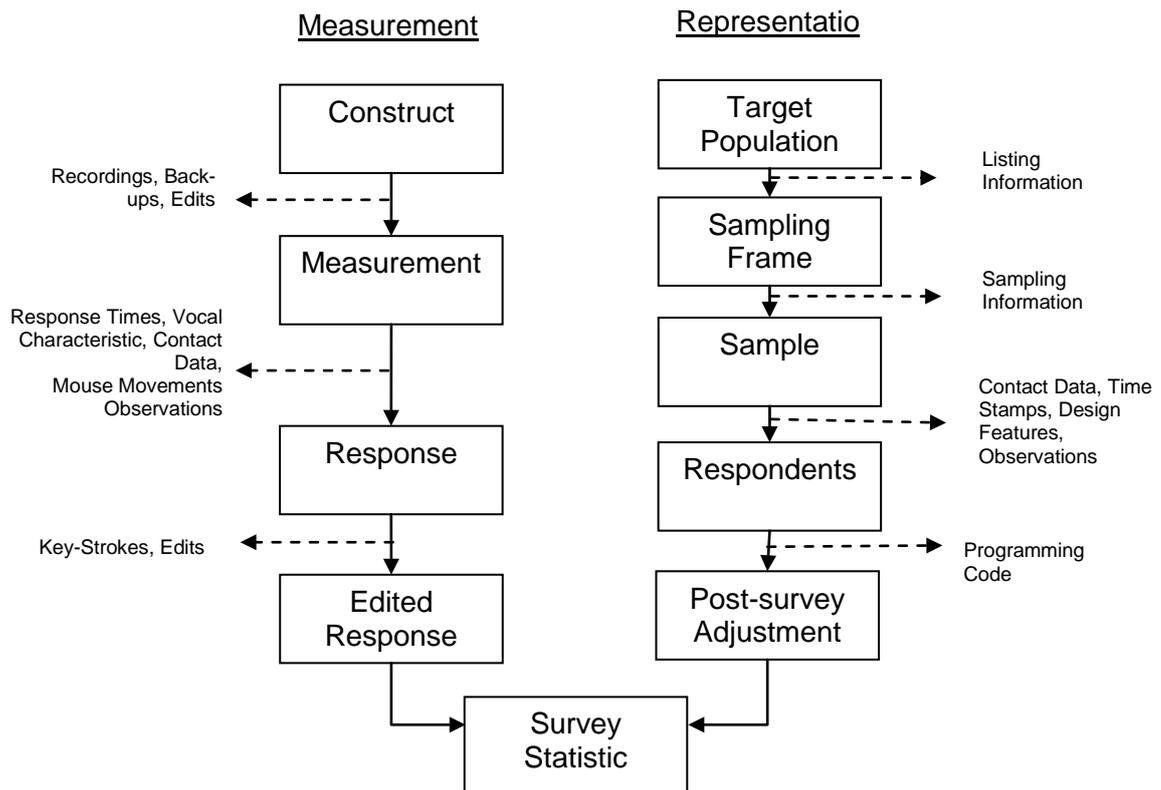


Figure 1: Paradata alongside the data collection process (see Groves et al. 2004, Kreuter & Casas-Cordero 2010)

On the representation side (on the right of the figure), a variety of paradata can be collected during the generation of sampling frames, ranging from data about the telephone number block to interviewer observations during the listing process (Eckman 2013). Very commonly collected are paradata generated during contact attempts to sampled households. In fact many of the large-scale international surveys are collecting and in part publishing such data together with the survey data themselves. For example, the Survey of Health, Ageing and Retirement in Europe (SHARE), or the European Social Survey (ESS) make detailed contact data available. In the case of the ESS, the call record data include call-level characteristics, such as the mode of the attempt (telephone, in-person, etc.). Other surveys might include information on whether an answering machine message was left (in telephone surveys), whether the interviewer left a 'sorry I missed you' card (in face to face surveys), or whether the case was offered incentives, among other

information. The U.S. National Survey of Family Growth (NSFG) made headlines in the survey community with its topical interviewer assessments on the presence of children in the household and the likelihood of sexual activity by the sampled person (Wagner et al. 2012). The beauty of many of these paradata is that they are (usually) available for both respondents and nonrespondents, thus allowing to assess or to correct the representation side.

Uses of Paradata

Just like with any Big Data stream, paradata are messy and require pre-processing before they can be analyzed. Groves (2012) pointed out that the era of Big Data is not called the era of Big Information for a reason. But with statistical models and with a specific purpose in mind, we have already seen successful uses of paradata to improve survey production or survey statistics.

Paradata are used to *detect problems with data collection* prior to implementing the actual instrument or *through monitoring the field work* and issuing interventions. Just to highlight a few examples: Reviewing several experimental studies that used response times, Olson & Parkhurst (2013) reported longer response times to be indicative of complex visual layouts, inconsistent response options, and lack of respondent engagement, and Horwitz (2014) found respondents' difficulty with specific survey questions to be correlated with certain mouse movements. Response times, break-offs and mouse clicks can be collected and assessed not only prior to the fielding of a survey, but they can also be monitored during the fieldwork process. Conrad et al. (2012) monitored response times and prompted survey respondents to give the question sufficient thought when they went too fast through the questions. Similarly, on the fieldwork side, contact history instruments (CHI), like the one implemented by the U.S. Census Bureau, are regularly used to monitor interviewer efforts when recruiting respondents and to detect falsification. The need to use the CHI was re-emphasized a few months ago in a hearing of the Committee on Oversight and Government Reform Subcommittee on Federal Workforce, U.S Postal Service, and the Census, of the U.S. House of Representatives (OIG 2014).

Paradata are also used to *improve data collection efficiency*. For years, CATI surveys have made extensive use of paradata to improve call scheduling. Such systems ensure that cases are called across different day parts, and parts of the week, and outcome codes on the calls are used to determine the interval length between calls. Such call scheduling algorithms are in place in most CATI studios. Enforcing similar rules or guidelines in face-to-face data collection is currently much harder, but the use of modern devices in fieldwork production may change this: in Poland interviewers are already using tablets with mapping and communication functions (Kurkowski 2013). Continuous monitoring of field work data collection with the help of paradata also allows the implementation of responsive (Groves and Heeringa 2006) or adaptive design procedures (Couper and Wagner 2011) with the goal of implementing the best possible design while maximizing data collection yield and minimizing costs. The NSFG was the first survey in which these ideas were tested. At a 2013 workshop on adaptive design organized by Statistics Netherlands, attempts using responsive or adaptive designs were presented for several European surveys, the Dutch and Norwegian Labour Force Surveys among them. The U.S.

Census Bureau is using variations of adaptive designs in many of their current surveys, among others the National Survey of College Graduates, the National Health Interview Survey, subsamples of the Economic Census, and the American Community Survey (Miller 2014). Statistics Canada has now switched (almost) all of its CATI data collection to responsive designs in which call record paradata play a key role. To support these efforts, several data collection organizations around the world are in the process of changing or building case management systems that allow a more flexible and timely use of incoming data and paradata (see the special issue of the Journal of Official Statistics on this topic in 2013).

Finally paradata are used to *adjust survey estimates*, either through the use of contact history information as it is done for example in the French Fertility, Contraception, sexual Dysfunction Survey (Charrance et al. 2014) or in using interviewer observations of housing unit characteristics as it was done in the PASS study (West et al. 2014). However, which paradata contribute to successful nonresponse adjustment will likely differ across surveys because proxy measures of key survey variables will thus naturally vary across surveys and survey topics (Kreuter et al. 2010). For the U.S. Health and Retirement Survey, Wagner et al. (2014) just recently found level of effort contact data not to be suited for nonresponse adjustment.

Big Data Characteristics in Paradata

Four characteristics, volume, velocity, variety, and lack of veracity, are commonly used to describe Big Data. Paradata share these characteristics, and survey statisticians are equally challenged by them.

- The *volume* of paradata can increase rapidly. For example if response times are analyzed across all variables and for all cases in a survey in a single analysis (e.g. Couper and Kreuter 2013) computation with complex models are challenged and require processing techniques designed for this purpose.
- The *velocity* describes the constant streams of data and calls for algorithms that allow learning and updating on a continuous basis, and of course the computing infrastructure to do so. In particular for adaptive designs models that allow updates with incoming data are key.
- The *variety* of the data (text, voice, observations) requires not only statistical methods that more easily allow for the combination of different data types collected at different levels, but often also considerable pre-processing of the data (e.g. Conrad et al. 2013).
- Finally, recent research has shown that the *veracity* of paradata is often lacking. Not all paradata are of equal quality, and some are even missing. For example, interviewers can misjudge features about the respondents, or can fail to record contact attempts altogether. (West and Sinibaldi 2013).

International Attention

Despite the challenges, there is great hope that the successful use of paradata will grow. For one, new technologies are being developed to deal with Big Data, but also the increasing ease to collect these data will further fuel the development. Furthermore, as the statistical world moves towards the implementation of quality

metrics, paradata will play an important role in creating new indicators. The overall interest in paradata is increasing, reflected in special interest and invited paper sessions at various conferences in the past, such as the International Workshop on Household Survey Nonresponse, the conferences of the European Survey Research Association (ESRA), the conferences of the American and the World Association of Public Opinion Research (AAPOR/WAPOR), the Joint Statistical Meeting (JSM) of the American Statistical Association, the Colloque Francophone sur les Sondages, the workshops for Comparative Survey Design and Implementation (CSDI), and, not least, in sessions of the bi-annual conferences of the International Statistical Institute (ISI).

Beaumont, J.F. (2014). Une procédure adaptative pour la priorisation des appels téléphoniques lors de la collecte des données. Presentation at 8^{ème} Colloque Francophone sur les Sondages, Dijon.

Callegaro, M. (2013). Paradata in Web Surveys. In: *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, 263-282. Hoboken, NJ: John Wiley.

Charrance, G., S. Legleye, and N. Razafindratsima (2014). Comparaison de trois méthodes de correction de la non-réponse totale utilisant les paradonnées dans l'enquête téléphonique FECOND. 8^{ème} Colloque Francophone sur les Sondages, Dijon.

Conrad, F., R., Tourangeau, M., Couper, and C., Zhang (2012). Interactive Intervention to Reduce Satisficing in Web Surveys. Westat, Rockville, MD, Jan 24th 2012.

Conrad, F.G., J.S. Broome, J.R. Benkí, F. Kreuter, R.M. Groves, D. Vannette, and C. McClain (2013). Interviewer Speech and the Success of Survey Invitations. *Journal of the Royal Statistical Society, Series A*, 176, 1, 191-210.

Couper, M.P. (1998). Measuring Survey Quality in a CASIC Environment. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 41-49.

Couper, M.P., and F. Kreuter (2013). Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 176, 1, 271-286.

Couper, M.P., and J. Wagner (2011). Using Paradata and Responsive Design to Manage Survey Nonresponse. *Proceedings of the 58th World Statistical Congress*.

Daas, P., and M. Puts (2014). Big Data as a Source of Statistical Information. *The Survey Statistician*, 69, 22-31.

Eckman, S. (2013). Paradata for Coverage Error Research. In: *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, 97-120. Hoboken, NJ: John Wiley.

Groves, R.M. (2012). Discussant: What happens when Big Data Meets Official Statistics. <http://live.worldbank.org>

- Groves, R.M., and S. Heeringa (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society, Series A*, 169, 3, 439-457.
- Groves, R.M., Jr.F. Fowler, M.P. Couper, J. Lepkowski, E. Singer, and R. Tourangeau (2004). *Survey Methodology*, Wiley and Sons, Inc.
- Horwitz, R. (2014). *Classifying Mouse Movements and Providing Help in Web Surveys*. Dissertation: University of Maryland.
- Kreuter, F. (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. New York: John Wiley.
- Kreuter, F. and C. Casas-Cordero (2010). Paradata. Working Paper, German Council for Social and Economic Data (RatSWD), No. 136.
- Kreuter, F., K. Olson, J. Wagner, T. Yan, T.M. Ezzati-Rice, C. Casas-Cordero, M. Lemay, A. Peytchev, R.M. Groves, and T.E. Raganathan (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys. *Journal of the Royal Statistical Society, Series A*, 173, 2, 1-21.
- Kurkowski, K. (2013). Using Internet and Hand-held Computers for Data Collection in Poland. Presentation at 5th International Workshop on Internet Survey and Survey Methodology, South Korea.
- Lane, J., V. Stodden, S. Bender, and H. Nissenbaum (2014). *Privacy, Big Data, and the Public Good. Frameworks for Engagement*. Cambridge: Cambridge University Press.
- Miller, P. (2014). What Does Adaptive Design Mean to You? Presentation at FedCASIC, Washington DC.
- OIG (2014). Testimony of The Honorable Todd J. Zinser Inspector General U.S. Department of Commerce before a hearing of the Committee on Oversight and Government Reform Subcommittee on Federal Workforce, U.S Postal Service, and the Census U.S. House of Representatives U.S. Census Bureau: Addressing Data Collection Vulnerabilities September 18, 2014.
- Olson, K., and B. Parkhurst (2013). Collecting Paradata for Measurement Error Evaluation. In: *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, 43-72. Hoboken, NJ: John Wiley.
- Couper, M.P., and J. Wagner. (2011). Using Paradata and Responsive Design to Manage Survey Nonresponse. *Proceedings of ISI 2011*.
- Wagner J, B.T. West, N. Kirgis, J.M. Lepkowski, W.G. Axim, and S.K. Ndiaye (2012). Use of Paradata in a Responsive Design Framework to Manage Field Data Collection. *Journal of Official Statistics*. 28, 4, 477-499.
- Wagner J., R. Valliant, F. Hubbard and L. Jiang (2014). Level of Effort Paradata and Nonresponse Adjustment Models for a National Face-to-Face Survey. *Journal of Survey Statistics and Methodology*, 2, 4, 410-431.

West, B., F. Kreuter, and M. Trappmann (2014). Is the Collection of Interviewer Observations Worthwhile in an Economic Panel Survey? New Evidence from the German Labor Market and Social Security (PASS) Study. *Journal of Survey Statistics and Methodology*, 2, 2, 159-181.

West, B.T., and J. Sinibaldi (2013). The Quality of Paradata: A Literature Review. In: *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, 339-59. Hoboken, NJ: John Wiley.

New and Emerging Methods – Call for Volunteers

If you're interested in contributing an article to the "New and Emerging Methods" section of a future edition of *The Survey Statistician*, please contact Mick Couper at mcouper@umich.edu.



Book and Software Review

Using Administrative Data for Statistical Purposes Dan Hedlin, Stockholm University

The use of administrative data for statistical purposes is nothing new. John Graunt's book "Natural and Political Observations upon the Bills of Mortality", published in 1662, may well be the first publication with truly register-based statistics. Data on number of births and deaths in London had been collected and made available since the early 17th century. There were also lists of causes of deaths, collected by "searchers" in every parish. Graunt made use of these data sources to, for example, estimate the population of London, the ratio of women to men in London and the ratio of new-born girls to boys. He also produced time series of causes of deaths. John Graunt's book sold extremely well; the third edition was published only three years after the first edition. Jorner (2008) points out that Graunt did use the data with a critical eye, and was transparent about his assumptions. In his calculation of the population of London he estimates the average annual number of births to 12,000 (taken over years for which Graunt believed that data was of good quality). Then he assumed that there were twice as many fertile women in London. He doubled that number again to obtain the total number of women in London, and upon multiplying that number by eight he arrived at the population of London: 384,000. For a modern official statistician these calculations may seem a tad advantageous. The number 384,000 is more like a Fermi estimate.

In the 16th and 17th centuries Sweden built up an impressive administrative data collection infrastructure. The purpose was two-fold: to collect taxes and to raise an army, which was extremely large for a country with a small population. The first National Statistical Office (NSO) in the world (Jorner 2008, p. 27), *Kungliga kommissionen över tabellverket*, was founded in 1756. Needless to say, the estimated population size was kept secret.

With Sweden's long history of administrative registers and a highly efficient administrative data collection infrastructure, it is not surprising that the first English language book devoted entirely to register-based statistics is written by two Swedish statisticians: Anders Wallgren and Britt Wallgren (Wallgren and Wallgren, 2007). A few years prior to the English version, Statistics Sweden had printed a Swedish language book, an "R&D Report" (Wallgren and Wallgren, 2004a), accompanied with the much shorter official guidelines, Wallgren and Wallgren (2004b). This article focuses on the second edition of *Register-based Statistics*, Wallgren and Wallgren (2014), in the sequel abbreviated to WW. In WW the authors draw on more than 20 years of experience at Statistics Sweden. The authors have also given many courses on register-based statistics and presented numerous conference papers on the topic.

What has been missing since the days of John Graunt is a statistical theory of register-based statistics. It has been an open issue whether general statistical theory can account for the needs of register-based statistics. Wallgren and Wallgren advocate repeatedly in their books that it cannot. Holt (2007) asserts that research is sorely needed to develop a framework for assessing the quality of administrative data. Now there is a difference between the terms *framework* and *theory*, which I will come back to.

The survey approach to register-based statistics

As John Graunt realised, to use an administrative register for statistical purposes you must not just use it as it is. However, a fairly common but naive way of using administrative data for statistical purposes is to accept the data as they are, after some basic validations. The total number of units in a target population is estimated by counting the number of records in the administrative register, bar identifiable overcoverage. A better way is to collect *all* administrative data that are relevant for the specific statistics to be produced and let the set of data go through a process that will make them suitable for statistics production. This is a time-consuming and costly process but for a large NSO it will pay off in the long run.

In WW, Section 1.5, typical steps in the transformation process from a set of administrative registers to a statistical register are listed. The objects of the administrative registers, matches and nonmatches, are identified and treated. New objects and new variables are derived (common examples are to form a household out of individuals and to amend the reference time for some variables). There will be editing, treatment of missing values and coding. These and other transformation steps are elaborated on in the remainder of the book. This is the *survey approach* to register-based statistics (WW p. 259). A *register survey* refers to the process of transforming one or several administrative registers to a *statistical register* and make use of it to produce estimates.

Requirements for a system of register-based statistics

WW (p. 6) formulate two principles that identify the main preconditions for using administrative registers for statistical purposes:

1. Unified systems of identity numbers are used in all administrative systems. The same identity number should follow an object over its lifetime.
2. A statistical office should have access to administrative registers kept by public authorities. This right should be supported by law and the protection of privacy must also be protected by law.

There is another, albeit tacit, precondition: the public authorities who maintain the administrative registers must not change the contents of the registers too often. And if they do, they must tell the NSO.

Holt (2008) stresses the crucial role of a well-functioning statistical infrastructure, which apart from administrative registers that have been made available to the production of official statistics, also includes classification systems. The maintenance of the statistical infrastructure is rather unglamorous but absolutely vital for the quality and trustworthiness of official statistics. The same may be said about statistical science and statistical methodology.

Systems approach

The *systems approach*, or *system-oriented thinking*, is central in register-based statistics.

So what is this system-oriented way of thinking? Simply put, it is to use *all* administrative data the NSO have collected when constructing a statistical register. This in contrast to sample surveys, which are often more or less produced in isolation from other sample surveys (sometimes referred to as ‘stove-pipes’). Matching samples from different social surveys can seldom be done since the samples are largely disjoint. (Data fusion is not discussed in WW, see Rässler 2004). Many NSOs, however, have a ‘large businesses unit’ where all information on the very large businesses obtained from all registers and sample surveys are collected and assessed.

Base registers and standardised variables

The *base registers* are the focal point in the register system. They define the populations of all surveys carried out by the statistical office (WW, p. 280). Note that the term survey includes register surveys as well as censuses and sample surveys. WW identify four base registers, see Figure 1.

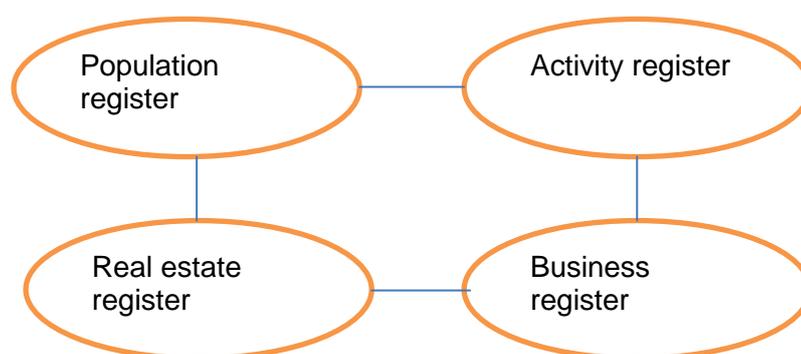


Figure 1. The base registers of a register-based survey system

Each base register contains more than one object type. For example, the population register contains the object types individual, event, family and dwelling. It is the dwelling that is the link to the real estate register, of which houses and flats are a proper subset. The activity register is a collection of relational objects that connect the objects in the population registers with those of the business register, such as ‘working’ or ‘studying’. The business register should perhaps be referred to as ‘business and organisation register’.

The base registers are not administrative registers; they are statistical registers. Particularly important variables are referred to as *standardised variables* in WW. One example is employment status in the employment register. The NSO should have a clear organisational structure for the maintenance of standardised variables.

Although WW make frequent references to Statistics Sweden, the description of the base registers should be understood as an abstract picture of the basis of a well-developed system for register-based statistics. Even more, Figure 1 – having been made complete with the relations between the base registers – is the core of a

conceptual model of a register system of statistics on society. (Actually, Statistics Sweden recognise three base registers, see Holmberg et al., 2011).

Protection of privacy

One of the benefits with register-based statistics is the relative ease by which estimates for small domains can be produced. However, there will be confidentiality issues. It may seem impossible to publish data for small domains and at the same time protect individuals' and businesses' confidentiality. The good news is that there is considerable research on this and there are solutions proposed in the literature, see for example Shlomo (2007 and 2009) and Hundepool et al. (2012).

Chapter 3 in WW addresses protection of privacy and confidentiality. This chapter is rather insufficient and does not contain modern references. Very little is said about statistical disclosure control (SDC), with suppression of sensitive table cells as the only method mentioned. Suppression, although common among practitioners, is in general not a method that can be recommended since it tends to lead to great losses of information. Many EU member states worked on SDC methods when preparing for the 2011 European Union censuses and the references above give a very useful overview.

Lack of validity – relevance error

The most severe downside with register-based statistics is often believed to be lack of validity (more commonly referred to as *relevance* in the survey community). The NSO can only have whatever there is in the administrative systems, as opposed to sample surveys where the NSO can (in theory!) define the target population and the study variables, although as for the feasibility to actually reach the sampled population units and to collect the desired data a lot can be said. WW give examples (e.g. p. 285) of how the model-error in a derived variable can be estimated and presented. The approach is quite simple; first you decide which data set you trust, that's your *test data*, to which the model is fitted, then the model is applied to another data set to predict values of *derived variables*. This frequently entails extrapolation beyond the range of data. WW note this fact but do not attempt any guidelines on what and how much extrapolation one can do.

To account for missing data, WW prefer imputation to weighting for practical and consistency reasons. They recommend random imputation for missing values in qualitative variables but rather mysteriously not for quantitative variables (p. 217 and 286).

Since prediction brings about statistical uncertainty one would have expected in WW some statistical treatment of this uncertainty. However, WW refrain from the issue with the following arguments (p. 285-286):

“Section 15.1 discusses statistical inference and register-based statistics. We are sceptical regarding confidence intervals for register-based estimates for two reasons:

- *the users of register-based statistics often base their conclusions on simultaneous comparisons of many, sometimes hundreds, estimates and*
- *there are non-random errors that invalidate confidence intervals.*

...These inference methods should not be used to publish thousands of standard errors for all estimates; this will only confuse users."

Other chapters

WW covers as far as I can see the main processes of register-based statistics, excluding perhaps presentation of the statistics in the final NSO publications. Other chapters than those I have touched upon above include a set of chapters (Ch. 6-9) going through the *"fundamental estimation methods"* (p. 202). All chapter headings in this part of the book start with "How to create a register" followed by, in turn, "Matching and combining sources", "The population", "The variables" and "Editing". After a chapter on metadata, there is a set of chapters (Ch. 11-14) with *"supplementary estimation methods"*. Their headings start with "Estimation methods"; they are "Introduction", "Missing values", "Coverage problems" and "Multi-valued variables". The book ends with Ch. 15 "Theory and quality of register-based methods" and Ch. 16 "Conclusions".

Intellectual basis

On page 11 in WW the authors make a distinction that I believe is important to understand the intellectual basis for this book:

"Sample surveys are based on a mathematical theory – probability and inference. ... Register surveys require a non-mathematical theory based on a systems approach."

On page 263 this distinction is elaborated on:

"The classical sample survey theory, as presented by Cochran (1963), for example, was clearly based on probability theory and statistical inference theory. We think that this theory has shaped the paradigm of many statisticians – a theoretical approach, whose principal aim is to develop expressions for bias and variance of estimators. The theory is expressed by probability distributions and formulas, and without probability distributions and formulas you have no theory. This paradigm is not adequate for register surveys and is no longer adequate for sample surveys for two reasons:

- *Sample survey theory and time series theory are today two separate and inconsistent fields within statistical science. If sample survey theory were considered complete, these two fields would be integrated, as the main purpose of sample surveys is to produce time series and estimate time series patterns.*
- *The link between academic inference theory and sample survey methodology is becoming increasingly weaker as we no longer can disregard the nonsampling errors. This is another reason why probability-based assessment of errors in estimates can no longer be considered as a distinguishing feature of sampling theory.*

We think that sample survey theory and register survey theory are similar, as the theory is (only) a collection of concepts and principles. The concepts and principles used for sample surveys are widely known, but the concepts and principles used for register surveys are not yet widely understood. However, as these principles are general, they should be regarded as a general survey theory; they hold for all surveys, all registers and all countries."

These quotes highlight one of the themes in WW: (classical) statistical inference is of limited use in register-based statistics. And, apparently, even in sample surveys. That's a rather unusual view to take.

One main aim of WW is to formulate a basis for a theory of register-based statistics (p. 1). A theory is "*a collection of concepts and principles*" (see above). In science, a theory is usually understood as a set of theorems and hypotheses, with of course a language to be able to express statements, which includes concepts and terminology. Viewing a theory as a set of concepts and principles is again rather unusual. (An example of a principle: "*Administrative registers should be transformed into statistical registers. All relevant sources should be used and combined during this transformation*". This is the "*Transformation principle*", p. 3). WW present a framework (concepts, principles, methods, terminology) rather than a theory in a scientific sense.

Comparing the reference lists in WW and that of Li-Chun Zhang's article on register-based statistics in the *Survey Statistician* (Zhang, 2014), one will find hardly any overlap, apart from standard references in the survey field such as Biemer and Lyberg (2003). Why? I think the main reason is the different intellectual bases. For Zhang it is natural to use probability theory and statistical inference. The references in Zhang (2014) not included in WW discuss a total error framework, linkage error, coverage error, estimation of reporting delays, confidentiality and relevance error.

Many opinions

In WW the authors do not shy from stating their opinions; however, in many cases they are less than convincing. One example was given above, under *Lack of validity – relevance error*. Let me give another example from WW, p. 29 (the cognitive processes referred to are those often attributed to Tourangeau (1984), that is, comprehension, retrieval, judgement, communication):

"Are the same cognitive processes relevant when persons report data to an administrative authority, either as a private person or as representative of an enterprise? These cognitive or psychological processes also exist in connection with administrative reporting, but we do not think they are important. Instead, administrative rules and legislation are important factors, and when reporting data from enterprises, accounting principles and practice are more important than psychology." The sole argument for this opinion seems to be (p. 34):

"Many administrative variables are precisely defined by administrative rules and legislation, and those who provide the data must follow these rules. In a sample survey, the answers depend on how the respondents interpreted the questions and our knowledge of such cognitive processes is as a rule very limited."

Of course, it is in WW acknowledged that not all administrative data are legally important and hence the quality of that kind of administrative data may not be very high.

Style of writing

As always with the writings of Anders Wallgren and Britt Wallgren, there are many excellent tables and graphs. Their use of contrasting examples will also be well known to their readers. In WW they have chosen to use mathematical formulae very

sparingly. The reason is probably to make the book more widely accessible but the downside is lack of precision. Not even adjustment of measurement error through a regression model (p. 56-57) is described with the help of mathematical notation. I find this passage hard to understand (prediction of a binary variable with linear regression, did I miss something?).

Instead of mathematical language, the methods are explained through many detailed, very 'hands-on' examples.

The text type is a mixture of exposition with many opinions stated and procedures with all steps to go through described in detail (although the theoretical arguments may be lacking) and discussion with pros and cons.

First edition vs second edition

While the first edition was similar to the Swedish version, Wallgren and Wallgren (2004a), the second edition is different from the 1st ed. The outline of the book has been restructured. There is a stronger emphasis on the systems approach, and the numerous principles have been amended and restructured. Some chapters are largely the same (e.g. Ch. 3, Protection of privacy and confidentiality) but many chapters have been re-worked. Although the list of references is surprisingly short for a book with such a wide scope, there has been a large turnover of references from the first edition.

Need to buy the 2nd edition?

So if you have access to the first edition, do you need to buy the second edition? Yes, if you are into the area, you probably do. It is and will be the standard reference book on register-based statistics. There might be an alternative some day, but today there isn't.

References

Biemer, P. P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*. Hoboken: Wiley.

Jorner, U. (2008). *Summa summarum: SCB:s första 150 år*. Statistics Sweden. (In Swedish).

Holmberg, A., Blomqvist, K., Engdahl, J., Irebäck, H., Lundell L.-G. and Svensson, J. (2011). A strategy to improve the register system to store, share and access data and its connections to a generic statistical information model (GSIM). Invited paper presented at Work Session on Statistical Data Editing, UNECE, Ljubljana, Slovenia, 9-11 May.

Holt, D. (2007). The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper. (With discussion). *The American Statistician*, 61, 1-15.

Holt, D. (2008). Official statistics, public policy and public trust. *Journal of the Royal Statistical Society, Series A*, 171, 1–20.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. and de Wolf, P.-P. (2012). *Statistical Disclosure Control*. Chichester: Wiley.

Rässler, S. (2004). Data Fusion: Identification Problems, Validity, and Multiple Imputation, *Austrian Journal of Statistics*, 33, 153-171.

Shlomo, N. (2007). Statistical Disclosure Control Methods for Census Frequency Tables. *International Statistical Review*, 75, 199-217.

Shlomo, N. (2009). Statistical Disclosure Control for the European Census Dissemination. Joint UNECE/Eurostat work session on statistical data confidentiality. 2-4 December, Bilbao, Spain.

Tourangeau, R. (1984). Cognitive Sciences and Survey Methods. In Janine, R. et al. (eds), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, Washington D.C. National Academy of Science, 73-100.

Wallgren, A. and Wallgren, B. (2004a). Registerstatistik– administrativa data för statistiska syften. R&D Report 2004:2. Statistics Sweden. (In Swedish).

Wallgren, A. and Wallgren, B. (2004b). Registerstatistik – en introduktion. Handbok 2004:1. Statistics Sweden. (In Swedish).

Wallgren, A. and Wallgren, B. (2007). *Register-based Statistics: statistical methods for administrative data*. Chichester: Wiley.

Wallgren, A. and Wallgren, B. (2014). *Register-based Statistics: statistical methods for administrative data*, 2nd Ed. Chichester: Wiley.

Zhang, L.-C. (2014). Data integration. *The Survey Statistician*. No. 70, July 2014. The International Association of Survey Statisticians.

We are interested in fostering review of books and software in the area of survey methods. This would include standard review of individual books or software packages. This may also include broader reviews of groups of text and monographs in specific sub-areas; or similarly broad reviews of available software. Of particular interest are some of the new R libraries that have been developed recently for survey methods. If you are able to write a review for this section, please contact Natalie Shlomo (natalie.shlomo@manchester.ac.uk).



AUSTRALIA

In June 2013 the Australian Bureau of Statistics (ABS) released a revised series of historical estimated resident population (ERP) data for the September 1991 to June 2011 period. This revision process is referred to as 'recasting' the data, and the scale of this change was unprecedented in the history of Australia's population estimates.

Why Recasting Was Necessary

Following each Census, the ABS rebases ERP to the latest Census count. As part of this process, the ABS adjusts ERP to take into account any 'intercensal difference' (i.e. the difference between unrebased estimates and the preliminary Census estimate) by distributing the difference evenly across each quarter over the five year period since the previous Census. However, preliminary rebasing of ERP after the 2011 Census could not credibly account for the large intercensal difference between 2006 and 2011, and the resulting ERP series showed implausible growth for the five-year period.

The large intercensal difference identified was primarily due to significant improvements in the methodology used in the 2011 Census Post Enumeration Survey (PES). The PES is conducted after each Census to assess coverage of Census counts, as represented by the key measure of net undercount. In 2011 the PES utilised the new methodology of Automated Data Linking (ADL), resulting in better linking and matching of PES and Census records, and therefore a better measure of net undercount. The ABS estimated that the 2011 net undercount was approximately 40% lower than it would have been if the previous PES methodology had been used, and if the ADL innovation had been available in 2006, the 30 June 2006 population estimates would have been lower by more than 200,000 people.

As a result of this large adjustment to population numbers, and following an extensive consultation process, the ABS made the decision to recast the ERP series back to 1991. This 20-year revision period was decided on as this produced estimates of population over the period that best reflected the growth observed in the historical data for population components (that is births, deaths and migration), which are the best source for measuring population change over time.

For information about the recasting process, see the feature article: [Recasting 20 Years of ERP in Australian Demographic Statistics, Dec 2012](#) (ABS cat. no. 3101.0). For information about the Post Enumeration Survey see [Census of Population and Housing - Details of Undercount, 2011](#) (ABS cat. no. 2940.0). Also see the Technical

Note on the Statistical Impact of ADL within the same publication under the Explanatory Notes tab.

For more information, contact Anthony Russo (anthony.russo@abs.gov.au)

BOSNIA AND HERZEGOVINA

Populaton and Housing Census after 22 years

Bosnia and Herzegovina has conducted Population and Housing Census in the period 1-15 October 2013. The census was conducted in a traditional way by interviewing people in 24,322 enumeration areas. More than 24,000 people were engaged in data collection process.

After more than two decades, the country will get basic statistical data on socio-economic and demographic characteristics of the population. The total Census budget is about 22 million euros and it includes all activities until the final data dissemination. First Census results were disseminated at the beginning of November 2013, while final data will be presented in stages in 2015.

In order to measure the quality of the Population and Housing Census data, the Post-enumeration survey (PES) was conducted in the period 02-10 November 2013. In 240 randomly selected enumeration areas, about 12,000 households or approximately 38,000 people was re-interviewed. The Post-enumeration survey will provide indicators of the census quality in terms of its coverage and content. First PES results will be available in the first semester 2015.

For more information, contact Edin Šabanović (edin.sabanovic@bhas.ba), Sector for Statistical Methodology, Standard, Planning, Quality and Coordination, Agency for Statistics of Bosnia and Herzegovina.

BRAZIL

The National School of Statistical Sciences Inaugurates a Doctorate on Population, Territory and Public Statistics

The National School of Statistical Sciences (ENCE) is an academic School linked and supported by the Brazilian Institute of Geography and Statistics (IBGE). The academic courses offered by ENCE are accredited by the Ministry of Education, and therefore its degrees have the same status as those offered by any of the conventional universities in Brazil.

ENCE is now inaugurating a research degree leading to a doctorate in Population, Territory and Public Statistics. The selection process for both the MSc and Doctorate

opens yearly in August-September. For each selection process, a call for applications is published at the Post-graduate space in the website: <http://www.ence.ibge.gov.br/>. Each academic year starts in March. Teaching and assessment is in Portuguese, but English language proficiency is also required.

More information about the Doctoral Programme is available (in Portuguese) at <http://www.ence.ibge.gov.br/index.php/pos-graduacao/mestrado-doutorado/apresentacao>

3rd IBGE Seminar on Methodology – SMI 2014 – Registrations close 27 October

The 3rd IBGE Seminar on Methodology - SMI2014 was held from 5 to 7 November 2014, in the Centre for Documentation and Dissemination of IBGE in the city of Rio de Janeiro.

The seminar's theme was "**Challenges and opportunities for obtaining data.**" Technical innovation and cultural change impose challenges, but at the same time, offer opportunities for statistical agencies to develop and adopt new approaches to obtain data. These include growing use of administrative records, combining data sources, mixed modes of data collection, technology intensive data collection, big data and automated data capture, amongst other approaches aimed at getting more detail, increased timeliness, broader coverage and better quality data, while at the same time reducing response burden.

More information is available at <http://eventos.ibge.gov.br/smi2014>

For more information, contact Mr. Cristiano Ferraz (cferraz@de.ufpe.br)

CANADA

New survey at Statistics Canada: Job Vacancy and Wage Survey (JVWS)

In order to fill gaps in labour statistics, a new survey called Job Vacancy and Wage Survey (JVWS) has been launched in July 2014. It will consist of a quarterly sample of 100,000 business locations in Canada on a rotation basis. All economic sectors will be represented, with the exception of public, provincial, and territorial administrations, which will be included at a later date. The main goal of the job vacancy component is to produce estimates of the number of job vacancies and vacancy rates by economic region and occupation. The annual wage component will measure average wages by occupation. Data from related surveys shows that having vacant positions is a semi-rare characteristic. At any given time, about 15% of smaller businesses and 30% of larger ones have vacant positions.

The JVWS questionnaire is based on a model developed by the National JVS Workgroup (USA) and most recently used in Kansas in 2013. Electronic questionnaires will be the only mode of collection for the survey. A short questionnaire will prompt respondents for basic information on each vacant position: type of position (full-time/part-time, permanent/temporary), offered wage, education

and experience levels required, recruitment strategy, and amount of time the position has been vacant.

Collection for the quarterly job vacancy component of the survey is planned to start in February of 2015 for a first release in August 2015. The first release for the annual wage component is planned for late 2016.

For further information, please contact: Danielle Lebrasseur (danielle.lebrasseur@statcan.gc.ca) or Étienne Rassart (etienne.rassart@statcan.gc.ca).

ESTONIA

The Workshop of the Baltic-Nordic-Ukrainian (BNU) Network on survey statistics was organized by Statistics Estonia and University of Tartu on 25-28 August 2014. The Network <http://wiki.helsinki.fi/display/BNU/Home> has been active since 1996. It has gradually grown and involves now also Belarus.

There were 50 participants attending the workshop: Estonia 10, Ukraine, Sweden and Finland 7, Lithuania and Latvia 6, Belarus 5, and 3 from other countries.

During four days there were two hour long tutorial lectures by three key speakers, presentations by six invited speakers and 26 short presentations by the participants. All contributed presentations had nominated discussants and it was followed by general discussion.

The key lectures covered new and very important areas of survey statistics: responsive design introduced by Prof. Carl-Erik Särndal, small area estimation and poverty indicators explained by Prof. Isabel Molina, and analytic inference for survey data presented by Prof. Jean Opsomer. The invited speakers covered very wide range of topics, as well. Among others was mixed mode data collection, sampling designs with unequal inclusion probabilities, use of mobile positioning data in official statistics, evidence based policies on aging.

The workshop has been and still is a meeting place for students, teachers and practitioners where students can present their research and ask for advice; practitioners can bring up the practical issues and give research ideas to people from academia. For participants from Ukraine and Belarus, but also from Baltic countries, it is a place to learn from experience of Nordic countries where survey statistics has much longer history.

The next event of the BNU Network will be international conference on survey statistics to be held in Finland in 2015.

We thank IASS for supporting this event and thus giving opportunity to several participants to attend the workshop.

Program and abstracts of the presentations can be found at <http://www.stat.ee/workshop-on-survey-statistics-theory-and-methodology-2014>

For more information about the event, please contact Maiki Ilves at maiki.ilves@stat.ee.

For more information, contact Ms. Gayatri Vishwakarma (c gayatri_v14@rediffmail.com)

INDIA

Dr. Gayatri Vishwakarma

With the growing population of statisticians and programmers in India, representatives from various multinational pharmaceutical companies and CROs took the initiative to establish a framework for an organization to provide a platform for statisticians and statistical programmers to meet and learn and to share knowledge and experience. Henceforth the Indian Association for Statistics in Clinical Trials (IASCT), which is the new initiative, was founded in 2007 in India. The mission of IASCT is to promote biostatistics and statistical programming in clinical research as career options for students of statistics and other technical disciplines in India. Also to enable professional development of statisticians and statistical programmers by organizing training sessions, meetings and conferences relating to statistical techniques used in drug development.

IASCT conducts several events round the year to help enhance and enrich the Statistics and Programming knowledge base in India, in line with its mission. IASCT Annual conference and **ConSPIC** are the major annual events which attract a lot of participation from the Biostatistics and Clinical Programming fraternity in India. Industry experts from India and abroad share their expertise and current global trends and advances.

The format of these conferences is very interactive and facilitates networking and learning from experts and peers. The annual conference caters to the wider clinical research community and thus provides a forum for interaction with other groups. IASCT conducts several workshops throughout the year wherein subject matter experts in specific areas of interest train participants in topics which are useful and applicable in their daily work. These focused workshops are great learning opportunities. Some of the recent workshops were on Survival analysis, Multiple Test Procedures for Gatekeeping, Statistics for non-statisticians, 'R' language for programmers. Recently IASCT conducted a successful **ConSPIC 2014** at Delhi (i.e. the Fifth **ConSPIC**) where more than 100 statisticians and programmers participated from government as well as private organizations. The next **ConSPIC 2015** will be held in November 2015 at Mysore, Southern part of India.

For more information, contact Ms. Gayatri Vishwakarm at (gayatri_v14@rediffmail.com).

NEW ZEALAND

Felibel Zabala

Linking methodology used in Statistics NZ's Integrated Data Infrastructure

Statistics NZ's Integrated Data Infrastructure (IDI) team has published a report, [Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure Project](#), explaining the linking methodology used in the IDI. This report contains details about how we are doing the linking, provides practical examples, and outlines some of the wider issues associated with using linked information in analysis. The report has been peer reviewed, including by the Office for National Statistics in the United Kingdom.

For further information on Statistics NZ's IDI, please contact Anapapa Mulitalo at anapapa.mulitalo@stats.govt.nz.

Using big data to measure price change

Compiling price indexes from 'big data', such as scanner data and web-scraped online data, presents particular methodological challenges for price measurement. The Prices Unit at Statistics NZ have been researching this area for the past five years.

Many products have their barcodes scanned when they are purchased. This retail transaction data – or 'scanner' data – records prices, quantities sold, and associated information for all transactions across a full reference period.

We are now being supplied with scanner data from market research company GfK for a range of key consumer electronics categories, such as televisions and computers. We intend to use this scanner data in the CPI from the September 2014 quarter, using the imputation Tornqvist rolling year GEKS (ITRYGEKS) method we developed with Statistics Netherlands. The co-authored paper on this approach, [Scanner data and the treatment of quality change in nonrevisable price indexes](#), was published recently in the *Journal of Business & Economic Statistics*. The ITRYGEKS method incorporates statistical models to adjust the measurement of price change appropriately for the changing mix of characteristics of the products being bought by consumers. We will be leading international practice in making such comprehensive use of consumer electronics scanner data in our CPI.

Prices senior researcher Frances Krsinich presented a paper at the UNECE/ILO meeting of the group of experts on consumer price indices in Geneva in May. 'Fixed-effects with a window splice: quality-adjusted price indexes with no characteristic information' describes the approach we are considering to estimate quality-adjusted price indexes from scanner data and online data where there is no information available on product characteristics. In this situation, explicit quality adjustment with methods such as the ITRYGEKS methodology mentioned above, cannot be used. The paper shows that the relatively simple fixed-effects (or time-product dummy) index is equivalent to a fully-interacted time dummy hedonic index based on all price-

determining characteristics of the products, despite those characteristics not being observed. In production, this can be combined with a modified approach to splicing that incorporates the price movement across the full estimation window to reflect new products with one period's lag, without requiring revision of previously published movements.

For more information email frances.krsinich@stats.govt.nz.

SPAIN

Cristina Prado and Carmen Guinea

Quality in the gathering of online data: Standardising online questionnaires, integration with administrative sources and development of bias control mechanisms

The Project for standardizing electronic questionnaires has allowed Eustat (Basque Country Statistical Office) the continuous improvement of data collection processes, adapting these processes to the use of Internet technology and new technologies to improve data quality, reduce response bias and non-responses and increase the precision of statistical data. Currently we used the on-line platform "WEB Channel" for all surveys, whatever the subject they have: economics, social-demographics, short times surveys or structural. In October, the Labor Force Survey was used in this on-line platform for the first time.

In "WEB Channel" the electronic questionnaires were designed from the following points of view:

- **Data quality:** With information collected via electronic questionnaires, the knowledge of the professionals who prepared the questionnaires had to be transferred to the online questionnaires so that the data quality was not affected, or could even be improved. In our project, the questionnaire quality mechanisms have centred on the following aspects:

Design of the questionnaire: One style guide, which include all the questionnaire design standards defined, was prepared. For that, we studied the recommendations made by international experts such as Mick Couper.

Information verification systems: In order to avoid erroneous data as far as possible, we have designed a validation and error control system that aims to minimise possible errors in answers. Three different types have been defined: Page, Cohesion and Length validations. The page and cohesion validations are organised into "hard" and "soft" in a manner that is user-friendly.

Browser: We have defined a browser logic that makes it possible to change pages using sequential progress and includes a graph control system.

- **Data security:** Regarding data security, a pre-questionnaire has been designed as a single starting point for all electronic questionnaires on all statistical

operations. It includes various security measures for both data protection and control of access, and records other information of interest that allows the "workload" of the surveyed individual in completing the questionnaire to be identified and analyzed.

- **Integration with administrative sources**

The administrative data is available, into the electronic questionnaires for Suggested response, as an optional response that can be modified, or for Response Verification. The administrative data are used to verify the answers given to certain questions as internal mechanisms for bias control.

- **Costs**

The system developed in turn allows for a reduction in costs both in the production of new electronic questionnaires and in the maintenance of existing ones, thanks to the creation of a software repository that is established as the single framework for building this software.

- **Results**

According to data, the percentage of data collection via the Web channel has increased over the years, especially the economic questionnaires with a very high percentage of responses exceeding 60% of the total. But the social and demographics surveys are not responded more than 5%, is that why we are focus on improved this rate with different measures which will be studied by a teamwork during next months in order to be applied next data collection period from spring on.

For more information, contact Mr.Miguel Angel Martínez Vidal (miguelangel.martinez.vidal@ine.es)

SWITZERLAND

Construction of commuter data for multi-establishment enterprises

In the past the Swiss population census, which was held every ten years (last time in 2000), provided for every working person the departure-to-work municipality and the start-working municipality. Thus commuter movements were known at municipality level.

Nowadays the traditional census was replaced by a register based approach where the unique information about commuter movements are based on the departure-to-work municipality and the employing enterprise. Therefore, for single-establishment enterprises (*consisting of only one local unit*) the needed information is directly available. But how to handle multi-establishment enterprises with establishments belonging to various municipalities? In this case, our approach consisted in imputing the number of commuters from its employee's municipalities to its establishments,

respecting certain constraints. Below we describe the available data and the algorithm we used in some detail.

There are three data sources related to multi-establishment enterprises available: Person data with sex and starting municipality variables, structural enterprise data of establishments with location (=ending municipality) and number of female and male employees variables, distance data with driving time between all Swiss municipalities. At the reference date, December 2011, there were 2500 different municipalities in Switzerland and nearly half of the 4 million active people worked in multi-establishment enterprises.

Our approach is based on a linear optimization algorithm: allocate all persons of an enterprise to its establishments respecting their sizes as much as possible and minimizing the sum of the resulting distances. This approach was refined by taking a given secondary residence as starting municipality if it was closer to the establishment municipalities than the primary residence, thus simulating weekly residents. Furthermore, this procedure was executed separately for each sex. Finally unrealistic distances (>120 minutes) were set to missing and the results were adjusted. In the case of a large number of combinations between starting and ending municipality the algorithm becomes computation intensive.

The results of this matching process were commuter movements that is to say the number of persons who did a certain starting-ending-municipality movement. The resulting commuter movements correlated well with direct estimations at an aggregated level based on the structural survey of the Swiss population census. For a single person the exact establishment location and hence the ending municipality is not known based on this matching process.

Further work will focus on using a more detailed propensity model, based on individual information like age, income and economic activity code of the establishments.

Please contact Monika Ferster (monika.ferster@bfs.admin.ch) at the Federal Statistical Office for further information.

UNITED KINGDOM

Peter Lynn

The Office for National Statistics recently re-established a programme of Quality Reviews of major statistical products and processes. The first review carried out under the new programme was of the Labour Force Survey, the first review of the LFS since 2002. The review (<http://bit.ly/1ETtV1e>) concluded that the survey is broadly fit for purpose in its current form, and is comparable in quality to the LFS in other European countries. However, recommendations were made regarding future improvements and research including assessment of potential alternative data collection periods and reference periods, consideration of introducing non-response

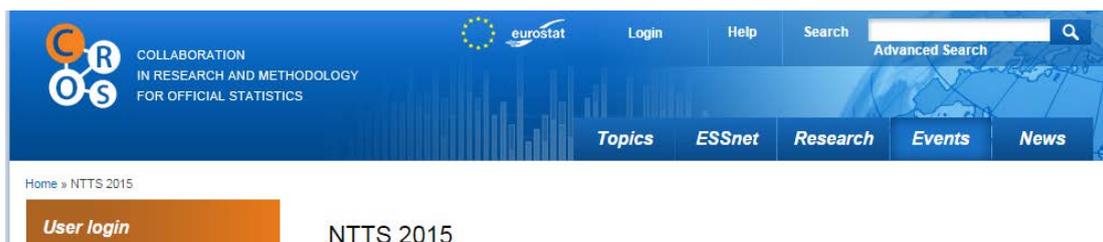
weighting, exploring ways of reducing sample attrition, and moving towards the introduction of web data collection.

The UK Government Statistical Service has a Methodology Advisory Committee, the purposes of which are, a) to provide a forum to allow government statisticians to obtain advice on methodological issues from a group of interested and experienced professional statisticians from outside government; b) to provide an opportunity to build and strengthen links between the Government Statistical Service (GSS) and the rest of the statistical profession. The committee meets twice a year and minutes and papers from all meetings are available on the ONS website (<http://www.ons.gov.uk/ons/guide-method/method-quality/advisory-committee/index.html>). Survey topics discussed recently include disclosure control methods for longitudinal survey data, and an experimental test of web survey methods for the Community Life Survey, a survey previously conducted by CAPI.

Understanding Society: the UK Household Panel Study is adding a boost sample of ethnic minorities and immigrants to the panel, using face-to-face screening of a disproportionately stratified probability sample of 19,000 residential addresses. The screening will take place throughout 2015, with a predicted yield of around 2,400 responding adults in each of three target groups: UK-born ethnic minorities, immigrant ethnic minorities, and other immigrants. The recruited respondents and their households will then be interviewed at annual intervals using the same survey instruments as the remainder of the sample. A working paper describing the sample design will be published online (<https://www.understandingsociety.ac.uk/research/publications/working-papers>).

The latest edition of the ONS Survey Methodology Bulletin (<http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-72/index.html>) includes articles on assessing the quality of data linkage, adjusting for over-count in the 2011 Census, improving advance notification letters for social surveys, and identifying optimal times for interviewers to visit addresses based on analysis of call record data.

For more information, contact Prof. Peter Lynn (plynn@essex.ac.uk)



New Techniques and Technologies for Statistics (NTTS) 2015

Organized by: Eurostat
Where: Brussels, Belgium
When: 10.03.2015 to 12.03.2015
Homepage: <http://www.cros-portal.eu/content/ntts-2015>

New Techniques and Technologies for Statistics (NTTS) is an international biennial scientific conference series, organised by Eurostat, on new **techniques** and methods for official statistics, and the impact of new **technologies** on statistical collection, production and dissemination systems.

The purpose of the conference is both to allow the **presentation of results** from currently ongoing research and innovation projects in official statistics, and to stimulate and facilitate the **preparation of new innovative projects** (by encouraging the exchange of views and co-operation between researchers - including the possible building of research consortia) with the aim of enhancing the quality and usefulness of official statistics and to prepare activities related to research in statistics within the European Framework Programme for Research and Development (Horizon 2020).

Online registration is mandatory if you wish to attend NTTS 2015. The Online registration can be found [here](#) and the registration interface is scheduled to be open from mid-December 2014.

30 Years with Journal of Official Statistics



We invite you to join survey research and official statistics colleagues for three days of conversation, networking and intensive learning at the JOS 30th Anniversary Conference.

30th Anniversary Conference organized by Journal of Official Statistics (JOS)

Organized by: Journal of Official Statistics Editorial Office
Where: Stockholm, Sweden.
When: 10.06.2015 to 12.06.2015
Homepage: <http://www.scb.se/jos-anniversary-2015>

Together we will be discoursing pressing issues of survey methodology for improving quality of official statistics across the globe as well as looking back 30 years of research addressed by Journal of Official Statistics.

An open access journal, JOS has published since 1985 theoretical and applied articles on survey methodology for statistics production within official statistics, policy decision making, and the economic and social sciences. In keeping with JOS tradition of high quality methodology research, the 30th JOS Anniversary Conference will provide programs, such as plenary sessions, panel discussion, and poster presentations as well as networking opportunities. Short courses offered by leading survey methodologists will cover: Bayesian method (Rod Little), questionnaire design and pretesting (Gordon Willis), survey research using emerging technologies (Trent Buskirk), and nonresponse (Kristen Olson).

And we also invite you to respond to the Call for Poster Sessions to present your work on current and pressing issues of research in survey methodology for official statistics.

Stockholm, the host city in Sweden for the JOS 2015 conference, offers a broad range of opportunities for spending time with colleagues and family. Registration and Course fees: One Course 2 500 SEK (you may attend maximum of two courses), Plenary Session Day 2 500 SEK. Early Bird discount (10%) for registrations before March 15, 2015.

For more information, please visit the homepage or contact jos@scb.se



ITACOSM 2015, The 4th ITALian CONference on Survey Methodology, A bi-annual international conference promoted by the Survey Sampling Group (S2G) of the Italian Statistical Society (SIS).

Organized by: Italian Statistical Society
Where: Rome
When: 24.06.2015 to 26.06.2015
Homepage: <http://itacosm15.sta.uniroma1.it/>

ITACOSM 2015 - 4th ITALian CONference on Survey Methodology is a bi-annual international conference promoted by the Survey Sampling Group (S2G) of the Italian Statistical Society (SIS).

ITACOSM 2015 focuses on themes related to survey sampling methodology, in human as well as natural sciences, with special emphasis on design and analysis of observational studies.

A (non-exhaustive) list of specific topics of interest for the conference is listed below.

- Sampling finite populations: design and data analysis;
- Survey sampling methodology;
- Small area estimation and related problems;
- Nonsampling errors / missing data;
- Opinion polls;
- Data production, data dissemination and data security, especially in Official Statistics;
- Data quality;
- Data integration;
- The use of big data in statistical surveys;

ITACOSM 2015 includes plenary (invited) sessions on hot themes, specialized (invited) session on specific topics, contributed sessions, and poster sessions.

Proposals of solicited sessions, i.e. spontaneous sessions composed by "homogeneous" talks of 3-4 authors, are encouraged. Among proposals (if any), a number of 2-3 solicited sessions will be chosen. The deadline for proposing solicited sessions is December 20, 2014. For more information, please visit the homepage or contact info_itacosm15@uniroma1.it



6th Conference of the
European Survey Research Association

Reykjavik, Iceland, July 13-17, 2015



The 6th Conference of the European Survey Research Association

Organized by: European Survey Research Association (ESRA)
Where: Reykjavik, Iceland
When: 13.07.2015 to 17.07.2015
Homepage: <http://www.europeansurveyresearch.org/conference>

The 6th Conference of the European Survey Research Association (ESRA) will take place 13th-17th July 2015 in Reykjavik, Iceland.

The scientific committee invites researchers who are active in the field of survey research and survey methodology to submit paper proposals.

To submit a presentation abstract, [sign up](#) or [log in to your ESRA account](#), click "Submit a paper abstract" and follow the instructions. You may submit a maximum of two papers on which you are enlisted as presenting author.

The closing date for submission of paper proposals is **15 January 2015**.

Paper proposals are invited in any area of survey methodology, or in substantive applications of survey research. We encourage proposals from researchers with a variety of backgrounds, including academic research, national statistics and market research.

The following are examples of topics that are of particular interest:

- Sample designs, coverage, and sampling
- Fieldwork processes
- Unit and item nonresponse
- Weighting and imputation
- Questionnaire development, testing and piloting
- Interviewers and interviewer effects
- Mixing modes and mode effects
- Online survey methods
- Surveys using mobile devices
- Linking survey data to auxiliary data sources
- Using paradata to evaluate survey quality
- Methods for cross-national and cross-cultural surveys
- Longitudinal surveys and longitudinal analysis techniques
- Methodological considerations specific to certain survey modes: face-to-face, phone, web, mail, etc.
- Analysing, monitoring and reducing the Total Survey Error
- Data documentation, archiving and data access
- Survey analysis techniques

- Election polling and public opinion
- Social indicators
- Substantive applications of survey research

A complete overview of all sessions that are organized can be consulted on the following link: <http://www.europeansurveyresearch.org/conf2015/sessions.php>.



Understanding Society's Scientific Conference 2015

Location: Colchester Campus, University of Essex
Date: 21-23 July 2015

The scientific conference provides an international forum for the exchange of research based on longitudinal data; in particular using household panel studies. An important aim of the event is to bring together people from different disciplines and to share research covering a broad range of themes.

The organising committee would particularly welcome abstracts that focus on:

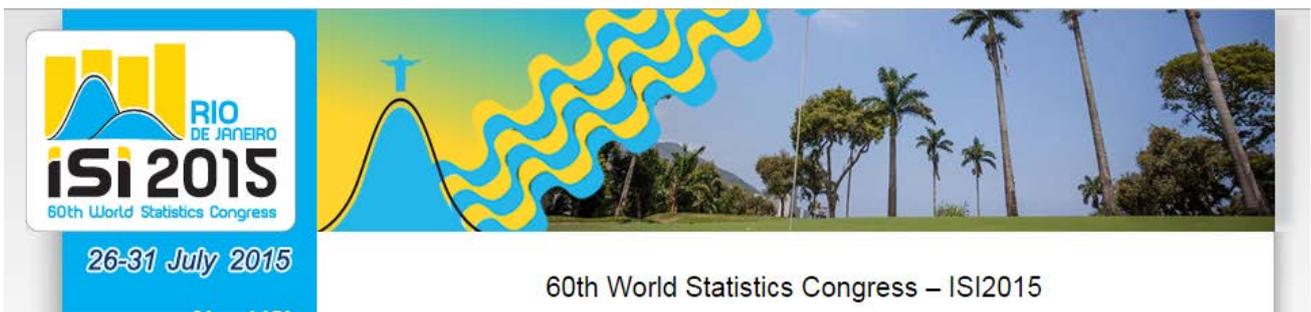
- Understanding Society and the British Household Panel Survey or other longitudinal studies;
- Health, education, employment, family and household, ethnicity, income, wealth, consumption and survey methodology.

Additionally they are keen to showcase comparative analyses of longitudinal data, innovative statistical methods and policy-relevant research

Further information

- [Event website](#)
- [Read about last year's conference](#)

For further information please email events@understandingsociety.ac.uk



60th ISI World Statistics Congress

Organized by: International Statistical Institute
Where: Rio de Janeiro, Brazil
When: 26.07.2015 to 31.07.2015
Homepage: <http://www.isi2015.org/>

We are delighted to invite you to the 60th ISI World Statistics Congress (WSC), which will take place in Rio de Janeiro, Brazil, during 26–31 July 2015.

The WSC is the flagship conference of the International Statistical Institute (ISI) and its seven associations. It is a biennial conference with a rich tradition, and IBGE is pleased to host and organize ISI2015 in Brazil.

The congress will bring together members of the statistical community to present, discuss, promote and disseminate research and best practice in every field of Statistics and its applications. The Scientific Programme of the ISI2015 will include a wealth of activities that will cover stimulating topics and will offer delegates innovative and well-balanced presentations, as well as plenty of opportunities for discussion and exchange.

A rich and exciting Social Programme is also being developed, with plenty to see and enjoy for participants and their accompanying persons, hoping to make your trip to Rio and taking part in ISI2015 a truly unforgettable experience.

The venue - Riocentro - is located in Barra da Tijuca, a district surrounded by natural beauty but also many sophisticated bars, restaurants and several malls and close to a variety of historical and cultural programs that only the Wonderful City can offer.

We are confident that all the ingredients are in place to ensure that the 60th ISI World Statistics Congress will be a memorable statistical event!

For further information please email Francisco Samaniego fjsamaniego@ucdavis.edu

First Latin American ISI Satellite Meeting on Small Area Estimation (SAE 2015)

Organized by: International Statistical Institute (ISI), the International Association of Survey Statisticians (IASS), the Sociedad Chilena de Estadística (SOCHE), the Instituto Nacional de Estadísticas (INE), the Ministerio de Desarrollo Social (MDS), and the Universidad Católica de Chile (Departamento de Estadística, Departamento de Salud Pública e Instituto de Sociología)

Where: Santiago, Chile

When: 03.08.2015 to 05.08.2015

Homepage: <http://www.encuestas.uc.cl/sae2015/index.html>

Welcome to the website of the First Latin American ISI Satellite Meeting on Small Area Estimation (SAE 2015), to be held at the Pontifical Catholic University of Chile on August 3-5 of 2015.

The SAE 2015 conference is co-sponsored by the International Statistical Institute (ISI), the International Association of Survey Statisticians (IASS), the Sociedad Chilena de Estadística (SOCHE), the Instituto Nacional de Estadísticas (INE), the Ministerio de Desarrollo Social (MDS), and the Universidad Católica de Chile (Departamento de Estadística, Departamento de Salud Pública e Instituto de Sociología).

The International Statistical Institute (ISI) is a non-profit, non-governmental organization and has consultative status by the Economic and Social Council of the United Nations since 1949. Officially, the ISI was established in 1885, though the initial international gatherings of statisticians started earlier - back in 1853. Therefore, we are one of the oldest scientific associations still active throughout the world today.

The International Association of Survey Statisticians (IASS – AISE) was founded in 1973 as a section of the International Statistical Institute. It aims to promote the study and development of the theory and practice of sample surveys and censuses. It also aims to increase the interest in surveys and censuses among statisticians, among governments and the public in the different countries of the world.

The Chilean Statistical Society (*Sociedad Chilena de Estadística*, SOCHE) was founded in 1967. Professional members of the society work in academia, however others work in the private sector and the government. The society also has Student members and Institutional members, usually from department of statistics/mathematics in Chilean universities.

The National Statistical Institute (*Instituto Nacional de Estadísticas*, INE) is one of the most long-standing institutions in Chile. Since its official creation in 1843, the institution has done several censuses, surveys and studies of the national reality. Today the INE provides to the country more than 70 indicators of great quality of a great diversity of subject including employment, prices, population, culture, public

safety, and economy, among other relevant aspects for decision-making in public policy.

The Ministry of Social Development (*Ministerio de Desarrollo Social*, MDS) is responsible for coordinating all Chilean social policies and articulating inter-ministerial initiatives. Its mission is to contribute to the design and implementation of policies, plans and programs on social development, especially those aimed at eradicating poverty and providing social protection to vulnerable persons or groups, promoting mobility and social integration. Likewise, the MDS must ensure coordination, consistency and coherence of policies, plans and programs for social development at national and regional levels. It also has to evaluate the pre-investment studies of investment projects seeking funding from the State to determine its social profitability, so that they answer to the strategies and policies for growth and economic and social development.

The Pontifical Catholic University (*Pontificia Universidad Católica*, UC) of Chile, established in 1888, is one of Chile's oldest universities and one of the most recognized educational institutions in Latin America. UC is a private, urban, multi-campus university. UC's 18 faculties are distributed through four campuses in Santiago and one regional campus located in southern Chile.

For more information, please visit the homepage or contact sae2015@uc.cl.



Conference Information Program EXPO 2015 Career Service Professional Development Sponsors Newsroom Home

SUBMIT ABSTRACTS



- invited posters
- topic-contributed papers and panels
- contributed papers and posters
- speed abstract

Joint Statistical Meetings

Organized by: American Statistical Association, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society, and International Statistical Institute

Where: Seattle, USA

When: 08.08.2015 to 13.08.2015

Homepage: <http://www.amstat.org/meetings/jsm/2015/index.cfm>

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the American Statistical Association, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society, and International Statistical Institute. Attended by more than 6,000 people, meeting activities include oral presentations, panel sessions, poster presentations, professional development courses, an exhibit hall, the Career Placement Service, society and section business meetings, committee meetings, social activities, and networking opportunities.

Seattle, Washington, the host city for JSM 2015, offers a wide range of options for sharing time with friends and colleagues or sightseeing with family. For information, contact meetings@amstat.org.

The 2015 Joint Statistical Meetings will be held August 8–13, at the Washington State Convention Center, 800 Convention Place, Seattle, WA 98101.

***The Fourth Baltic-Nordic Conference on Survey Statistics -
BaNoCoSS-2015 -
will be held on 24-28 August 2015 in Helsinki, Finland***



Fourth Baltic-Nordic Conference on Survey Statistics BaNoCoSS 2015

Organized by: University of Helsinki
Where: Helsinki, Finland
When: 24.08.2015 to 28.08.2015
Homepage: <https://wiki.helsinki.fi/display/banocoss2015/4th+Baltic-Nordic+Conference+on+Survey+Statistics>

The Fourth Baltic-Nordic Conference on Survey Statistics - BaNoCoSS-2015 - will be held on 24-28 August 2015 in Helsinki, Finland.

BaNoCoSS-2015 is a scientific conference presenting developments on theory, methodology and applications of survey statistics in a broad sense.

The conference provides a platform for discussion and exchange of ideas for a variety of people. These include, for example, statisticians, researchers and other experts of universities, national statistical institutes, research institutes and other governmental bodies, and private enterprises, dealing with survey research methodology, empirical research and statistics production. University students in statistics and related disciplines provide an important interest group of the conference.

BaNoCoSS-2015 is organized by the Baltic-Nordic-Ukrainian Network on Survey Statistics, University of Helsinki, Statistics Finland and The Finnish Statistical Society.

The Year 2015:

University of Helsinki celebrates its 375th anniversary.

Statistics Finland celebrates its 150th anniversary.

Former conferences on survey statistics organized by the BNU network include:

[BaNoCoSS-2011 in Norrfällsviken, Sweden](#)

[BaNoCoSS-2007 in Kuusamo, Finland](#)

BaNoCoSS-2002 in Ammarnäs, Sweden

Additional information on the activities of the Baltic-Nordic-Ukrainian Network on Survey Statistics is available at <http://wiki.helsinki.fi/display/BNU>.

email: banocoss-info@helsinki.fi

Welcome to the web pages of **EESW15**, the fourth European Establishment Statistics Workshop

Poznań, Poland, 7-9 September 2015

- hosted by [Poznań University of Economics](#) and [Statistical Office in Poznań](#) -



EESW15, The Fourth European Establishment Statistics Workshop

Where: Poznan, Poland
When: 07.09.2015 to 09.09. 2015
Homepage: <http://enbes.wikispaces.com/EESW15>

We invite you to a workshop aiming to provide a strong offering in latest results on methods and practices for producing business, economic and organisational statistics. The workshop gives the opportunity for official statistics methodologists, academic researchers and practitioners from the business sphere to interact with colleagues and exchange experiences on the topics of common interest, both through ample allocated discussion times and on the informal fringes of the workshop. The workshop covers best methodologies and practices for all stages of the statistical production process: specifying needs, designing sampling and estimation procedures, data collection instruments, building systems, collecting data, processing (including editing, imputation and estimation), analysis, dissemination, creating and using process data, quality measures, and so on.

We especially invite contributions on new developments such as:

- supra-national integration of business statistics
- improving consistency of business statistics over unit types
- design and maintenance of business registers from a statistical perspective
- business profiling: process and effects
- methodology for coordinated sampling frames
- estimation based on the combination of surveys and administrative data
- changes and trends in structure and organisation of businesses
- business statistics and national accounts.

Proposals addressing issues concerning statistical units are especially welcomed.

The workshop will be held at the Poznań University of Economics, Poland. Previous EESWs have been held in Stockholm in 2009, in Lausanne in 2011, and in Nuremberg in 2013.



2015 International Total Survey Error Conference

Where: Baltimore, USA
When: 19.09.2015 to 22.09.2015
Homepage: <https://www.tse15.org/ehome/90248/TSE15home/?&>

TSE15 is an international conference focused on survey quality and the challenges of big data. The Total Survey Error concept summarizes the ways a survey estimate may deviate from the corresponding value in the population.

The conference is for statisticians, survey managers and methodologists, pollsters, public opinion researchers, and marketing research professionals from around the world who are concerned about data quality and analytics. Presentations and short courses cover tools and approaches for understanding the sources of error and developing methods for reducing total error. We look forward to your participation at TSE15 in Baltimore.

For general info about TSE15, email info@tse15.org.



ICES V, The Fifth International Conference on Establishment Surveys

Organized by: Swiss Federal Statistical Office
Where: Geneva, Switzerland
When: 20.06.2016 to 23.06.2016
Homepage: <http://www.portal-stat.admin.ch/ices5/>

The Fifth International Conference on Establishment Surveys (ICES-V) will be held in Geneva, Switzerland, on June 20-23, 2016. Continuing in the traditions of ICES-I to ICES-IV, ICES-V intends to explore new areas of establishment statistics as well as to reflect state-of-the-art at the time of holding the conference.

Situated for the first time in Europe, in the beautiful surroundings of the canton and city of Geneva, ICES-V is expected to be attractive to professionals and researchers in the area of statistics on businesses, farms and institutions throughout the world. The conference is planned to include:

- Strong offering of short courses on different levels (introductory, intermediate, advanced)
- Introductory overview lectures on important and timely topics
- Selection of contributed papers and posters
- A keynote speaker and reception
- Poster sessions and software demonstrations

All Conference activities will be held at the International Conference Center, Geneva. The participants will have a large variety of choices of hotels of different categories in Geneva. The participants will be provided with a free Geneva Transport Card upon check-in to their hotel, valid for the entire duration of their stay.

This site will be updated with new information as we progress in our steps towards the conference, so please do visit it occasionally. Alternatively, send an email to ices-v@bfs.admin.ch with the subject line "Please add to ICES list" to be kept abreast per email of events related to the ICES conference series.

The preceding conferences in the ICES series have regularly attracted around 350-400 participants, and we trust that ICES-V – placed in Europe and close to Africa and Asia – will through the participation reflect the relevance of business, agricultural, economic and other establishment statistics in these geographic regions.

Dear visitor, we do hope that based on this information the conference will be of sufficient interest for you to contribute to it as an author, session organizer, sponsors, exhibitor or participant!

For more information, please visit the homepage or contact ices-v@bfs.admin.ch



3MC International Conference 2016

Events

[▶ 2016 3MC Conference](#)

Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC)

Organized by: CSDI (Comparative Survey Design and Implementation)
Where: Chicago, Illinois
When: July 2016
Homepage: <http://csdiworkshop.org/v2/index.php/3mc-2016>

As part of an ongoing effort to promote quality in multi-population surveys and to raise the level of methodological expertise in various applied fields of comparative survey research, the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts will be held July 2016 in Chicago (3MC 2016).

This conference will bring together researchers and survey practitioners concerned with survey methodology and practice in comparative contexts. It will provide a unique opportunity to discuss and present research that contributes to our understanding of survey needs and methods in multi-cultural, multi-national, and multi-lingual contexts. Conference contributions will help document current best practices and stimulate new ideas for further research and development.

For more information please visit <http://csdiworkshop.org/v2/index.php/3mc-2016>



In Other Journals

Journal of Survey Statistics and Methodology

VOLUME 2 / NUMBER 4 / DECEMBER 2014

<http://jssam.oxfordjournals.org/content/current>

Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation

C.-E. Särndal and P. Lundquist

Allocation for Dual Frame Telephone Surveys with Nonresponse

S. L. Lohr and J. M. Brick

Level-of-Effort Paradata and Nonresponse Adjustment Models for a National Face-to-Face Survey

J. Wagner, R. Valliant, F. Hubbard, and L. Jiang

Interviewer Effects on Latent Constructs in Survey Research

K. Beullens and G. Loosveldt

Comparison of Three Modes for a Crime Victimization Survey

S. Laaksonen and M. Heiskanen

Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK

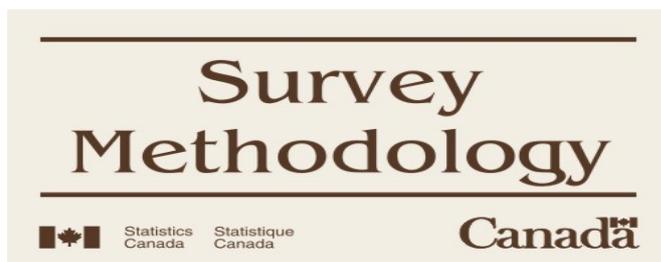
M. Brown and L. Calderwood

Mobile Web Survey Design: Scrolling versus Paging, SMS versus E-mail Invitations

A. Mavletova and M. P. Couper

What Are You Doing Now? Activity-Level Responses and Recall Failures in the American Time Use Survey

T. A. Baghal, R. F. Belli, A. L. Phillips, and N. Ruther



December 2014, VOL 40, NO 2

<http://www.statcan.gc.ca/pub/12-001-x/12-001-x2014002-eng.htm>

From multiple modes for surveys to multiple data sources for estimates

C. F. Citro

Frequentist and Bayesian approaches for comparing interviewer variance components in two groups of survey interviewers

B. T. West and M. R. Elliott

Bagging non-differentiable estimators in complex surveys

J. C. Wang, J. D. Opsomer and H. Wang

Fractional hot deck imputation for robust inference under item nonresponse in survey sampling

J. Kwang Kim and S. Yang

Potential gains from using unit level cost information in a model-assisted framework

D. G. Steel and R. r. Clark

Optimal solutions in controlled selection problems with two-way stratification

S. W. Kim, S. G. Heeringa and P. W. Solenberger

On aligned composite estimates from overlapping samples for growth rates and totals

P.I Knottnerus

The estimation of gross flows in complex surveys with random nonresponse

A. Gutiérrez, L. Trujillo and P. L.D. N. Silva

Chi-squared tests in dual frame surveys

Y. Lu

Estimation methods on multiple sampling frames in two-stage sample designs

Guillaume Chauvet and Guylène Tandeau de Marsac

Combining information from multiple complex surveys

Q. Dong, M. R. Elliott and T. E. Raghunathan

December 2014, Vol. 30 No. 4

<http://www.jos.nu/entry.asp>

Preface – Introduction to the Special Issue on Establishment Surveys

In Search of Motivation for the Business Survey Response Task

V. Torres van Grinsven, I. Bolko, M. Bavdaž

An Adaptive Data Collection Procedure for Call Prioritization

J-F Beaumont, C. Bocci, D. Haziza

Measuring Representativeness of Short-Term Business Statistics

P. Ouwehand, B. Schouten

Does the Length of Fielding Period Matter? Examining Response Scores of Early Versus Late Responders

R. Sigman, T. Lewis, N.D. Yount, K. Lee

The Utility of Nonparametric Transformations for Imputation of Survey Data

M. W. Robbins

Modeling Nonresponse in Establishment Surveys: Using an Ensemble Tree Model to Create Nonresponse Propensity Scores and Detect Potential Bias in an Agricultural Survey

M. Earp, M. Mitchell, J. McCarthy, F. Kreuter

Detecting and Treating Verified Influential Values in a Monthly Retail Trade Survey

M.H. Mulry, B.E. Oliver, S.J. Kaputa

The Impact of Sampling Designs on Small Area Estimates for Business Data

J.P. Burgard, R. Münnich, T. Zimmermann

On Precision in Estimates of Change over Time where Samples are Positively Coordinated by Permanent Random Numbers

A. Lindblom

Analytic Tools for Evaluating Variability of Standard Errors in Large-Scale Establishment Surveys

M.J. Cho, J.L. Eltinge, J. Gershunskaya, Larry Huff

Estimation of Mean Squared Error of X-11-ARIMA and Other Estimators of Time Series Components

D. Pfeffermann, M. Sverchkov

Data Smearing: An Approach to Disclosure Limitation for Tabular Data

D. Toth



VOL 7, NO 5 (2014)

<http://www.surveypractice.org/index.php/SurveyPractice/issue/view/59>

Can Interviewer Behaviors During ACASI Affect Data Quality

B.T. West, E. Peytcheva

Open-Ended Survey Questions: Item Nonresponse Nightmare or Qualitative Data Dream?

A. L. Miller, A.D. Lambert

"Here comes the sun" : Evidence of the Effect of Sun on Compliance to a Survey Request

N. Guéguen, C. Jacob

The Questionnaire Looks Self Explanatory, so to Save Time I'll Just Skip the Skip Instructions

M. Roszkowski, M.I Glatzer, M. Glatzer, M. Soven



Vol. 8, No. 3 (2014)

<https://ojs.ub.uni-konstanz.de/srm/>

The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement

J. Smyth, K. Olson, A. Kasabian

The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency

F. Wolter, B. Laier

Is Vague Valid? The Comparative Predictive Validity of Vague Quantifiers and Numeric Response Options

T.A. Baghal

Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective

Y. Pan, M. Fond

Evaluation of weighting methods to integrate a top-up sample with an ongoing longitudinal sample

N. Watson



Statistical Journal of the IAOS: Journal of the International Association for Official Statistics

VOL 30, NO 4 (2014)

<http://www.iospress.nl/journal/statistical-journal-of-the-iaos/>

Using design principles to deliver strategy in a new organisational structure

C. A. Davies and G. Brown

Interview with Jay Siegel

Kirsten K. West

Median-based measures of inequality: Reassessing the increase in income inequality in the U.S. and Sweden

J. L. Gastwirth

Sustaining improved estimates of poverty in Armenia: An application of multivariate raking ratio adjustment

A. Arushanyan, K. Kouyumjyan, D. Martirosova, A. Mushtaq and F. Scheuren

A nationally representative economic survey five months after the Haitian earthquake: Radical changes in household members and gender discrepancy in employment retention

R. S. Kim, J. D. Ashley and M. E. Corcoran

Comment on: A nationally representative economic survey five months after the Haitian earthquake: Radical changes in household members and gender discrepancy in employment retention by R.S. Kim, J.D. Ashley and M.E. Corcoran

N. S. Gunaratna

Towards an international framework for statistical councils of national statistical systems

R. Alldritt

Laying the foundations for a new approach to Census taking in Ireland

J. Dunne and S. MacFeely

Comments on: Laying the foundations for a new approach to Census taking in Ireland by John Dunne and Steve MacFeely

A. O'Hara

Introduction to the special section "Experiences when building the methodological architecture for statistics production"

E. Elvers

Methodology architecture: A roadmap for new methodological directions in the Australian Bureau of Statistics

S.-M. Tam

Strengthening methodological architecture with multiple frames and data sources

J. M. Harris and C. Z.F. Clark

Methodology and technology directions at Statistics New Zealand

A. J. Seyb

Discussion: Methodology Architecture – an area under construction

M. Karlberg and W. Radermacher

Ethnicity, race and Māori life expectancy in Aotearoa New Zealand

J. Waldon and K. Dunstan

The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics

B.F.M. Bakker, J. van Rooijen and L. van Toor

Data quality evaluation in statistical data processing

D. Škiltere and S. Jesiljevska

Redesign in a statistical office – sociodemographic statistics in the Office for National Statistics

P.A. Smith



VOL 82, ISSUE 3 (DECEMBER 2014)

<http://onlinelibrary.wiley.com/doi/10.1111/insr.v82.3/issuetoc>

Fifty Years of Classification and Regression Trees

W.-Y. Loh

Disclosure-protected Inference Using Generalised Linear Models

J. O. Chipperfield and C. M. O'Keefe

Combining Probability Distributions by Multiplication in Metrology: A Viable Methods?

D. Grientschnig and I. Lira

The Deviance Information Criterion in Comparison of Normal Mixing Models

T. Fung, J. J. J. Wang and E. Seneta

Akeike's Information Criterion, C_p and Estimators of Loss for Elliptically Symmetric Distributions

A. Boisbunon, S. Canu, D. Fourdrinier, W. Strawderman and M. T. Wells

Local Powers of Optimal One-sample and Multi-sample Tests for the Concentration of Fisher-von Mises-Langevin Distributions

C. Ley and T. Verdebout

Small-scale Inference: Empirical Bayes and Confidence Methods for as Few as a Single Comparison

D. R. Bickel



Volume 6, Issue 1 (2014)

<http://repository.cmu.edu/jpc/>

On the 'Semantics' of Differential Privacy: A Bayesian Formulation

S. P. Kasiviswanathan and A. Smith

Bayesian Estimation of Disclosure Risks for Multiply Imputed, Synthetic Data

J. P. Reiter, Q. Wang, and B. Zhang

An Evaluation Framework for Privacy-Preserving Record Linkage

D. Vatsalan, P. Christen, C. M. O'Keefe, and V. S. Verykios

Noise Multiplication for Statistical Disclosure Control of Extreme Values in Log-normal Regression Samples

M. Klein, T. Mathew, and B. Sinha



Volume 7, Issue 3 (2014)

<http://www.tdp.cat/issues11/vol07n03.php>

DPCube: Differentially Private Histogram Release through Multidimensional Partitioning

Y. Xiao, L. Xiong, L. Fan, S. Goryczka, H. Li

Censors for Boolean Description Logic

T. Studer, J. Werner

PeGS: Perturbed Gibbs Samplers that Generate Privacy-Compliant Synthetic Data

Y. Park, J. Ghosh

Truly Anonymous Paper Submission and Review Scheme

C.-I Fan, M.-T. Chen, Y.-K. Liang, L.-S. Chen

Attribute Based Group Key Management

M. Nabeel, E. Bertino

A Systematic Comparison and Evaluation of k-Anonymization Algorithms for Practitioners

V. Ayala-Rivera, P. McDonagh, T. Cerqueus, L. Murphy



January 2015, Vol 178, Issue 1

<http://onlinelibrary.wiley.com/doi/10.1111/rssa.2014.178.issue-1/issuetoc>

Hierarchical models for estimating state and demographic trends in US death penalty public opinion

K. E. Shirley and A. Gelman

Inference and forecasting in the age–period–cohort model with unknown exposure with an application to mesothelioma mortality

M. D. M. Miranda, B. Nielsen and J. P. Nielsen

Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials

C. Forni, M. Marcellino and C. Schumacher

Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach

R. Vassallo, G. B. Durrant, P. W. F. Smith and H. Goldstein

Broader health coverage is good for the nation's health: evidence from country level panel data

R. Moreno-Serra and P. C. Smith

Local cost surface models of distance decay for the analysis of gridded population data

C. D. Lloyd

Diffusion of innovations in dynamic networks

C. C. Greenan

A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League

S. J. Koopman and R. Lit

Fitting age-specific fertility rates by a flexible generalized skew normal probability density function

S. Mazzuco and B. Scarpa

Modelling reporting delays for outbreak detection in infectious disease data

A. Noufaily, Y. Ghebremichael-Weldeselassie, D. Gragn Enki, P. Garthwaite, N. Andrews, A. Charlett and P. Farrington

Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout

L. Keele, R. Titiunik and J. R. Zubizarreta

Diagnostics for respondent-driven sampling

K. J. Gile, L. G. Johnston and M. J. Salganik

Bayesian inference for transportation origin–destination matrices: the Poisson–inverse Gaussian and other Poisson mixtures

K. Perrakis, D. Karlis, M. Cools and D. Janssens



**Journal of the American
Statistical Association**



Volume 109, Issue 507 (2014)

<http://amstat.tandfonline.com/toc/uasa20/109/507>

A Bayesian Nonparametric Modelling Framework for Development Toxicity Studies

K. Fronczyk & A. Kottas

Mechanistic Hierarchical Gaussian Processes

M. W. Wheeler, D. B. Dunson, S. P. Pandalai, B. A. Baker & A. H. Herring

Identifying Genetic Variants for Addiction via Propensity Score Adjusted Generalized Kendall's Tau

Y. Jiang, N. Li & H. Zhang

Estimating Ocean Circulation: An MCMC Approach With Approximated Likelihoods via the Bernoulli Factory

R. Herbei & L. M. Berliner

Multivariate Analysis of Longitudinal Ordinal Data With Mixed Effects Models, With Application to Clinical Outcomes in Osteoarthritis

C. M. Laffont, M. Vandemeulebroecke & D. Concordet

Causal Inference for fMRI Time Series Data With Systematic Errors of Measurement in a Balanced On/Off Study of Social Evaluative Threat
M. E. Sobel & M. A. Lindquist

Bayesian Generalized Low Rank Regression Models for Neuroimaging Phenotypes and Genetic Markers
H. Zhu, Z. Khondker, Z. Lu & J. G. Ibrahim

Estimation and Accuracy After Model Selection
B. Efron

Bayesian Aggregation of Order-Based Rank Data
K. Deng, S. Han, K. J. Li & J. S. Liu

Inference From Intrinsic Bayes' Procedures Under Model Selection and Uncertainty
A. J. Womack, L. León-Novelo & G. Casella

Adaptive Confidence Bands for Nonparametric Regression Functions
T. T. Cai, M. Low & Z. Ma

Efficient R-Estimation of Principal and Common Principal Components
M. Hallin, D. Paindaveine & T. Verdebout

Spatially Varying Coefficient Model for Neuroimaging Data With Jump Discontinuities
H. Zhu, J. Fan & L. Kong

Testing Second-Order Dynamics for Autoregressive Processes in Presence of Time-Varying Variance
V. Patilea & H. Raïssi

Filtering With Heavy Tails
A. Harvey & A.a Luati

Generalized Gaussian Process Regression Model for Non-Gaussian Functional Data
B. Wang & J. Qing Shi

Space-Filling Fractional Factorial Designs
Y.-D. Zhou & H. Xu

Weighted M -statistics With Superior Design Sensitivity in Matched Observational Studies With Multiple Controls
P. R. Rosenbaum

Multiply Robust Estimation in Regression Analysis With Missing Data
P. Han

Targeted Local Support Vector Machine for Age-Dependent Classification
T. Chen, Y. Wang, H. Chen, K. Marder & D. Zeng

On an Additive Semigraphoid Model for Statistical Networks With Application to Pathway Analysis
B. Li, H. Chun & H. Zhao

**Functional Principal Component Analysis of Spatiotemporal Point Processes
With Applications in Disease Surveillance**

Y. Li & Y. Guan

**Optimal Tests of Treatment Effects for the Overall Population and Two
Subpopulations in Randomized Trials, Using Sparse Linear Programming**

M.I Rosenblum, H. Liu & E.-H. Yen

**Sequential Lasso Cum EBIC for Feature Selection With Ultra-High Dimensional
Feature Space**

S. Luo & Z. Chen

**Segmented Model Selection in Quantile Regression Using the Minimum
Description Length Principle**

A. Aue, R. C. Y. Cheung, T. C. M. Lee & M. Zhong

The Sparse MLE for Ultrahigh-Dimensional Feature Screening

C. Xu & J. Chen

**Nonparametric Independence Screening in Sparse Ultra-High-Dimensional
Varying Coefficient Models**

J. Fan, Y. Ma & W. Dai

Interaction Screening for Ultrahigh-Dimensional Data

N. Hao & H. H. Zhang

**Martingale Difference Correlation and Its Use in High-Dimensional Variable
Screening**

X. Shao & J. Zhang

Some Comments on Copula-Based Regression

H. Dette, R. Van Hecke & S. Volgushev

BIOMETRIKA

Volume 101, issue 4 (2014)

<http://biomet.oxfordjournals.org/content/current>

Classification with confidence

J. Lei

When does more regularization imply fewer degrees of freedom? Sufficient conditions and counterexamples

S. Kaufman and S. Rosset

Variable selection in regression with compositional covariates

W. Lin, P. Shi, R. Feng, and H. Li

Censored rank independence screening for high-dimensional survival data

R. Song, W. Lu, S. Ma, and X. J. Jeng

Transformed sufficient dimension reduction

T. Wang, X. Guo, L. Zhu, and P. Xu

Interactive model building for Q-learning

E. B. Laber, K. A. Linn, and L. A. Stefanski

Estimation of a semiparametric natural direct effect model incorporating baseline covariates

E.J. Tchetgen and I. Shpitser

Robust Estimators for nondecomposable elliptical graphical models

D. Vogel and D.E. Tyler

Nonparametric Bayes dynamic modelling of relational data

D. Durante and D.B. Dunson

A class of improved hybrid Hochberg-Hommel type step-up multiple test procedure

J. Gou, A.C. Tamhane, D. Xi and D. Rom

A distribution-free two-sample run test applicable to high-dimensional data

M. Biswas, M. Mukhopadhyay, A.K. Ghosh

Testing independence and goodness-of-fit in linear models

A. Sen and B. Sen

Circular designs balanced for neighbours at distances one and two

R.E.L. Aldred, R.A. Bailey, B. D. McKay and I.M. Wanless

Nearly orthogonal arrays mappable into fully orthogonal arrays

R. Mukerjee, F. Sun and B. Tang

Analytical p-value calculation for the higher criticism test in finite-d problems

I.J. Barnett and X. Lin

Generalized cornfield conditions for the risk difference

P. Ding and T.J. Wandersweele

Tests for Kronecker envelope models in multilinear principle components analysis

J.R. Schott

Semiparametric maximum likelihood inference by using failed contact attempts to adjust for nonignorable nonresponse

J. Qin and D.A. Follmann

Robust Bayesian variable selection in linear models with spherically symmetric errors

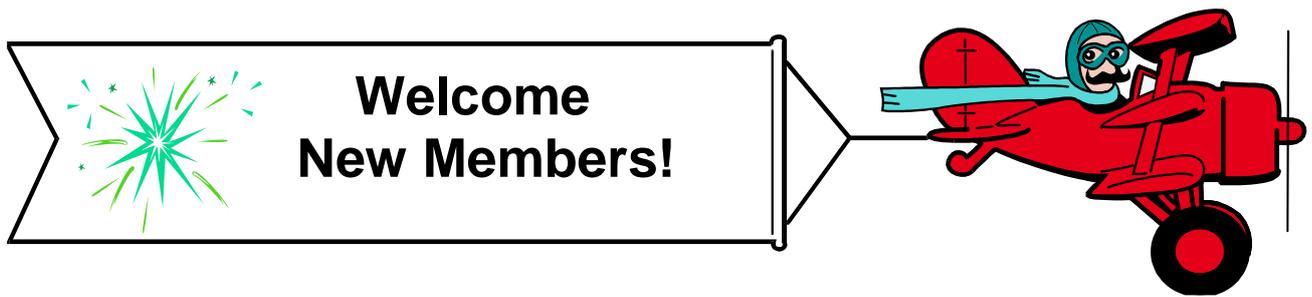
Y. Maruyama and W.E. Strawdermann

General type-token distribution

S. Hidaka

On exact forms of Taylor's theorem for vector-valued functions

C.Feng, H. Wang, T. Chen and X.N. Tu



We are very pleased to welcome the following new members!

Member	Country
Dr. Adedayo Amos Adewara	Nigeria
Dr. Zahoor Ahmad	Pakistan
Mr. Rizwan Bashir	United States
Dr. Dragos Cimpoies	Moldova, Republic of
Dr. Ton De Waal	Netherlands
Ms. Carolina Franco	United States
Dr. Gutiérrez Rojas Hugo Andres	Colombia
Ms. Maiki Ilves	Estonia
Mr. Karol Krotki	United States
Mr. Gareth Minshall	New Zealand
Ms. Isabel Molina Peralta	Spain
Mr. Letsema G. Motsemme	Botswana
Mrs. Melike Oguz Alper	United Kingdom
Mr. Diego Andres Perez Ruiz	United Kingdom
Mr. Bryan Sayer	United States
Ms. Tersa Schmidt	South Africa
Dr. Iwan Aschwin Sno	Suriname
Mrs. Eva Tsouparopoulou	Greece
Dr. James Wagner	United States
Dr. Emilio Lopez Excobar	Mexico

IASS Officers and Council Members

Executive Officers

President:	Danny Pfeffermann (Israel)	msdanny@huji.ac.il
President-elect:	Steve Heeringa (USA)	sheering@isr.umich.edu
Vice-Presidents:	Jairo Mabil Oka Arrow (South Africa) Geoffrey Lee (Australia)	jairo.arrow@gmail.com geoff.lee99@bigpond.com
Scientific Secretary:	Mick Couper (USA)	mcouper@umich.edu
Council Members (2011-2015):	Christine Bycroft (New Zealand) Ka-Lin Karen Chan (China) Olivier Dupriez (Belgium/USA) Natalie Shlomo (UK) Marcel de Toledo Vieira (Brazil) Alvaro Gonzalez Villalobos (Argentina)	christine.bycroft@stats.govt.nz klchan@censtatd.gov.hk odupriez@worldbank.org natalie.shlomo@manchester.ac.uk marcel.vieira@ufff.edu.br alvarun@gmail.com
Council Members: (2013-2017)	J. Michael Brick (USA) Daniela Cocchi (Italy) Jack Gambino (Canada) Risto Lehtonen (Finland) Ralf Münnich (Germany) Jean Opsomer (USA)	mikebrick@westat.com daniela.cocchi@unibo.it gambino@statcan.ca risto.lehtonen@helsinki.fi muennich@uni-trier.de jopsomer@stat.colostate.edu

Committee Chairs

Chair of the Rio 2015 Programme Committee	Christine Bycroft (New Zealand)	christine.bycroft@stats.govt.nz
Chair of the committee for the Cochran Hansen prize award	Risto Lehtonen (Finland)	risto.lehtonen@helsinki.fi

Ex Officio Members

ISI Operation Manager/Assistant Director, supporting the IASS Transition Executive Director:	Shabani Mehta	s.mehta@cbs.nl
Treasurer:	Catherine Meunier (France)	katherine.meunier@orange.fr
Finance:	Ada van Krimpen (The Netherlands)	an.vankrimpen@cbs.nl
Webmaster:	Mehmood Asghar and Olivier Dupriez	odupriez@worldbank.org
IASS Secretariat Membership Officer	Margaret de Rooter-Molloy (The Netherlands)	m.deruitermolloy@cbs.nl

2 International Organisations

AFRISTAT
EUROSTAT

15 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA (IBGE)
CANADA – STATISTICS CANADA
DENMARK – DANMARKS STATISTIK
FINLAND – STATISTICS FINLAND
GERMANY – STATISTISCHE BUNDESAMT
ITALY – INSTITUTO NAZIONALE DI STATISTICA (ISTAT)
KOREA, REPUBLIC OF – STATISTICS KOREA
LUXEMBOURG – EUROPEAN COMMISSION – EUROSTAT
MACAO – DIRECÇÃO DOS SERVIÇOS DE ESTATÍSTICA E CENSOS
MALI – AFRISTAT
MEXICO – INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFIA (INEGI)
MEXICO – NUMERIKA-MEDICION Y ANALISIS ESTAD. AVANZADO, SC
MAURITIUS – STATISTICS MAURITIUS
NEW ZEALAND – STATISTICS NEW ZEALAND
NORWAY – STATISTICS NORWAY
PORTUGAL – INSTITUTO NACIONAL DE ESTADÍSTICA (INE)
SWEDEN – STATISTICS SWEDEN
UNITED STATES – NASS, RESEARCH & DEVELOPMENT DIVISION
UNITED STATES – NATIONAL CENTER FOR HEALTH STATISTICS
UNITED STATES – WESTAT INC.
UNITED STATES – RESEARCH TRIANGLE INSTITUTE
UNITED STATES – SURVEY RESEARCH CENTER

5 Universities, Research Centers, Private Statistics Firms

USA – CENTERS FOR DISEASE CONTROL AND PREVENTION
USA – RESEARCH TRIANGLE INSTITUTE
USA – SURVEY RESEARCH CENTER, UNIVERSITY OF MICHIGAN
USA – U.S. DEPARTMENT OF AGRICULTURE
USA – WESTAT

INTERNATIONAL ASSOCIATION
OF SURVEY STATISTICIANS

CHANGE OF ADDRESS FORM



If your home or business address has changed, please copy, complete, and mail this form to:

IASS Secretariat Membership Officer
Margaret de Rooter-Molloy
International Statistical Institute
P.O. Box 24070, 2490 AB The Hague,
The Netherlands

Name: Mr./Mrs./Miss/Ms. _____ First name: _____

E-mail address (please just indicate one): _____

May we list your e-mail address on the IASS web site?

Yes No

Home address

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number: _____

Fax number: _____

Business address

Company: _____

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number and extension: _____

Fax number: _____

Please specify address to which your IASS correspondence should be sent:

Home Business

Read
The Survey Statistician
online!



<http://isi-iass.org/home/services/the-survey-statistician/>