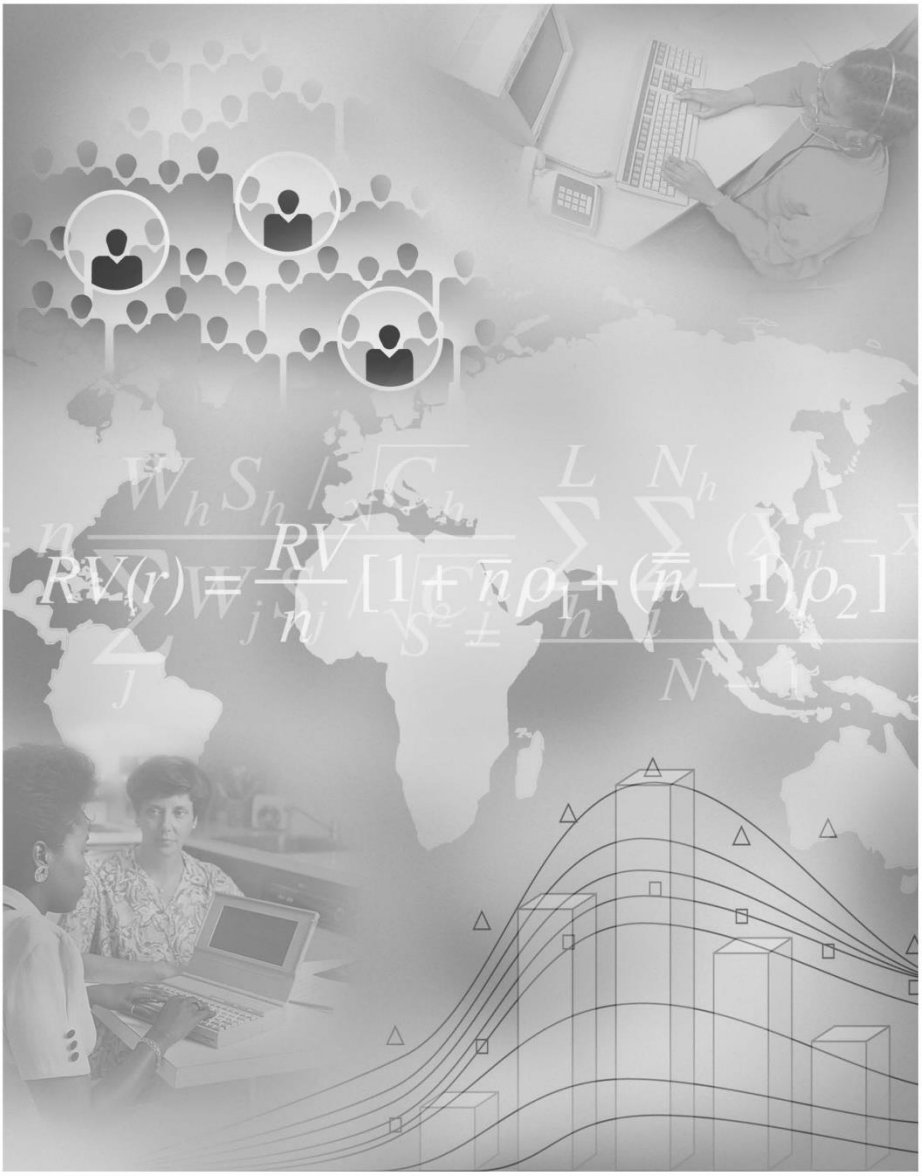
the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 69 **January 2014**



INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

International Statistical Institute

INTERNATIONALE ASSOCIATION D'ENQUETES DES STATISTICIENS

Institut International de Statistique

the Survey Statistician
The Newsletter of the International Association of Survey Statisticians
No. 69 January 2014

International Statistical Institute
Institut International de Statistique

Editors
Natalie Shlomo and Frank Yu

Section Editors
Pierre Lavallée — Country Reports
Robert Clark — Ask the Experts
Mick Couper — New and Emerging Methods
Natalie Shlomo — Book and Software Review

Circulation/Production
Henry Chiem
Yovina Joymungul Poorun
Courtney Williamson
Justin van der Veeke

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 frank.yu@abs.gov.au or mailed to:
Frank Yu
Australian Bureau of Statistics
Locked Bag 10
Belconnen ACT 2616
Australia

In This Issue

3 **Letter from the President and Appendix**

9 **Letter from the Editors**

11 **Report from the Scientific Secretary**

12 **News and Announcements**

17 **Ask the Experts:**
When should a multiple frame survey be used?
By Sharon L. Lohr

22 **New and Emerging Methods:**
Big Data as a Source of Statistical Information
By Piet J.H. Daas and Marco J.H. Puts

32 **Book and Software Review:**
Nonresponse in Social Science Surveys: A Research Agenda,
By R. Tourangeau and T. J. Plewes,
Reviewed by Peter Lynn

34 **Country Reports**
- Argentina
- Australia
- Canada
- India
- New Zealand

41 **Upcoming Conferences and Workshops**

47 **In Other Journals**

59 **Welcome New Members**

60 **IASS Officers and Council Members**

61 **Institutional Members**

62 **Change of Address Form**

Letter from the President

Dear Friends and Colleagues,

It is my pleasure to write to you as the new president of our esteemed and important association, the IASS, which represents survey statisticians from across the globe. Our profession has application and implications for a wide range of fields, and the professional work which we carry out and the collaboration between us by means of our 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 analyzing big data sets with numerous unknown parameters, employing advanced computing algorithms, what is referred to these days as "data science". But often there is little control on how these data are collected and which population they actually represent, and so the need for well-designed probability samples with careful attention to questionnaire design, mode of response and corresponding nonresponse, and consequently the use of proper estimation and analysis techniques will always remain. Sit back and relax, our profession is secured for many years to come but of course, there are many new intriguing challenges facing us in all aspects of our work, whether in theoretical research or in application. It is here where our society can help very significantly and I look forward to working with you to achieve these goals.

I shall return to these introductory remarks below, but first I would like to express in the name of all of you my profound thanks to my predecessor, Professor Ray Chambers, for his masterful leadership of the IASS for the last two years. During his term as President, Ray led very skillfully the transition of the IASS from a French-registered association to the new Dutch version of the IASS. It was only in Hong Kong that I became fully aware of all the big issues and complications involved in this transition and I can only hope that no similar transition will be required during my two years of presidency. So thank you Ray and even though you are no longer part of the formal executive of the IASS, I know that I can count on you to help me with good advice and help whenever I shall need it. Ray will tell you that he wouldn't have succeeded in this almost impossible transition mission without the big help of our outgoing Executive Director, Ms. Catherine Meunier, and we are all grateful to her as well for all that she has done for the IASS. As of May of 2014 and under the new version of the IASS, we shall no longer have an Executive Director. Instead, Ms. Shabani Mehta from Statistics Netherlands (s.mehta@cbs.nl) will serve as the IASS liaison at the ISI and I can testify already that she is a great asset for us. Related to the aforementioned transition of the IASS, I would like to express our gratitude to INSEE in France for hosting the IASS for so many years, always supporting our goals and providing all our needs, and to Survey Methodology for providing the French translation of our official publications. Unfortunately, French is no longer an official language of the IASS but we do hope to have at least some of our new website content translated into several languages (see below). Finally, I wish to thank the outgoing council members for all their contributions to the development of our association and of course, welcome the new members. I very much look forward to working with you and the continuing members in advancing the goals of our association. Details of the continuing and new Executive and council members appeared in the previous edition of The Survey Statistician.

Next, I would like to outline my priorities for the IASS activities until the end of my term in 2015. As first priority I think that we need to at least double our membership. The IASS has now some 385 individual members and 23 institutional members. The numbers have declined steadily over the years and what is even more worrying is that our membership is dominated by elderly

people (myself included), with almost no young survey statisticians, which our future as a society depends on. The young survey statisticians are extremely bright and knowledgeable; with many new innovative ideas, but for some reason they are not registered and are not active in the society. We need a much larger membership if we truly want to achieve the goals set up in my introductory comments. The fruits of our work pervade so many fields of endeavor and we should have a broad representation from all areas of research and applications. A large membership will allow us greater latitudes in our joint collaboration efforts and will increase dramatically our budget. This, in turn, will facilitate more and varied activity like sponsoring international and regional conferences, running short courses and workshops, and providing consultation in developing countries. With a larger membership, we may also increase the number of invited paper sessions in future ISI congresses. In an appendix to this letter, I propose some ideas of how we can possibly achieve the goal of doubling our membership. The most important component of this campaign (and probably the most effective one) is a "Member-brings- a Member" plan. No doubt, each one of us has a colleague, a student or another contact for whom it would be appropriate to become an IASS member. I am asking you to speak to those potential members about the benefits of being an IASS member and encourage them to sign up. The membership fees are extremely low: thirty (30) Euros for members from developed countries (2.5 Euros per month!!!), and fifteen (15) Euros for members from developing countries. If each one of us brings in only one new member (and many of us can bring in many more new members), we already achieve our goal. Please read the appendix and join the campaign. If you have other ideas of how we can increase personal and institutional membership, write to me to msdanny@cbs.gov.il and I shall share them with all of us.

A second priority of high importance is to participate in the ISI Statistical Capacity Building effort in developing countries. Our contribution to this effort can be by teaching sample survey statistics and by consulting. In fact, I proposed that the IASS gets involved in this type of activity more than a year ago as part of an effort to increase our institutional membership. The idea is that the IASS will pay the cost of developing appropriate courses and the travel expenses of the instructors, while the local institutions will cover the local costs and in return for the courses or consulting mission, they will become institutional members of the IASS. I am very happy to tell you that our council member, Dr. Olivier Dupriez has already secured money from the World Bank to fund travel expenses and per diems for several such projects. Now it all depends on you. We need volunteers to travel for a week or two and teach a course or provide consulting. Your service will be voluntary but as I just mentioned, all your costs will be covered. Several people already volunteered when I first proposed this idea. I am going to write to many heads of statistical bureaus in developing countries to propose this "deal", asking that they define what they are mostly interested in. I shall soon set up a committee to overlook this important operation. In the meantime, you can write to me with related ideas and most important, in order to volunteer.

A third priority is to boost the teaching of survey sampling in statistics departments at universities and colleges around the world. Unlike in the previous century, survey sampling is now hardly taught as a formal compulsory or even optional course and other, perhaps more attractive courses get the priority. But as I said in my introductory comments, sample surveys will continue to constitute a major component of statistical practice in every country and it is critical to train new generations of survey statisticians. I plan to write to as many as possible heads of statistics department around the world and encourage them to include sample surveys in their regular curriculum, offering them our help, but I rely on you to help me in this campaign, wherever you are and whatever your affiliation is.

Finally, but not less important, is the redesign of the IASS website (http://isi-iass.org/home/). If I am not mistaken, this project was initialized by Ray, and our new webmaster is no other than Olivier from the World Bank (odupriez@worldbank.org), whose contributions to our association I have already mentioned. I propose that you write directly to Olivier if you have any suggestions regarding the new site. One idea of Olivier, which I have already mentioned is to try and translate the important contents of the website to several languages, including French, Spanish and Chinese, thus making the site transparent to non-English speaking members. In the new website you will find also registration forms to the IASS, which you can use for recruiting new members. Needless to say that the Website is the prime means of communication between us and new members, and we look forward to it being more efficient, more attractive and more user-friendly than our present site. Hopefully, the new website will be versatile enough to allow future developments such as on-line discussion groups, "Ask the experts", Calendar of events, job advertisements and on-line seminars. Some of these options appear already in the current site, but they haven't been activated. Another idea proposed by Olivier is to have a "directory of members", which will contain a searchable directory of members with their email address, accessible only to IASS members.

As you can see, we have many ambitious plans for the IASS for the next two years, but we can only implement them with your help. This is our association, and I expect each one of you to be active in making the IASS a lively organization.
Let me finish this long letter with some important happy news.

1. As many of you are aware of, our program committee, headed by Ms. Christine Bycroft, has finalized a list of proposed "invited papers sessions" (IPS) for the next ISI congress in Rio De Janeiro and I take this opportunity to thank Christine and her colleagues for their hard work. The list has been sent to the central ISI program committee and we are hopeful that all our proposals will be accepted. We plan to have also a special tutorial on a topic of general interest. If you have a suggestion for a topic or instructor, let me know.

2. In Rio de Janeiro we shall have for the first time two special fascinating sessions, with no parallel IPS sessions. The first session is an "IASS President's Invited Speaker Session" and I am happy to tell you that Professors Jon Rao and Wayne Fuller accepted my invitation to present a joint paper on "Sample Surveys, Past, Present and Future Directions" (title is tentative). The session will last 70 minutes. The second special session is a "Journal Papers Session" and I have invited the editors of the Journal of Survey Statistics and Methodology (JSSM) and Survey Methodology to select papers for this session. This session will last 100 minutes, 30 minutes for each paper, 15 minutes for a discussant (one for each paper) and 10 minutes for floor discussion. I am grateful to Professor Vijay Nair, President of ISI for suggesting and facilitating these two important sessions.

I plan to write to you quite regularly in the future (shorter letters), reporting on the progress we make with our plans. In the meantime, I wish you all a very happy and productive year.

Danny Pfeffermann,
IASS President

Appendix

Membership Campaign

A. Preface

The current membership of the IASS consists of 385 individual members and 23 institutional members. This is far from being satisfactory and we need to at least **double** that number if we want our association to be lively and effective.

The larger we are, the more we can do. Let us remind ourselves what our goals are: *to promote the study and development of the theory and practice of sample surveys and censuses, and to Increase the interest in surveys and censuses among statisticians, governments and the public in countries around the world*. This can be done by providing support for international and regional conferences, provision of short courses and workshops, and by providing expert advice to other members and organizations who ask for it. The Association also publishes The Survey Statistician twice a year and sends it to all our members at no extra cost. For this we need to have a strong association, and strong means many members and an extensive budget.

IASS membership consists of Individual and Institutional members. As of November 2013, the IASS consists of 385 individual members and 23 institutional members. The breakdown of the total individual membership by gender shows 279 males and 106 females. A further demographic breakdown of our membership is shown below.

Age Group	Males	Females	Total
0 - 29	0	0	0
30 - 39	9	6	15
40 - 49	29	4	33
50 - 59	24	22	46
60 - 69	61	18	79
70 +	76	9	85
Total, known age	199	59	258
Unknown age	80	47	127
Total	279	106	385

As can be seen, we only have 38 male members and 10 female members under the age of 50, only 9 male members and 6 female members under the age of 40 and no members under the age of 30. **We are aging** very rapidly. Where is the young generation of survey statisticians? Why aren't they regular members of our association? If the trend continues, soon our association will disappear. We need to do something, and now is the time.

B. What should we do?

Every National Statistics Bureau around the world includes many survey statisticians. Several Universities (unfortunately, not too many) have professors and students specializing in sample surveys. There are survey statisticians in the private sector as well. This should be our target population.

1. A face to face, Members – Bring - Members campaign, by which colleagues will bring in colleagues, Professors will bring in students etc. (suggested also by our Council member, Eva Elvers). Under this campaign we ask all current members to approach people they consider potential members and talk to them about the advantages of membership and try to sign them up.

2. Approach heads of statistical offices and urge them to join the IASS as institutional members in addition to encouraging their employees to join as Individual members. I am sure that many such organizations can help in paying the individual fees as well. I plan to write to the heads of NSO's in many countries, asking that they become institutional members but please, don't rely on me and take the initiative.

3. Make sure to have membership desks at conferences, workshops and big seminars.

4. At the last ISI meeting in Hong Kong it was decided that conferences and workshops sponsored by the IASS will be required to offer reduced registration fees for IASS members. This on its own will already make it beneficial to join the IASS.

5. Before and during the ISI meeting in Hong Kong we decided to offer special courses and consulting for Statistical Bureaus in developing countries. This will be part of, and coordinated with the ISI Capacity Building Campaign. Many of our members already volunteered to provide this kind of service free of charge. The idea is that the IASS will pay for the travel costs and the local organizations will take care of local expenditures. The condition for providing this kind of service will be **institutional membership** by the corresponding organizations.

Please share with us other ideas that we shall circulate to all our members.

Professor Mick Couper, one of our council members made the following suggestions regarding institutional membership:

For all institutions, the primary benefit is that of building a network for their survey statisticians, developing the profession, and contributing to the public good. We might need to do different things for developed countries (including private companies and university departments) than for developing countries. The benefits for developed countries may focus more on acknowledging their support and providing access to improved member benefits (e.g., job market, etc.). For developing countries this may require more tangible financial

benefits for membership. Below are some benefits that could be considered for institutional members:

a) Publish the list of institutional members the IASS website.

b) Have each institution nominate several (say, up to 5) individuals who would receive all communications from the IASS, and thus be informed about conferences, short courses, sponsorship opportunities, etc.

c) Reduced rate for short courses and other training events for employees of institutional members e.g., up to 5 employees from institutional members get reduced rate).

d) Free individual memberships for a set number of individuals in such organizations. At the IASS Council meeting in Dublin, it was noted that many countries with foreign currency exchange regulations have difficulties in paying individual dues. It may be administratively easier for an institution to pay both for institutional membership and for a set number of individual memberships (possibly at a reduced rate) in a single payment.

Letter from the Editors

The January 2014 issue of *The Survey Statistician* 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.

In the New and Emerging Methods Section (edited by the Scientific Secretary Mick Couper), Piet Daas and Marco Puts from Statistics Netherlands have contributed an article titled 'Big Data as a Source of Statistical Information'. In the article, they address the research carried out at Statistics Netherlands on the use of Big Data for official statistics, including methodology, privacy and security concerns and the skills required for employing Big Data. In the Ask the Experts Section (edited by Robert Clark), Sharon Lohr from Westat, Inc. answers the question of 'When should a multiple frame survey be used?'. For the Book and Software Review Section, Peter Lynn from ISER, University of Essex has contributed a review of the Committee on National Statistics Report on 'Nonresponse in Social Science Surveys: A Research Agenda'. On behalf of the IASS membership, we wish to thank the authors and editors of these sections for their important contributions to *The Survey Statistician.*

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 Robert Clark (rclark@uow.edu.au). 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 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 Lavallée (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. To facilitate the country reports, we have included the list of current country representatives and their email addresses in this issue. Please contact Geoff Lee (geoff.lee99@bigpond.com) if there is any change or addition to the list of country representatives.

This issue of *The Survey Statistician* includes the first letter from our new IASS President, Danny Pfeffermann. The letter also includes an important appendix on proposals for an IASS Membership Campaign. Please send your comments to Danny at msdanny@cbs.gov.il on how you can contribute to this important task of increasing our membership. We also include the first report from our new Scientific Secretary, Mick Couper.

In the News and Announcement Section, we congratulate Ken Brewer on his prestigious Jo Waksberg Award. We also have news from our IASS Scientific Programme Committee Chair, Christine Bycroft, who is in the process of submitting the proposed IASS Invited Sessions to the ISI Scientific Programme Committee for the WSC 2015 in Rio de Janeiro. In addition, we include a report by Partha Lahiri on

the highly successful satellite meeting on Small Area Estimation held in Bangkok that followed the Hong Kong WSC 2013 conference.

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 2014. We also thank Henry Chiem and Courtney Williamson 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 Henry Chiem, Courtney Williamson and Yovina Joymungul Poorun of 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.

Frank Yu frank.yu@abs.gov.au

Natalie Shlomo natalie.shlomo@manchester.ac.uk

Report from the Scientific Secretary

IASS was well-represented at the 59th World Statistics Conference held in Hong Kong in August 2013. Thanks to the efforts of Eric Rancourt, chair of the Program Committee, IASS held 11 invited paper sessions:

- New developments in multilevel model inference from complex sample survey data (organiser: J.N.K. Rao)
- Issues related to major redesigns in national statistical offices (organiser: Xiuhua Tian)
- Model-assisted approaches to combining information from different surveys (survey data integration) (organiser: Jae Kwang Kim)
- Administrative censuses: Approaches when national population registers are not complete, or not available (organiser: Christine Bycroft)
- Matrix sampling, split-questionnaire: Design and estimation (organiser: Takis Merkouris)
- Rankings, use of ranks, and survey sampling (organiser: Tommy Wright)
- The challenge from web panel surveys (organiser: Jörgen Svensson)
- Response to natural disasters (organiser: Gary Dunnet)
- New developments in small area estimation and applications (organiser: Mike A. Hidiroglou)
- Analytic inference for data from complex surveys (organiser: Jean D. Opsomer)
- Recent developments in imputation (organiser: Yves Tillé)

IASS members were also well-represented in special topic sessions and contributed paper and poster sessions.

In addition, while IASS no longer runs a separate program of short courses at the WSC, our own Steve Heeringa chaired the ISI committee on short courses, which meant that a number of short courses of relevance to IASS members were offered in Hong Kong. These included:

- Business Survey Design (Wesley Yung and Mike A. Hidiroglou)
- Practical Tools for Designing and Weighting Survey Samples (Richard Valliant and Jill A. Dever)
- Editing and Imputation of Survey Data (Eric Rancourt and Jean-Francois Beaumont)
- Analysis of Complex Sample Survey Data (Kirk M. Wolter, F. Jay Breidt, and Jean D. Opsomer)

The program committee for the 60th WSC to be held in Rio de Janeiro in July 2015 is already hard at work, under the leadership of Christine Bycroft. We look forward to an excellent program.

In addition to participation in the WSC, the IASS also promotes and provides limited financial support for regional or satellite conferences or workshops of interest to survey statisticians. Please contact me for further information.

Mick P. Couper
mcouper@umich.edu

News and Announcements

First Asian International Statistical Institute Satellite Meeting on Small Area Estimation, Bangkok, Thailand, Sept. 1-4, 2013

By Raymond Chambers, Partha Lahiri and Jiraphan Suntornchost

The Department of Mathematics and Computer Science at Chulalongkorn University, Bangkok, Thailand hosted the First Asian International Statistical Institute Satellite Meeting on Small Area Estimation (SAE2013) during September 1-4, 2013. The event was co-sponsored by the International Association of Survey Statisticians (IASS), the Survey Research Methods Section (SRMS) of the American Statistical Association, the Institute for the Promotion of Teaching Science and Technology (IPST, Thailand) and the Mathematical Association of Thailand under the Patronage of His Majesty the King, Thailand.

The aim of SAE2013 was to assess the current state of development and usage of small area methodology. The meeting served as a bridge between mathematical statisticians and practitioners working on small area estimation in academia, private and government agencies. Although there have been a number of conferences on small area estimation in the recent past, they have been in Europe and North America, and have had a focus on practitioners in that part of the world. Holding the meeting in Bangkok gave researchers in south-east Asian countries an opportunity to learn about state-of-the-art small area estimation techniques from the experts in the field.

There were 89 meeting participants from 29 countries representing 5 continents. The meeting was organized so that there were no parallel invited sessions, giving participants the opportunity to attend all invited sessions. The 7 invited and the 8 contributed sessions together covered a wide range of theoretical and applied topics in SAE. A special attraction of the meeting was a three-hour invited panel on the evaluation of small area methodology in government programs. Panelists from six different international survey organizations discussed their experiences in implementing small area estimation system in their organizations in an informal setting that encouraged free exchange of ideas. The meeting was preceded by an overview half day workshop by Ray Chambers on small area estimation methods and was followed by a half day course on SAE using R by Santanu Pramanik. The programme, abstracts and slides/papers for the presentations at the meeting are available on the meeting website http://www.math.sc.chula.ac.th/sae2013.

A notable feature of the meeting was the delicious Thai food and drinks that were served throughout the meeting. In addition, the local organizers hosted a fantastic banquet that included a cultural program featuring Thai traditional music and dance. A pre-meeting tour of Buddhist temples in Bangkok was organized especially for the foreign visitors.

SAE2013 was a great success. However, this would not have been possible without inputs from many people. In particular, we would like to thank the organizers of the different invited sessions, the many students and faculty of Chulalongkorn University who helped with the efficient organization of the meeting and of course the participants, especially those who came from faraway places.

Next year, the tradition of SAE conferences continues with SAE2014 scheduled for September 3-5 in Poznan, Poland (www.sae2014.ue.poznan.pl). We wish the organisers of this conference every success. Given the success of SAE2013, this should be assured.

The 2013 Waksberg Awardee: Ken Brewer

The *Survey Methodology* journal has established an annual invited paper series in honour of Jo Waksberg to recognise his contributions to survey methodology. In 2013 the Jo Waksberg Award was given to Dr Ken Brewer (see photograph below). The topic of his paper is "Three Survey Sampling Controversies". This paper was presented to the Statistics Canada Methodology Symposium in November 2013 by Michael Hidiroglou on Ken's behalf. *The Survey Statistician* would like to congratulate Ken for a well-deserved recognition of his contribution to survey statistics.

The editors thank Ray Chambers, Stephen Horn and Bill Gross for the following contribution:

Ken Brewer was born in London in 1931. Although he completed his schooling and early university degrees in the UK (Honours Degree in Physics and Mathematics from Imperial College followed by an MSc in Astronomy from the University of London), his professional statistical life only really started following his move to Australia in 1952 to take up a PhD scholarship in Astronomy at the Australian National University (ANU). Two years into this scholarship, however, he decided that he wasn't cut out to be an astronomer, and so in 1954 he resigned it to join the Commonwealth Bureau of Census and Statistics or CBCS (which subsequently became the Australian Bureau of Statistics or ABS), as a graduate clerk. He was the fifth person to become a member of the new Sampling Section, founded in 1952 under the inspirational leadership of Ken Foreman.

Under Foreman's tutelage, Ken rapidly became a competent survey statistician, and within the space of a few months was designing and analysing samples on his own, with minimal supervision. In fact, by the late 1950s, Ken, with three others to help him, became responsible for all the other survey work of the Sampling Branch (which the Sampling Section had by then evolved into). The CBCS got its first computer in 1959, and Ken was one of the first two people trained up to use the computer.

It was about this time that Ken started to become interested in the conceptual problems that were associated with the statistical theory of sampling at the time. In particular, he became interested in the problem of designing a without replacement sampling scheme with a specified set of inclusion probabilities, and his first two published papers, in 1962 and in 1963, addressed this issue.

Shortly afterwards, and also in 1963, he published what is now regarded as the first serious investigation of the use of population models in sample survey inference. In this paper Ken argued that so-called model-based sample survey inference, i.e. inference based on the variability implicit in a statistical model for surveyed population, could serve as the basis for both efficient sample design as well as estimation. This ran contrary to the prevailing paradigm for survey inference, which held that the only source of variability in survey sampling was that associated with the sample selection process. Ken's 1963 paper showed however that if one were willing to characterise the unknown values making up the finite population of interest as being generated by a random process, then one could model this process and base any subsequent sample inference on the variability imparted by this model. In particular, he showed that if the population values followed the simple linear regression through the origin model, then one could demonstrate that this variance was minimised if one chose the sample deterministically rather than probabilistically. This result was independently discovered by Richard Royall some years later and the discovery led to a heated debate between design-based and model-based survey statisticians.

In the meantime Ken had moved on in his career at the CBCS and had been appointed as the Commonwealth Statistician's Personal Representative for the design and conduct of the 1966 Census of the Australian Territory of Papua New Guinea. This Census was the first with a total coverage of the entire population of Papua New Guinea, and included a Sample Census of the indigenous villages. With the strong support of Foreman in Canberra, Ken was able to bring this enterprise to a successful conclusion.

Ken continued to work on more theoretical considerations in survey sampling, again focussing on methods for unequal probability sampling and also attempting to reconcile the model-based and design-based approaches to survey inference. This last interest was one that cropped up again and again in Ken's research over the next 30 years, always with the aim of providing a unifying perspective that would eliminate the logical differences between these two approaches. In essence, Ken achieved this by first using a population model to motivate an efficient estimator of the population quantity of interest, but then requiring that this estimator also possess good design-based properties. This approach is now referred to as model-assisted approach, which underpinned Ken's 2002 book *Combined Survey Sampling Inference: Weighing Basu's Elephants*.

Ken retired from the Australian Public Service in 1992, and, with characteristic determination, decided to use his increased free time for statistical research. He therefore immediately commenced a PhD in Statistics at ANU on "Reconciliation of Some Apparently Incompatible Approaches to Statistical Inference", which focussed on developing an integrated design-based and model-based approach to survey sampling inference, as well as an exploration of ways of integrating frequentist and Bayesian inference. After obtaining his PhD degree in 1996, he took an honorary appointment at the university, allowing him to carry out research on the fundamental statistical inference problems that he had started to deal with in his doctoral research and also to interact with the statisticians there.

His focus now was on more basic issues than sample survey inference, aiming in effect to provide a resolution to the logical problems that arise in the use of Bayesian methods for hypothesis testing, and in particular, resolution of the so-called 'Lindley Paradox', where he was convinced that a solution could be found that did not require the use of a subjective prior. His investigations into this problem eventually led him to develop an alternative way of measuring the False Discovery Rate that gave values that were intermediate between the widely used AIC and BIC measures of model fit, and seemed more sensible. Furthermore, in the case of normal data and a single free parameter, this approach could be shown to be equivalent, to a Bayesian Hypothesis test as well as being implied by an extended version of Benford's Law of Numbers.

ISI World Statistical Congress, Rio 2015

Report by IASS Programme Committee Chair, Christine Bycroft

Along with other ISI associations, the IASS submits a limited number of Invited Paper Session topics to the ISI for the Rio 2015 scientific programme. Many thanks to those who responded to the IASS call for topic proposals. The number of very good proposals made for some difficult decisions in selecting IASS sponsored topics. The final selection will be made by the ISI Rio Scientific Programme Committee at a face to face meeting in May 2014.

The Rio website is now up and running http://www.isi2015.ibge.gov.br/. There is still plenty of time to think about organising a Special Topic Session - proposals close 16 May 2014. Abstracts for Contributed Paper Sessions are due by 13 Feb 2015. Other important dates can be found at http://www.isi2015.ibge.gov.br/index.php/isi2015/scientific-programme.

One of the changes for 2015 is the on-line submission system. Another is a change in timing for technical sessions, which are now 100 minutes and 120 minutes, offering more flexibility on the number of speakers and whether to include a discussant or not.

Ask the Experts

When should a multiple frame survey be used?

Sharon L. Lohr

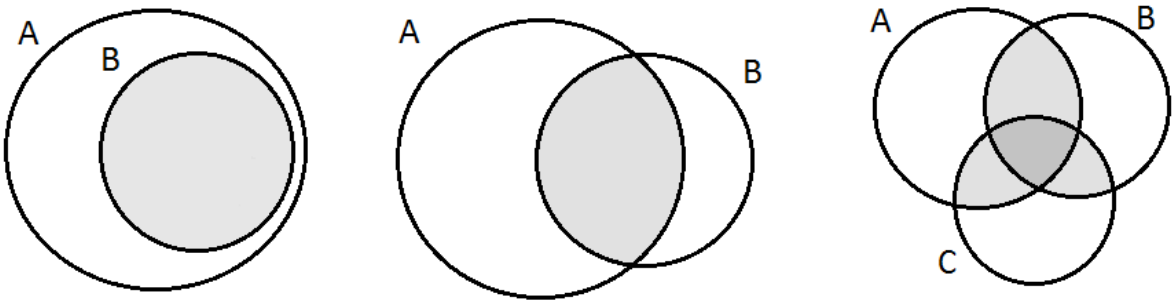
Sharon Lohr is a vice president at Westat, 1600 Research Boulevard, Rockville MD 20850, USA, sharonlohr@westat.com.

Uses and Examples of Multiple Frame Surveys

In classical probability sampling, a sample is selected from a single well-defined sampling frame. The sampling frame is assumed to include the entire population of interest. In a multiple frame survey two or more frames are used, and a probability sample is drawn *independently* from each frame. Each frame can be incomplete, although the union of the frames is assumed to cover the population of interest. We refer to the frames as A, B, C, etc. Three general types of multiple frame surveys are depicted in Figure 1.

Figure 1. Examples of multiple frame surveys. (a) Frame A is complete, and frame B is incomplete. The shaded portion belongs to both frames. (b) Both frames A and B are incomplete. The shaded portion belongs to both frames. (c) Frames A, B, and C are all incomplete. The light shaded portions belong to exactly two frames, and the darker shaded portion belongs to all three frames.

(a) (b) (c)



Examples of multiple frame surveys that correspond to the depictions in Figure 1 include:

- Complete and incomplete frames (Figure 1a). One frame may have full coverage but be expensive to sample, while other frames may be less expensive to sample but incomplete. This is frequently the case when the frame with full coverage is a general population or an area frame, while the incomplete frames are lists. In an agricultural dual frame survey, frame A is often an area frame in which the primary sampling units consist of land areas, while frame B is a list of known agricultural holdings.
- A special case of combining a complete frame with incomplete frames (Figure 1a) occurs when the population of interest is relatively rare within the complete frame (see Kalton 2009, for a review of methods that can be used for sampling rare populations). The rare population of interest may be

persons with a hearing impairment, agricultural holdings on which tobacco is grown, or households containing children who are under two years of age. List frames of persons with a hearing impairment, for example, may be available from support organizations. A high percentage of persons on the list will be in the population of interest, but the list will not contain persons who do not elect to join the organization.

- Dual frame telephone surveys with design corresponding to Figure 1(b) are needed in many countries to obtain complete coverage of the household telephone population, because neither frame A (cellular telephone numbers) nor frame B (landline telephone numbers) covers the entire population.
- Multiple frames may be used to cover a population of interest, as in Figure 1(c). For example, to sample registered nurses who are currently working in a certain state, frame A may be a list of hospitals in that state, frame B may be the list of licensed nurses in that state, and the other frames may be lists of licensed nurses from neighboring states (who might also work in the state of interest). None of the frames, taken separately, covers the entire population, but it is hoped that the union of the frames includes the majority of registered nurses working in the state.

Multiple frames may also be used at other stages in the sampling procedure. For the Survey of Retail Stores described in Hansen et al. (1953), a list frame and an area frame were constructed within each primary sampling unit, so that the subsampling within each primary sampling unit was conducted as a separate dual frame survey.

Estimation

In order to decide whether a multiple frame survey should be used, it is necessary to consider how estimation is typically done. Consider the design in Figure 1(b) with two incomplete but overlapping frames. Three domains are formed by the regions in the frame. Using the notation of Hartley (1962), let a denote the part of the population in frame A but not frame B, let b denote the part of the population in frame B but not frame A, and let ab denote the shaded part of the figure that belongs to both frames. The population total may be written as

$$Y = Y_a + Y_{ab} + Y_b,$$

where Y_d is the population total in domain d, for $d \in \{a, ab, b\}$. The estimator of Y_{ab} needs to account for the fact that individuals in the overlap domain could be sampled either from frame A or from frame B.

The simplest way to account for that multiplicity is to take a weighted average of the estimated population totals in the overlap domain, with $\hat{Y}_{ab} = \lambda \hat{Y}_{ab}^A + (1-\lambda)\hat{Y}_{ab}^B$, where $\hat{Y}_{ab}^A$ and $\hat{Y}_{ab}^B$ are the estimators of Y_{ab} from the surveys taken from frames A and B, respectively. This adjustment multiplies the sampling weights for observations in domain ab of frame A by λ, and multiplies the sampling weights for observations in domain ab of frame B by $1-\lambda$. Other estimators that have been proposed (see Lohr 2011 for a review) minimize the variance of the estimated population total (Hartley 1962), rely on individual inclusion probabilities from the separate samples (Bankier 1986), or use pseudo-maximum-likelihood (Skinner and Rao 1996) or empirical likelihood (Rao and Wu 2010) methods to achieve higher efficiency. Steel and Clark (2010) describe estimators that may be used when frames A and B coincide. All of the methods result in adjusting the weights of observations in the overlap domains.

This weight adjustment has two implications for a multiple frame survey. First, information needs to be collected for each observation that will allow the proper weight adjustment to be made. Typically, the domain membership is unknown before sampling for at least one of the frames. In a dual frame landline/cellular telephone survey, respondents are asked about their cellular and landline telephone usage so that they can be classified into domain a, ab, or b.

The second implication is that even if an optimal survey design is used, the multiple frame survey will be less efficient than a comparable stratified sample from a consolidated unduplicated frame. Consider the situation in Figure 1(a), where A is an expensive complete frame and B an inexpensive incomplete frame. In a dual frame survey, some of the observations sampled from frame A will be in domain ab, and those observations could have been sampled more cheaply from the other frame. If the domain membership of units in frame A is known before the sample is selected, then a stratified sample may be drawn from the two strata a and ab. The inexpensive frame B may be used for all of the sample from stratum ab, so that relatively more of the survey resources may be directed toward improving the accuracy for stratum a.

Design Considerations

Multiple frame surveys can be more complicated to design and more complicated to analyze than single frame surveys. They also provide more opportunities for nonsampling errors. Nonresponse patterns may differ in the samples (see Brick et al. 2011). As seen above, estimates of population totals rely on weight adjustments to compensate for multiple chances of selection in the overlap domains, so accurate information on domain membership is needed to be able to construct unbiased estimators. Samples from the different frames might be collected using different modes or survey procedures, resulting in measurement error. A multiple frame survey can be less efficient than a single frame survey because individuals in the overlapping parts of the frames are not necessarily sampled with the optimal sampling fraction. The following questions can help the survey planner decide whether to use a multiple frame survey.

Is it possible to identify domain membership for survey respondents? A key to obtaining unbiased estimates of population quantities is to be able to tell which survey respondents could have been selected from one or more of the other frames. Sometimes matching methods can be used to determine whether a unit sampled from an area frame is also in a list frame. In other cases, as with dual frame landline/cellular telephone surveys, the respondent must be asked about membership in the other frame. If you cannot determine the information needed to construct an unbiased estimator, then you should not use a multiple frame survey.

Can multiple frames be consolidated into one frame before sampling? In some situations, the frames under consideration may be membership lists from different organizations. For example, the membership lists of different statistical societies might be used as frames for conducting a survey of statisticians. If it can be done, merging and unduplicating the frames and then sampling from the consolidated single frame will usually be more efficient than employing a multiple frame survey.
The resulting merged list can be stratified using information available from the constituent frames. If record linkage methods are used to merge the frames, it may be desired to take a conservative approach and place unresolved, possibly duplicated, records in a separate stratum. The records in this stratum can be resolved later by asking respondents about multiplicity and adjusting the weights for respondents who report different membership information than is recorded in the frame.

Is it necessary to use multiple frames to obtain complete coverage of the target population? If yes, and if merging and unduplicating the frames is not an option, then a multiple frame survey is needed to avoid undercoverage bias. This is the case with telephone surveys in many countries, because neither the landline nor cellular telephone frames include the entire telephone population. In other countries, however, nearly all of the telephone population has cellular service so that Figure 1(a) more closely resembles the situation, with frame A as the cellular frame and frame B as the landline frame. In that case, the cost of a multiple frame approach should be compared with the cost of sampling solely from the cellular frame.

When is it better to use a dual frame survey than to take the entire sample from a complete frame? Suppose that frame A covers the entire population and frame B is incomplete, as in Figure 1(a). The relative efficiency of the dual frame survey depends on the relative costs to sample from the two frames, the proportion of the population that is in frame B, and the variances in the different domains. Hartley (1962) showed that a dual frame survey can result in appreciable efficiency gains if the cost to sample from frame B is much less than the cost to sample from frame A, if frame B contains a large fraction of the population, and if the optimal design and estimator are used. If the costs are comparable, however, or frame B contains a small fraction of the population, the extra administrative costs from a dual frame survey may exceed the potential savings from using frame B.

Should a screening or overlap survey be used? In a screening survey, individuals sampled from one frame are asked if they are also in the other frame; if so, the interview is not conducted. In an overlap survey, interviews are obtained from all eligible respondents from the different surveys and the information is combined later. The decision depends on the costs of screening and obtaining an interview in the frames and on the amount of overlap between the frames; Brick and Lohr (2013) described situations in which a screening survey is more efficient because it avoids the cost of obtaining the interview from persons sampled from frame A who are identified to be in domain ab.

Can multiple frame methods be used with convenience samples? Yes, but the resulting estimates will not have attributes of estimates from a probability sample. For example, suppose a probability sample is taken from frame A but the frame B survey asks individuals to respond to a solicitation on a web page. From a probability sampling viewpoint, individuals responding to the second survey represent only themselves: frame B is the respondents to the web survey. They can be considered a stratum sampled with certainty, but that will generally be a very small part of the population.

Multiple frame surveys will improve coverage or efficiency in many situations. They are not a cure for all problems, however, and they must be designed carefully to balance costs, coverage, and complexity.

References

Bankier, M.D. (1986). Estimators Based on Several Stratified Samples with Applications to Multiple Frame Surveys. *Journal of the American Statistical Association*, 81, 1074-1079.

Brick, J.M., Cervantes, I.F., Lee, S., and Norman, G. (2011). Nonsampling Errors in Dual Frame Surveys. *Survey Methodology*, 37, 1-12.

Brick, J.M. and Lohr, S. (2013). Allocation for Dual Frame Surveys with Nonresponse. Paper presented at the Statistics Canada 2013 International

Methodology Symposium on Producing Reliable Estimates from Imperfect Frames, Ottawa ON.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory, Volume 1.* New York: Wiley.

Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Social Statistics Section, American Statistical Association,* 203-206.

Kalton, G. (2009). Methods for Oversampling Rare Populations in Sample Surveys. *Survey Methodology,* 35, 125-141.

Lohr, S. (2011). Alternative Survey Sample Designs: Sampling with Multiple Overlapping Frames. *Survey Methodology,* 37, 197-213.

Rao, J.N.K., and Wu, C. (2010). Pseudo-empirical Likelihood Inference for Dual Frame Surveys. *Journal of the American Statistical Association*, 105, 1494-1503.

Skinner and Rao (1996). Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association,* 91, 349-356.

Steel, D.G. and Clark, R.G. (2010). Combining Survey Results. *The Survey Statistician,* No. 62 (July), 9-14.

Ask the Experts - Call for Questions

If you'd like to ask the experts a question, please contact Robert Clark at rclark@uow.edu.au.

New and Emerging Methods

Big Data as a Source of Statistical Information[1]

Piet J.H. Daas and Marco J.H. Puts

Abstract
Big Data is an extremely interesting data source for statistics. Since more and more data is generated in our modern world and is digitally stored, it could certainly be used to replace traditional sources or provide additional information for official statistics. Especially given declines in survey response rates, information gathered from Big Data is an interesting addition. However, extracting statistically-relevant information from Big Data sources is not an easy task. In this paper the current state of the art of research on the use of Big Data for official statistics at Statistics Netherlands is described. The paper is based on the real world experiences of the authors obtained during these studies. The topics discussed are related to Big Data methodology, privacy and security concerns and the skills required for successfully employing Big Data.

Introduction
Big Data is a term that one hears more and more often at conferences, meetings and seminars. Since its first introduction in 1997, in a conference paper by Cox and Ellsworth (1997), it has really become a hot topic. This is understandable if one realizes that between the introduction of the term Big Data and the present, the world has changed from a 'data-poor' environment to a world in which data is abundant (Global Pulse, 2012). This is mainly due to the fact that during this period increasing amounts of data have been generated on the web and by sensors in the ever growing number of electronic devices surrounding us. Because of the ongoing decline in the costs of disk storage this data is no longer thrown away but remains stored. As such, Big Data has the potential to provide information on statistically-relevant populations at high frequency, at a high degree of granularity, and from a wide range of angles, narrowing both time and knowledge gaps. This enables the production of more relevant and timely statistics and can result in proxy indicators that enable richer, deeper insights into human experience than traditional sources of official statistics can (Glasson et al., 2013; Global Pulse, 2012).

Anyone who is able to access and analyze Big Data could – potentially – extract meaning from them and gain a competitive edge. This realization has prompted many commercial companies to write white papers and blogs on the huge potential of Big Data. These stories, however, do not always withstand a rigorous scientific analysis and – unfortunately – tend to place the use and potential of Big Data near the edge of the scientific realm. We agree with Glasson et al. (2013), that Big Data has serious potential as it is a very interesting (secondary) data source for official

[1] The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

statistics. However, a rigorous scientific approach is needed to uncover its true potential.

This paper is based on the real world experiences of the authors obtained during the Big Data case studies performed at Statistics Netherlands. An overview of the case studies and their initial findings can be found in the papers of Daas and van der Loo (2013) and Daas et al. (2013). Extracting statistically-relevant information from Big Data sources is not an easy task. In this paper the current state of the art of research on the use of Big Data for official statistics is described. It is the start of the development of Big Data methodology, which is data driven due to the lack of a proper scientific foundation. In addition, the skills needed to perform this work and deal with privacy and security issues to enable this research are briefly discussed.

Big Data characteristics

Big Data is often characterized as data of increasing Volume, Velocity and Variety; the famous 3 V's (Manyika et al., 2011). The volume of what is considered 'big' depends on the capabilities of the organization managing the data and on the capabilities of the applications that are traditionally used to process and analyze it. For a survey oriented organization, the limits of processing Big Data are quickly approached, since in most surveys the sample size is optimized, minimizing the amount of data needed. In this respect, statistical organizations that also process administrative data have a head start. At our office Big Data sources that comprise of up to 100 million records a day are studied. The latter constitutes a file of 25 Gigabytes. Compared to the size of the files routinely used by astronomers and climate researches, i.e. 1 or more Terabytes, these are not enormous amounts of data. However, for a statistically oriented organization they certainly are.

The frequency at which Big Data is generated is the second V (velocity). Statistical offices dealing with a continuous stream of data may initiate the production of more speedy statistics (think of weekly and daily figures) and perhaps even of 'real-time' statistics (Glasson et al., 2013; Struijs and Daas, 2013). It should be noted that, in combination with the data volume, velocity can place high demands on the communication bandwidth available.

The third V refers to variety. This results from the increase in the many different sources that could potentially be used and the variability of the data in these sources. Big Data is often largely unstructured, meaning that it has no predefined data model and/or does not fit well into conventional relational databases. However, the lack of structure can also refer to the fact that little or no information is available on the relationship between data elements. In most cases, this V will increase computational complexity.

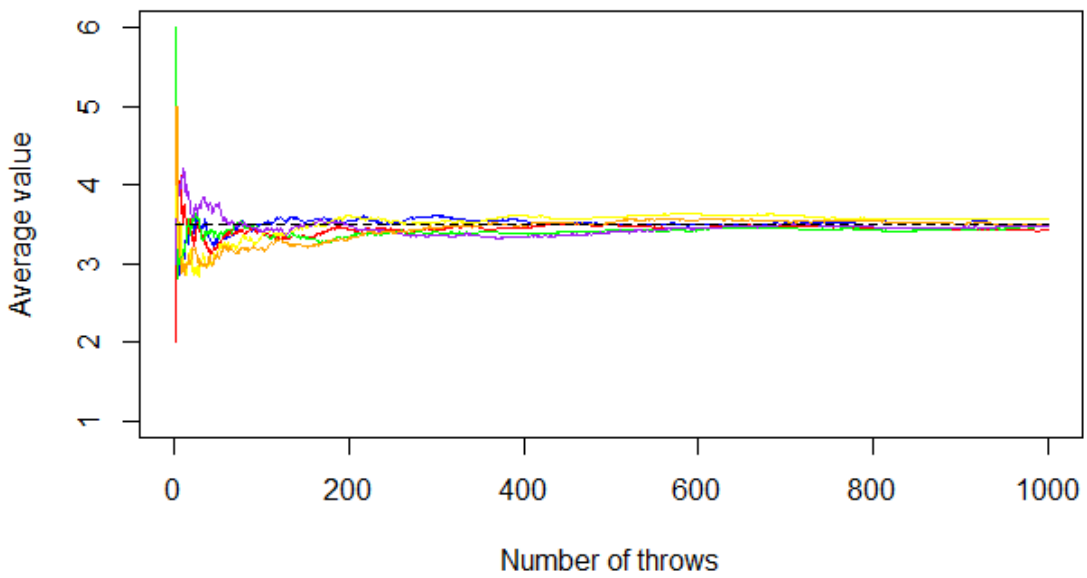
The magnitude of a Big Data source can be seen as the product of the three V's and hence analyzing it can be quite a daunting task. Even more important, however, is the need for a different mindset when performing this task. Working with Big Data requires an open mindset and the ability not to see all problems a priori in terms of sampling theory, e.g. the reduction of variance. Particularly for survey statisticians, who have become accustomed to a data poor environment and – as a result – have developed a focus on extracting the maximum amount of information from (very) small data sets, this may take some time to get used to. However, survey statistics and its methods have adapted to external changes before (Groves, 2011). In our opinion Big Data initiates the need for a change of such magnitude that it truly represents a 'paradigm shift' within the field of statistics (Kuhn, 2012) or even the emergence of a new field of statistical science. Others have a slightly more nuanced view (Walker and Fung, 2013).

Differences between survey and Big Data

The difference between the mindset of using sample surveys and Big Data is illustrated by a plot ordinarily used to demonstrate the law of large numbers. In figure 1 the average of a particular run of throws of a single dice is shown. As the number of throws increases, the average of the values of all the results approaches 3.5; the expected value. Different runs will show a different shape over a small number of throws but over a large number of throws they start to behave very similar. It may take a considerable number of throws to achieve this. The leftmost side of the graphs in figure 1 corresponds to the traditional survey sampling approach.

When small amounts of data are present it is quite difficult to accurately estimate the expected value of the population variable under study. Methods to assure estimates of high accuracy and precision are sought after. When one attempts to use Big Data for statistics the situation more resembles that on the right of figure 12. Large amounts of data are present that, at first, suggest an accurate and precise estimate of the expected value. Note that even after 1000 throws the expected value of 3.5 is not yet reached in several of the runs. The situation will improve when even more runs are performed. This, however, describes the situation in an ideal world.

Figure 1. Development of the average value of 1000 subsequent single dice throws for six different runs. The dotted line represents the expected value (3.5)



Another difference between sample surveys and Big Data becomes clear when looked at the approach traditionally used to derive estimates for population quantities. Suppose a source consists of billions of records, like social media. Big Data sets provide measurements of phenomena at a level of detail far exceeding that of sample surveys. The variability of the phenomena at this level of detail is often found to be large. Sampling of the Big Data set to reduce its size, with the aim to speed up of processing and analysis, only increases the variance of the estimates of the phenomena, thereby forfeiting the benefits Big Data has to offer. In addition, efficient sampling schemes aimed at minimizing the increase in variance are difficult

[2] This comparison is probably not entirely correct, but kept for simplicity. The Big data situation more likely resembles throwing hundreds of dice of all different shapes, with different numbers of sides, while attempting to estimate the average value of the one perfectly cube-shaped dice.

if not impossible to accomplish with Big Data, since stratified or complex sampling is not possible because of a lack of background characteristics of the data records. Simply casting Big Data as a – potentially large – sample is therefore rarely a suitable approach. Alternative data reduction techniques are needed, which optimally reduce the data files in size while retaining as much of their information content as possible.

During our studies we ran into several issues that seriously affect the usability of Big Data for official statistics. Some are methodological, some are practical and some are related to other issues. They are all considered important. A recent UNECE paper confirms these findings and provides a structured overview (Glasson et al., 2013).

The essential issues identified by us are:

1. the ability to process and analyze large amounts of data
2. dealing with noisy, dirty and unstructured data
3. dealing with selectivity
4. ways to go beyond correlation
5. skills needed to perform these tasks
6. legal issues when studying Big Data.

We have not solved all issues yet, but we will describe how we deal or plan to deal with them in the remainder of this paper.

Big Data issues

Processing large amounts of data

When one wants to analyze Big Data, one needs to have a computing environment that enables the rapid processing of large amounts of data. We found – and this is important – that for some data sources analyzing samples or parts of data did not suffice. In many Big Data sources the information content is low, meaning that not all data included is relevant for the research question under study. The low information content is the result of a combination of reasons, the most important of which are: the fact that the data is not generated for the particular purpose the user has in mind (it is secondary data); the data is often unstructured and noisy (more on this below); only a limited number of variables are available; the phenomenon in which the researcher is interested does not occur often. To make successful use of Big Data large amounts of data need to be processed (Boyd and Crawford, 2011). Analyzing a small portion of data may be a good way to start your 'Big Data' studies but the findings derived (if any) certainly need to be verified by those obtained from analyzing the whole dataset. It is through using all data that the true value of Big Data comes out, certainly when you are interested in rare events.

Analyzing Big Data is routinely performed in our office with R or Python, but if a researcher is more comfortable in another programming environment this is no problem. The most important skill here is knowing how to write a program that is able to access all the data in a Big Data set within a reasonable amount of time. Having a secure computer environment with many fast processors, large amounts of RAM and fast disk access certainly helps. Several important considerations are described in Scannapieco et al. (2013) and NAS (2013). Parallel processing could be a way to speed things up. We are currently using (multi-core) general-purpose computing on graphics processing units and are looking at distributed computing, such as (our own) secure private cloud or a local cluster. When studying large amounts of data, creating visual representations is a good way to start to get an idea of their content (Frankel and Reid, 2008).

Dealing with noisy, dirty and unstructured data

As is the case for many secondary data sources and certainly for sources containing huge amounts of data, not all records are relevant for the purpose the researcher has in mind. The irrelevant records, the 'noise', may even negatively affect the relevant ones. We have analyzed quite a number of Big Data sources and have noticed that the signal-to-noise ratio is often rather low. In general the data can be considered as mostly noise; only a mere fraction of the data is of interest – the signal (Silver, 2012). As such, one could consider a lot of Big Data studies as finding a needle in a haystack. However, sometimes the needle also resembles hay. Finding ways to reduce the noise in Big Data, thereby increasing the signal-to-noise ratio, is vital for obtaining a successful result. Aggregating or applying a filter, such as a Kalman or a filter based on Poisson distributed noise (Manton et al., 1999), have been successfully applied, as are constructing queries that prefer the inclusion of relevant records. Another approach routinely used is removing all records with data that are clearly corrupt; dirty records. If one analyzes 100 million records, removing 100 erroneous records is easier than considering the relevance of each 'outlier', their effect on the estimate, and attempting to correct them. One needs to be pragmatic when analyzing Big Data. Applying dimension reduction methods, such as principal component analysis, factor analysis or self-organizing maps, are other ways to decrease the size of the data without losing much information (Hastie et al., 2009).

For unstructured data, such as texts or pictures on web pages, the first step of analysis differs. This type of data is usually transformed into a form more appropriate for statistical analysis; such as word frequencies and their distribution or the clustering of pictures into similar groups. Methods for performing this initial step can be found in areas of science more accustomed to dealing with this kind of data; such as machine learning (Murphy, 2012; Breiman, 2001). We can also learn a lot from Google or other internet giants here (Scott et al., 2013; Spector et al., 2012).

Information theory (Shannon, 1948) can provide more insight into the nature of Big Data. Besides noise, as mentioned above, redundancy also contributes significantly in the total volume of Big Data. Whereas noise reduction can lead to a 'lossy' compression, i.e. part of the data is removed, redundancy reduction leads to a lossless compression of Big Data. In the latter case the data can be reconstructed perfectly. By defining the signal-to-noise ratio in terms of bit-reduction, we could get some appreciation of target information hidden in Big Data.

In all cases, information is extracted from Big Data, usually reducing its size and bringing it more in line with the size of the files statisticians are accustomed to dealing with. In subsequent steps many familiar statistical methods can be used.

Dealing with selectivity

Despite the huge amounts of data present in Big Data sources they may still not cover the complete target population considered. Big Data may therefore be selective. This is an important issue in relation to using Big Data for official statistics where it is – especially when viewed from the perspective of sampling theory – a prerequisite for valid inference. It is therefore essential that this issue is addressed in the context of Big Data. There are several important points to consider.

The first one is the realization that selectivity may vary tremendously per Big Data source and per target variable. Although for some sources this is indeed a major issue, for others coverage is almost or essentially complete, due to the nature of the process through which the data come about. Even when coverage is partial, this may still result in a considerable amount of data for that particular group. Methods that could be used likely resemble post-stratification without having variance issues. This

seems to suggest that the complete absence of a particular part of the population is a more important issue. It will be a challenge to accurately determine this.

The second point is related to the first one. As there is no sample design for Big Data, ways to correct for missing data are needed and will very likely require a model-based or an algorithmic based approach (Buelens et al., 2012). Model-based methods require estimating the model parameters. This pivotal task is challenging for data sources with hardly any auxiliary variables. When models are not tenable, approaches inspired by non-probability correction methods (Baker et al., 2013) or by the area of data-mining and machine-learning (Breiman, 2001) may provide solutions.

The third point is – when applicable – an issue that needs to be addressed first. A lot of Big Data sources actually register events or – more correctly – aggregates of events. This is the reason why many of these sources are big. Examples of (aggregates of) events are: the content and time when a social media message is written, the start and end times and location(s) of a call made by a particular phone and the number of vehicles passing a road sensor at a particular location at a particular point in time. In fact, a major part of the Big Data sources studied by the authors were found to be event-based and hardly contained information on the statistical units of interest. Since the events stored are (indirectly) caused by the statistical units, e.g. people or businesses, dealing with them in a traditional way requires these events to be converted to the corresponding units first. This may not be easy, as a limited amount of identifying information is available. Perhaps the additional use of other (Big Data) sources may assist here. Considering the above, this suggests a two-step approach in which the first step consists of deriving profiles from events to identify units or subpopulations (groups of units). The subsequent step uses inference methods based on the information provided by these. In this context, it is however also important to realize that the part of the population included in Big Data sets might be representative for the whole population for a particular variable studied. Or that, even when it is not, Big Data might still be used to produce (biased) estimates providing that they strongly correlate with existing statistics, for example to improve accuracy and speed of the existing statistics or merely to reduce the sample size.

Correlation and causation

Because of the huge amounts of data available, comparing it to data from another (survey) source may very likely result in a correlation between a particular Big Data variable and a survey variable. However, a high correlation does not always imply causation. In fact having access to more data increases the chance that correlations are found. It is the difficult task of the researcher to fully investigate this ‘relation’ and try to distinguish a true from a false correlation which may also be referred to as coincidental or spurious. It is best to attempt to falsify the relationship (before anyone else does) by performing additional analysis. Because of its sheer volume this may take some time for Big Data. Patience can be a virtue here, as a longer data series may provide clues to the stability of the relation observed. However, one also needs to realize that the correlation observed may provide a hint at something very interesting. If the correlation cannot easily be falsified, more rigorous analysis aimed at confirming a causal relationship needs to be performed (Pearl, 2009). Be aware that this can be challenging as it requires analyzing lots and lots of data and may even require combining several (Big) data sources. We use cointegration and structural time series studies as our first next step (Krieg and van den Brakel, 2012).

Competencies needed

All of the above merely describes part of the skills currently assigned to survey statisticians. Some of these skills are also new to statisticians experienced in the study of administrative data. Certainly knowledge of high performance computing approaches and algorithmic ways of inference are not types of expertise routinely observed in official statistical environments. Usually survey statisticians, 'register' statisticians and IT-personnel are found in such surroundings. This, in combination with the need for a different mindset and a data-driven (pragmatic) way of analyses prompts us to the sexy "new kid on the block" of statistics; the elusive Data Scientist. This jack of all trades does indeed harbor many of the skills mentioned above. The data science skills usually mentioned are mathematics, statistics, machine learning, computer science and high performance computing, visualization, communication and business/domain expertise (Schutt and O'Neill, 2013). Combined with problem solving skills, perseverance, creativity and an open mind set, these scientists certainly seem suited for the task at hand. However, despite the lack of muscle strength, they seem more to resemble supermen or superwomen. One may seriously consider if a single person is able to gain sufficient expertise in all of these fields. We think that there is a need to create multidisciplinary teams in which people are involved that, as a group, cover all data science skills mentioned above. Such a group should pragmatically tackle Big Data – with an open mind set – from various viewpoints to extract information with the aim of creating statistics. At our office we are currently at the verge of constructing such a group. Depending on its success this group may increase in size.

Privacy and security issues

The Dutch data protection act allows scientific and statistical research on data sources such as Big Data provided appropriate security measures are taken when dealing with privacy-sensitive data. This enables us to perform our research studies on the potential use of Big Data for official statistics. However, the routine production of statistics based on Big Data is another matter. There are several issues, real or perceived, that may impede its routine use (Struijs and Daas, 2013). Data ownership and copyright may be an issue, along with the purpose for which data are registered. Even if data are publicly accessible, for instance on websites or as social media messages that do not have access restrictions, questions of ownership and purpose of publication can be raised. Even the collection of internet data via web robots can be negatively perceived as it causes a burden on the providers of the sites. And even if there are no legal impediments, the perception of the public is a factor to take into account. These concerns have to be taken seriously. Fortunately, there are measures that can be taken to overcome at least some of the obstacles, for example, by anonymizing unique identifiers, removing the privacy sensitive part of a Global Positioning System track (e.g. the first and last 100 meter) or by using informed consent. If a reduction of response burden can be offered, this can be very helpful, also in getting the support of the general public. For the long run, changes in legislation may be considered, to ensure continuous data access for official statistics. But it remains important to stay in line with public opinion, because credibility and public trust are important assets. Within the European Union, changes in the European legislation must also be considered. In addition to the national laws, European laws or regulations can impede the collection of data, even if the Dutch legislation does not present any problem.

Final conclusions

From the above, it is clear that the use of Big Data as a data source for official statistics has considerable implications. Most of all it will certainly affect the work field of statisticians engaged in that area. For them, new skills that go beyond the ones traditionally considered for statisticians, are needed to unleash the true potential of

Big Data. With this, data science becomes introduced in statistical organizations. In addition, and to assure the valid use of Big Data, the need emerges to develop Big Data methodology (West, 2013). Methods specific for the statistical analysis of Big Data must be developed to solve the important methodological challenges describes in this paper. Without such methods, only a limited number of Big Data sources could be used for statistics in an appropriate way. In our opinion, this requires two major changes. The first is shifting the focus of statisticians towards secondary data. The second is changing the mindset to a state that enables collaboration with experts in areas of science with a different statistical culture (Breiman, 2001; Kass, 2011). There are a number of research areas in which considerable expertise on the analysis of large data sets has been developed already. In this respect it is encouraging to see that Big Data gets increasing attention within the international statistical community. At the recent Joint Statistical Meeting and the World Statistics Congress a number of Big Data sessions were held. In addition, several international taskforces have been formed, in which plans for cooperation in this exciting field of research are emerging. Together, the statistical community can certainly face the future with confidence, provided there is a willingness to adapt. Exciting times lie ahead indeed!

Acknowledgments
The authors gratefully acknowledge their Statistical Netherlands colleagues Bart Buelens, Chris de Blois, Rob Willems, Elke Moons, Jan van den Brakel, Peter Struijs, Barteld Braaksma and Marton Vucsan for stimulating discussions and constructive remarks. Mick Couper is gratefully acknowledged for inspiring us to write this paper.

References

Baker, R., Brick, J.M. Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J.,

Tourangeau, R. (2013). Report on the AAPOR Task Force on Non-Probability Sampling. AAPOR report, May.

Boyd, D., Crawford, C. (2011). Six Provocations for Big Data. Paper for A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 21.

Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science 16 (3), pp. 199-231.

Buelens, B., Boonstra, H.J., van den Brakel, J., Daas, P. (2012). Shifting paradigms in official statistics: from design-based to model-based to algorithmic inference. Discussion paper 201218, Statistics Netherlands, The Hague/Heerlen.

Cox, M., Ellsworth, D. (1997). Application-Controlled Demand Paging for Out-of-Core Visualization. Proceedings of the 8th IEEE Visualization '97 Conference, pp 235-244.

Daas, P.J.H., Puts, M.J., Buelens, B., van den Hurk, P.A.M. (2013). Big Data and Official Statistics. Paper for the 2013 New Techniques and Technologies for Statistics conference, Brussels, Belgium.

Daas, P., van der Loo, M. (2013). Big Data (and official statistics). Paper for the 2013 Meeting on the Management of Statistical Information Systems (MSIS 2013), Paris - Bangkok, France - Thailand.

Frankel, F., Reid, R. (2008). Big Data: Distilling meaning from data. Nature 455, p. 30.

Glasson, M., Trepanier, J., Patruno, V., Daas, P., Skaliotis, M., Khan, A. (2013). What does "Big Data" mean for Official Statistics? Paper for the High-Level Group for

the Modernization of Statistical Production and Services, United Nations Economic Commission for Europe, March 10.

Global Pulse (2012). Big Data for Development: Opportunities & Challenges. White paper, Global pulse, New York, USA, May.

Groves, R.M. (2011). Three Eras of Survey Research. Public Opinion Quarterly 75, pp. 861- 871.

Hastie, T., Tibshirani R., Friedman J. (2009). The elements of statistical learning: data mining, inference, and prediction. Second Ed., Springer.

Kass, R.E. (2011). Statistical Inference: The Big Picture. Statistical Science 26 (1), pp. 1-9.

Krieg, S., Van den Brakel, J.A. (2012). Estimation of the Monthly Unemployment rate for Six Domains through Structural Time Series Modelling with Cointegrated Trends. Computational Statistics and Data Analysis 56 (10), pp. 2918-2933.

Kuhn, T. (2012). The Structure of Scientific Revolutions: 50th Anniversary Edition. University of Chicago Press, Chicago, USA.

Manton, J.H., Krishnamurthy, V., Elliott, R.J. (1999) Discrete Time Filters for Double Stochastic Poisson Processes and Other Exponential Noise Models. Int. J. Adapt. Control Signal Process. 13, pp. 393-416.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. (2011). Big Data: The next frontier for innovation, competition, and productivity. Report of the McKinsey Global Institute, McKinsey & Company.

Murphy, K.P. (2012). Machine Learning: a Probabilistic Perspective. The MIT Press, USA.

NAS (2013). Frontiers in Massive Data Analysis. The National Academies Press, Washington, D.C., USA.

Pearl, J. (2009). Causal inference in statistics: An overview. Statistics Surveys 3, pp. 96–146.

Scannapieco, M., Virgillito, A., Zardetto, D. (2013). Placing Big Data in Official Statistics: A Big Challenge? Paper for the 2013 New Techniques and Technologies for Statistics conference, Brussels, Belgium.

Schutt, R., O'Neill, C. (2013). Doing Data Science: Straight Talk from the Frontline. O'Reilly Media, Sebastopol, CA, USA.

Scott, S.L., Blocker, A.W., Bonassi, F.V., Chipman, H.A., George, E.I., McCulloch, R.E. (2013). Bayes and Big Data: The Consensus Monte Carlo Algorithm. Bayes 250, to appear. Available at: http://static.googleusercontent.com/media/research.google.com/nl//pubs/archive/41849.pdf

Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27, pp. 379–423 & 623–656, July & October.

Silver, N. (2012). The Signal and the Noise: Why So Many Predictions Fail — but Some Don't. Penguin Group, New York, USA.

Spector, A., Norvig, P., Petrov, S. (2012). Google's Hybrid Approach to Research. Communications of the ACM 55 (7), pp. 34-37.

Struijs, P., Daas, P.J.H. (2013). Big Data, Big Impact? Paper for the Seminar on Statistical Data Collection, Geneva, Switzerland.

Walker, D., Fung, K. (2013). Big Data and big business: Should statisticians join in? Significance 10 (4), pp. 20-25.

West, G. (2013). Wisdom in Numbers. Scientific American, May, p. 14.

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

Nonresponse in Social Science Surveys: A Research Agenda,
R. Tourangeau and T. J. Plewes
National Academies Press, Washington DC (2013)

Reviewed by Peter Lynn

This publication reports the findings of a panel that was convened by the US National Research Council's Committee on National Statistics. The panel was convened in response to concerns about the threats to statistical inference from the problems associated with declining response rates in traditional social science surveys. An initial planning meeting determined that the objectives of the panel's work should be a) to assess what is known about the causes and consequences of increasing nonresponse, the current state of survey methodology, and methods designed to improve response for surveys in the government, academic, and private sectors, and b) to identify high-priority research that can answer important unresolved questions about survey response and determine the most cost-effective ways to improve response and the quality of survey data for the advancement of knowledge in the social sciences. The focus was on surveys of households and individuals in USA, though most of the findings apply also to other situations (a focus on finding suitable alternatives to telephone surveys is one of the few peculiarly American topics).

The objectives were achieved by an extensive literature search and by convening two workshops to which experts were invited to make presentations on specific topics. The first workshop, which took place in February 2011, featured reviews of the state of knowledge about the role of field operations in achieving high response rates, the current status of research on mode effects, evidence on effectiveness of incentives, research on postsurvey adjustments for nonresponse, and new metrics for nonresponse. Papers from that workshop and the initial planning meeting were published in a special issue of *The Annals of the American Academy of Political and Social Science*, "The Nonresponse Challenge to Surveys and Statistics," (Volume 645, January 2013).

The second workshop, in April, 2011, reviewed international research on nonresponse; the role of interviewers in achieving high response rates; models for survey costs; current issues and practices in mixed-mode survey research; and nonresponse in social network surveys and respondent-driven sampling methods.

This volume draws together all the information gathered by the panel about the current state of knowledge, current practices, and methodological research and then proceeds to propose an ambitious but thoughtful research agenda that aims to identify how practices might be improved in the future in order to prevent non-response from doing serious damage to the utility of survey research. The first four chapters of the report summarises knowledge in the following areas: "The growing problem of nonresponse" (33 pages), "Nonresponse bias" (11 pages), "Mitigating the consequences of nonresponse" (10 pages), and "Approaches to improving survey

response" (40 pages). The fourth of these chapters includes useful discussion of multiple mode designs, responsive designs, incentives, the role of interviewers, and the use of paradata and administrative data, among others.

The intentions of the panel, and of the Russell Sage Foundation, who proposed and funded the work of the panel, are laudable indeed. The scope of the work is impressive. Collectively, this report and the special issue of *The Annals of the American Academy of Political and Social Science* referred to above provide an accessible and authoritative overview of what is known about the most important aspects of survey nonresponse. This encompasses trends in the level and nature of nonresponse, the changing nature of survey practice, and the range of procedures embraced by survey researchers to reduce nonresponse and nonresponse bias, and what is known about the effects of these procedures. These two publications provide an invaluable reference source for survey researchers. A pdf of the panel report can be downloaded for free from http://www.nap.edu/catalog.php?record_id=18293. (The Annals are unfortunately only available to subscribers, though the contents page and abstracts can be read for free at http://ann.sagepub.com/content/645/1.toc).

With such a wide range of related topics to cover, it is inevitable that some important details are omitted. For example, in some places discussion of the effects of various design features concentrates on the overall, sample-wide, effect, with little or no discussion of the possible heterogeneity of the effect. The heterogeneity of effects is certainly an under-researched topic and better understanding could help the research community to identify how to effectively target various design features at different types of surveys or different types of respondents.

However, it is the short chapter 5 of the report (4 pages) that could have the biggest influence. This is where the panel summarises a proposed research agenda for filling gaps in our knowledge about nonresponse and methods for tackling it (more details can be found in the relevant earlier sections of the report). No fewer than 26 broad research topics are identified. The proposed research includes basic research designed to help understand the nature of the problem, research to clarify the factors that provide positive motivation, research into the cost implications of nonresponse and how to capture cost data in a standardized way, research into adjustment methods, research into the effects of response enhancement methods on nonresponse bias, and research into cost-error trade-offs, responsive designs, and other approaches that could help to identify policies and procedures that would assist in constraining the problem of nonresponse. The panel does not suggest how the research should be funded or by whom it should be carried out, but anticipates a range of sources and contributions. Let's hope that some major funders take up the initiative. The work of this panel certainly deserves to bear fruit and the survey research world benefit from authoritative research on these topics.

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).

Country Reports

ARGENTINA

Veronica Beritich

Survey on Sexual and Reproductive Health

The National Institute of Statistics and Censuses (Instituto Nacional de Estadística y Censos - INDEC) conducted, on May 18-July 5, 2013 the field work for the first Survey on Sexual and Reproductive Health (Encuesta sobre Salud Sexual y Reproductiva) in Argentina.

The general objective of this survey is to generate information about sexual and reproductive health of men and women throughout the country. Its geographic scope encompasses all towns of 2,000 or more inhabitants. The target population is composed of women aged 14 to 49 years and men aged 14 to 59 years, living in dwellings visited by the Annual Survey of Urban Households (Encuesta Anual de Hogares Urbanos - EAHU) during the third quarter of 2012.

The topics consulted were the following:

• Sociodemographic, labour and educational characteristics.
• Knowledge and use of contraception at first intercourse, and at present.
• Preventive practices by women.
• Fertility.
• Care and controls during pregnancy of the last child born alive.
• Lactation.
• Information about sexually transmitted infections.

General information on this survey can be found at www.indec.gov.ar.
For further information, please contact ces@indec.mecon.gov.ar.

AUSTRALIA

Ross Watmuff and Bernard Baffour

Measuring impact of introducing web forms to Australian Bureau of Statistics' surveys

The Australian Bureau of Statistics (ABS) is introducing web forms to many of its surveys, and the method of measurement of the impact on data has differed between household surveys and business surveys.

In the case of the monthly household Labour Force Survey (LFS) which produces the nation's key economic employment indicators, respondents are now encouraged to complete the survey themselves via a web form but also have the option of computer-aided face-to-face interview (generally for the first month in sample) or computer aided telephone interview. The introduction of a new collection mode for some sample households required a change in the processing strategy e.g. how long to wait for a web response before switching to follow-up using another mode. The potential statistical impact of the process change was therefore not limited to an impact arising solely from the different reporting behaviour of respondents across alternative modes, but includes a wholesale change to collection strategy.

The area-based LFS sample of dwellings is divided into eight 'rotation groups', each of which remains in the sample for eight months. However, each month the dwellings from one rotation group are replaced by a new sample of dwellings, meaning 7/8ths of the sample remains common in consecutive months. Each rotation group is designed to be a representative sample of the population.

Web forms were introduced in an embedded experiment with half of the dwellings in each rotation group being offered the option to complete the survey online, beginning with their first month in sample, and the other half remainder as a mixture of face-to-face and telephone. The differences in the estimates from each of these twin halves of rotation groups formed the base estimate of impact from which a whole of sample estimate of impact was formed. Each of the half rotation groups continues to be subject to their initial collection process allocation - web form offer, or no web form offer - during their eight months in the sample. Data from rotation groups were pooled across months to maximise the power of the statistical test to detect any difference between the two groups. As of September 2013 the measurement period is half-way completed and to date it has not found any significant impact caused by the introduction of web forms.

In the vast majority of business surveys, any possible web form impact was expected to be less than that for household surveys, as the move is from one mode of self-enumeration (paper) to another (web). Consequently, web forms have been offered to the whole sample at the same time and, given the lack of option for a controlled experiment, the ABS employed a propensity scoring method used by Statistics Canada to assess how the introduction of web forms has impacted on business surveys. The probability that a business elects to respond using a web form rather than a paper form was modelled based on characteristics of that business, such as its industry and its number of employees. By comparing businesses with similar propensities for responding over the web, the difference between web and paper response values was estimated and the significance of this difference calculated. The underlying assumption is that, by using the propensity model to control for

differences in business characteristics, any remaining differences between web and mail responses can be attributed to the response mode.

In practice, there have been several limitations on our use of this method to identify significant differences between response modes. First, the probability that a business elects to use a web form depends on factors that have not been available for inclusion in the propensity model, such as characteristics of the person who completed the form. This may introduce bias to estimates of the differences between response modes. Second, this method is most powerful where there is a good mix of web and paper responses, but for most ABS business surveys about 80-90% of respondents have used the web form where available. The small numbers of paper form respondents may prevent real mode differences from being identified as statistically significant and indeed, to date, there has been no statistical evidence of a mode impact for the business surveys that have moved across to web forms.

For further information please contact Ross Watmuff, Australian Bureau of Statistics via email: ross.watmuff@abs.gov.au.

Enhancing social research in Australia using dual-frame telephone surveys

Using national surveys for social research is important for the measurement of social and health related outcomes. The social research industry typically conducts surveys by sampling from lists of landline telephone numbers. However, it is estimated that one in five people are currently resident in mobile phone only households in Australia, and this population appears to be increasing. Further, people who use mobile phones tend to have different characteristics and outcomes from people who use landline phones only. To ensure full-coverage of the population and accurate estimates, a dual sampling framework is required that includes both landline and mobile phone numbers. National sampling frames for mobile phone numbers have not been readily available to the social research industry due to confidentiality and privacy issues, and numbers generated by random digit dialling are not linked to geographical indicators. Additionally, total mobile phone usage by geographical area is not available for weighting purposes.

The research landscape is very limited and concentrated heavily in the USA. But country-specific differences in culture, socio-economic conditions, demographic factors and availability of mobile phone usage data, make it imperative that Australian researchers develop dual-frame telephone surveys that are tailored to local needs. Further, existing research has been concentrated on the theoretical development of dual-frame surveys and there is a substantial gap in knowledge of the practical application of dual sampling frames in social survey research.

Therefore, the Institute for Social Science Research at the University of Queensland and the Social Research Centre are collaborating together to investigate alternative approaches to obtaining dual sampling frames for achieving full-coverage of the population and derive optimal strategies for producing improved statistical population estimates of socially important outcomes in Australia. This project will be the first in Australia to allow research academics and survey practitioners to work together in developing the design, conduct and analysis of a survey using data from a combined landline and mobile phone sampling frame. The objectives of the project are, firstly, to establish a sampling methodology for undertaking dual frame telephone surveys in Australia using both landline and mobile phone data; secondly, to identify the optimal strategies for combining survey data obtained from dual landline and mobile phone sampling frames; and thirdly, to present the methodologies developed into easy to

follow practical guidelines for the mainstream application of dual telephone sampling frames in Australia.

For more information contact Bernard Baffour at b.baffour@uq.edu.au.

CANADA

Martin Renaud

The Industry Statistics Branch Monthly Surveys System Integration Project at Statistics Canada

Statistics Canada conducts a fair number of monthly economic surveys. Among these surveys are the Monthly Survey of Manufacturing (MSM), the Monthly Retail Trade Survey (MRTS), the Monthly Wholesale Trade Survey (MWTS), and the Monthly Survey of Food Services and Drinking Places (MFS). Up until approximately 2010, these four surveys were using their own custom-built survey processing systems. However, under Statistics Canada's corporate business architecture initiative which promotes the use of generalized systems to reduce the number of systems supported by the corporation, it was decided to undertake a project to harmonize the survey processing of the four surveys by making use of generalized systems. At the same time, given that all four surveys share common features in their sample design, several aspects of their respective post-collection methodology were targeted for harmonization. In particular, outlier detection methods, edit and imputation strategies and estimation were three components which harmonized as much as it was possible without having to go through a complete redesign of the surveys.

The implementation of this new integrated processing environment occurred in three stages. After some intensive testing, the first survey to successfully enter the new system was the MFS in December 2011. The second stage was devoted to the MSM which also went through rigorous testing before finally being integrated in the new system in September 2012. Finally, given the many similarities between the data processing and methodology of the MRTS and the MWTS, the last stage consists of integrating both surveys into the new system. Their full integration is planned for January 2014.

Now that this project is nearing completion, a new component has recently been added. The main goal of this component is to harmonize as much as possible the methods used before data collection for all four surveys. Aspects of the sample design such as the sampling unit, the level and detail of stratification, and the way administrative data is used, among others, are being considered. Upon completion of this component, which is scheduled for 2015, all four economic surveys will be fully integrated under one common methodology and data processing system.

INDIA

Dr. Gayatri Vishwakarma

A very interesting report on "Scenarios: Shaping India's Future" is issued by Planning Commission of India in July 2013. In this report the relationship between the fundamental forces that emerges from Indian systems analysis is presented adequately. The system analysis reveals three scenarios of India. They can be described under the headings

1. Insufficient Action (or "Muddling Along"),
2. Policy Logjam ("Falling Apart"), and
3. Strong, Inclusive Growth ("The Flotilla Advances")

The first scenario talks about reforms. The second scenario shows that this scenario emerges from India remaining stuck in a centralized governance system whose theory-in-use is to exert control in the face of demands for devolution, by centralized mega-schemes and projects, and by "redistribution" of wealth through a system of "handouts" and subsidies.
The last scenario demonstrates the India's future with a federal governance system in which the wheels begin to mesh more smoothly, local governance institutions and small enterprises are nurtured and grow effectively.
One of the chapters of the report tells about "India at a Turning Point: Which Scenario Will Emerge?" represented nicely and it is worth reading. In a substance report enlighten that India's current situation is described as Insufficient Action or Muddling Along. If India now focuses on implementing the overdue governance reforms, it can expect a speeding up of the India flotilla's progress and cohesion. If India does not implement governance and institutional reforms very soon, it can expect a further falling apart of the flotilla and the India growth story.

NEW ZEALAND

Lynsey Hayes

Geocaching puzzle a novel way of promoting New Zealand's new census data
New Zealand's first *census of population and dwellings* in seven years was conducted on 5 March 2013. The first results from the 2013 Census were released in October 2013. One novel approach for spreading the word about Statistics NZ and our census data was based on the 'sport' of geocaching.

Geocaching is a recreational activity where people use GPS devices to locate boxes that are hidden all around the world. People post the coordinates of their box, or 'geocache', online, and then other people try and find the container, sign their name in the logbook, and return the geocache for others to find. On census night this year, a census themed geocache was hidden and geocachers had to use the Interactive Boundary Maps tool on the Statistics NZ website to solve a puzzle to find out where the box was located.

Online push strategy for future census
Statistics NZ is investigating options for how the *future census* will be carried out. As part of this work we ran a data collection pilot in one census collection district, Oamaru, during the 2013 Census. The major focus of the pilot was to promote the Internet as people's first choice of response mode (an 'online push strategy') and learn as much as we could about maximising response in the live, rather than test, census environment.

Using an address list compiled from administrative data sources, we sent people in Oamaru a letter with the census web address and individual ID numbers so they could choose to complete the census online. Paper forms were only provided when people requested them from the contact centre. As a result, over 65 percent of census forms from Oamaru were received online, compared to around 35 percent across the rest of the country.

Overseas experience and our Oamaru pilot make us confident that this proposed collection model can work in the New Zealand environment. It also showed us that implementing new methods across New Zealand will require meticulous design and testing work to maintain current census targets. We have learnt a lot about the strengths and weaknesses of the proposed changes and made significant progress in our thinking and understanding – although we still have a lot to learn. For further information, please email Pat Coope: pat.coope@stats.govt.nz.

Developments in household surveys
Statistics NZ maintains an *area sampling frame* for use in household surveys, consisting of primary sampling units (PSUs) which are mostly contiguous aggregations of Statistics NZ meshblocks. A meshblock is the smallest geographic unit for which statistical data is collected by Statistics NZ. It is usual to review the PSUs following each census, and the review following the 2013 Census marks the first time the reformation of these PSUs will be automated using the ArcGIS software in conjunction with the Python scripting language. In addition to the PSU reformation, a number of other changes are intended, including a major review of stratum definitions and overlap control methodology. Statistics NZ is particularly committed to minimising respondent burden across the Official Statistics System, and a review of the overlap control methods makes explicit coordination of survey loads across other government departments more feasible. Increasing financial pressures also mean that investigations into existing field practices are being reviewed, with a particular focus on the potential of administrative address data to reduce in-field enumeration activity and highlight areas of high dwelling growth. For more details, email Chris Hansen: chris.hansen@stats.govt.nz.

As part of Statistics NZ's major review of our areal sampling *overlap control methodology*, a new method is being considered to help control response burden in household surveys.

Currently we quarantine PSUs for a year after the completion of a survey.
We are now experimenting with a conditional overlap control method that allows overlap, but with conditional inclusion probabilities that minimises the extent of overlap. This method, due to Philip Bell (Australian Bureau of Statistics), selects units for a survey with probability conditional on their selection in earlier surveys. The target unconditional probability is preserved, however. To simplify the calculation of the required conditional probabilities for additional surveys, the system saves the PSU selection histories. This method can also be modified to handle reformation of the geographic areas in the sampling frame. For more information, email Vic Duoba: vic.duoba@stats.govt.nz.

Significant cross-agency initiative integrates several important datasets
Statistics NZ's *Integrated Data Infrastructure* (IDI) allows for statistical outputs and research on the transitions and outcomes of people through education, labour market, benefits, justice, health and safety, migration, and business data.
The IDI is primarily based on administrative data and also contains a number of surveys undertaken by Statistics NZ and other agencies. Our Statistical Methods team has been involved in integrating a number of datasets into the IDI, for example, Statistics NZ's Survey of Family Income and Employment data, Ministry of Education secondary school achievement and tertiary education data, Inland Revenue data, and Ministry of Justice charges data. We applied probabilistic matching techniques when integrating these datasets.

Better outcomes for New Zealanders
This is the goal of the *Analysis for Outcomes* (AFO) programme Statistics NZ is working on together with a range of other government organisations and departments. Analysis for Outcomes will enable government agencies to identify new opportunities and improve value for money from initiatives and services by better leveraging the public sector's administrative data.

In August 2013, Cabinet confirmed its decision to establish a central agency Analytics and Insights team and expand Statistics NZ's IDI to deliver the AFO initiative. What this means is that there will be a significant investment in expanding the data services Statistics NZ provides through the IDI, in particular, providing more services to the core public sector as the platform to enable data-sharing for matching, anonymising, and accessing person-centred data.

This work has the potential to make a real difference to New Zealanders. To ensure public services are focused in the right place, and deliver the best outcomes, the State sector needs to understand how New Zealanders use these services and the outcomes they get, and where the opportunities exist to improve both services and outcomes.

Read the Analysis for Outcomes paper.

For more information on the IDI, email Lynsey Hayes: info@stats.govt.nz.

Upcoming Conferences and Workshops

2014 Joint Statistical Meetings (JSM)

Date: 2 – 7 August 2014
Venue: Boston Convention and Exhibition Center, Massachusetts, USA
Homepage: http://www.amstat.org/meetings/jsm/2014/

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America.

Attended by more than 6,000 people, meeting activities include oral presentations, panel sessions, poster presentations, continuing education courses, an exhibit hall (with state-of-the-art statistical products and opportunities), career placement services, society and section business meetings, committee meetings, social activities and networking opportunities.

Key Dates

Jan 17, 2014	Computer Technology Workshop (CTW) proposals are due for consideration for the 2014 program
Dec 3, 2013 – Feb 3, 2014	Online submission of abstracts (invited poster, regular and topic contributed abstracts)
Mar 31 - Apr 17, 2014	Online Abstract Editing Open
May 1, 2014	Registration & Housing Open
May 12, 2014	Draft Manuscript Deadline
May 29, 2014	Early Registration Deadline
May 30 - Jul 1, 2014	Regular Registration (increased fees apply)
Jul 2, 2014	Housing Deadline
Jul 2 - Jul 17, 2014	Late Registration (increased fees apply)

Comments and suggestions are welcome at meetings@amstat.org.

The Royal Statistical Society 2014 International Conference

Organized by: Royal Statistical Society
Where: Sheffield, UK
When: 1 – 4 September 2014
Homepage: http://www.rssconference.org.uk/

The conference programme will include 4 plenary sessions with speakers including Peter Hall (University of Melbourne) and Lord Allan of Hallam (Director of Policy in Europe for Facebook), as well as around 30 invited sessions covering a broad range of topics.

The broad streams for the conference will be as follows:

- Bioinformatics, Genomics & Biostatistics
- Communication of statistical ideas (including data visualisation)
- Data science (including experimental design)
- Emerging topics (including big data)
- Environment and ecology
- Industry and commerce
- Medical, clinical trials and epidemiology
- Public sector and policy evaluation (including open data)
- Statistics in sport
- Statistical methods and theory

There will also be a stream of professional development workshops running through the conference.

The deadline for contributed talk abstracts is 30 March 2014 while submissions for posters will be accepted until 30 June.

The draft programme listing all invited sessions will be available during February 2014.

The 2014 IAOS Conference on Official Statistics

Organized by: General Statistics Office, Vietnam
Where: Pullman Danang Beach Resort, Da Nang, Vietnam
When: October 8 – 10, 2014
Homepage: http://iaos2014.gso.gov.vn/

Meeting the Needs of a Changing World

The conference will focus on how official statistics need to constantly adapt to meet the needs of a changing world.

Themes for the conference include:

- Keeping up with changing needs of users for easily accessible, relevant, reliable information
- Exploiting new technology, especially use of the Internet for data collection and dissemination
- Coping with shrinking budgets
- Dealing with changing expectations on access to detailed data while protecting respondents' confidentiality
- Collaborating with other sectors, including education, to improve communication with users
- Developing new sources of information, such as administrative records, for producing official statistics

The conference will be hosted by the General Statistics Office, Vietnam. Please contact iaos2014@gso.gov.vn for more information.

2014 International Methodology Symposium

Statistics Canada
October 29-31, 2014
Palais des congrès de Gatineau,
Quebec, Canada

Beyond Traditional Survey Taking: Adapting to a Changing World

Call for Contributed Papers

Statistics Canada's 2014 International Methodology Symposium will take place at the Palais des congrès de Gatineau (5 minutes from downtown Ottawa) from **October 29 to 31, 2014**.

The theme of the Symposium is **"Beyond Traditional Survey Taking: Adapting to a Changing World"**. All members of the statistical community are invited to attend, whether they work in private research organizations, government or universities, particularly if they have an interest in methodological challenges resulting from the use of non-traditional survey methods.

The Symposium will include both plenary and parallel sessions that cover a wide variety of topics. Additional research and results may be presented via poster sessions.

We are soliciting contributed papers that examine **methodological issues resulting from the use of non-traditional survey methods**. Topics may include the following:

- Big data
- Record linkage
- Administrative data
- Web (panel) surveys
- Psychology of respondents
- Optimization of data collection
- Mode effects
- Electronic questionnaire
- Model-based approach
- Balanced sampling
- Microsimulations
- Time series
- Measurement errors
- Total survey errors
- Confidentiality

Please submit your proposal by email to symposium2014@statcan.gc.ca by **March 21, 2014**. It must include the following: title, an abstract (in English or French) of approximately 250 words, three to six keywords and your full contact information.

We will contact you before **May 9, 2014** to inform you whether your proposal was accepted. For all accepted proposals, you must submit the final version of your presentation (in English or French) by **September 8, 2014**. Proceedings from the conference will be published and distributed to participants. You must submit your final paper to us by **December 19, 2014**.

Please visit our Web site regularly in order to get more detailed and updated information: http://www.statcan.gc.ca/conferences/symposium2014/index-eng.htm

Statistique Canada Statistics Canada Canada

10th International Conference on Transport Survey Methods (ISCSTC10)

Organized by: The International Steering Committee Travel Survey Conference series
Where: Leura, Australia
When: 16 – 21 November 2014
Homepage: http://www.regodirect.com.au/isctsc10/home/

Embracing technological and behavioral changes

The International Steering Committee for Travel Survey Conferences (ISCTSC) organizes conferences with the aim of offering transport professionals (researchers, practitioners, modelers, planners, and others) the possibility to present their work, exchange information, network, promote international collaboration, and serve as a forum for the presentation of workshops, papers, and posters.

Rapidly evolving problems and policy contexts are compelling us to advance the state-of-the-art of survey methods, tools, strategies and protocols, while assuring the stability and coherence of the data from which trends can be tracked and understood. This year's conference will build on outputs from previous conferences and try to address these emerging research issues.

It will place particular emphasis on two related elements; new technologies, and decision and behavioral processes.

Important Dates

a. 1 November 2013: Abstracts due
b. 23 January 2014: Notification of abstract acceptance
c. 1 February 2014: Registration Opens
d. 15 September 2014: Full papers due
e. 31 July 2014: Close of Early-bird Registration
f. November 16-21, 2014: ISCTSC 10th International Conference on Transport Survey Methods

In Other Journals

Journal of Survey Statistics and Methodology

VOL 1, ISSUE 2 (2013)
http://jssam.oxfordjournals.org/content/current

Editors' Note
J. Sedransk, R. Tourangeau

Summary Report of the AAPOR Task Force on Non-probability Sampling
R. Baker, J. M. Brick, N. A. Bates, M. Battaglia, M. P. Couper, J. A. Dever, K. J. Gile, and R. Tourangeau

Comments
R. Valliant
D. Rivers
C. A. Gotway Crawford
G. Terhanian
G. Langer

Rejoinder
R. Baker, J. M. Brick, N. A. Bates, M. Battaglia, M. P. Couper, J. A. Dever, K. J. Gile, and R. Tourangeau

On Sampling Design Issues when Dealing with Zeros
B. King and A. Madansky

Encouraging Record Use for Financial Asset Questions in a Web Survey
M. P. Couper, M. B. Ofstedal, and S.Lee

Alternative Sequential Mixed-Mode Designs: Effects on Attrition Rates, Attrition Bias, and Costs
P. Lynn

Survey Methodology

Statistics Canada Statistique Canada

Canada

JUNE 2013, VOL 39, NO 1

http://www.statcan.gc.ca/pub/12-001-x/12-001-x2012002-eng.htm

Objective stepwise Bayes weights in survey sampling
J. Strief and G. Meeden

Optimizing quality of response through adaptive survey designs
B. Schouten, M. Calinescu and A. Luiten

Automatic editing with hard and soft edits
S. Scholtus

Sparse and efficient replication variance estimation for complex surveys
J. K. Kim and C. Wu

Estimation of the variance of cross-sectional indicators for the SILC survey in Switzerland
A. Massiani

Combining cohorts in longitudinal surveys
I.A. Carrillo and A. F. Karr

Indirect sampling applied to skewed populations
P. Lavallée and S. Labelle-Blanchet

On the performance of self benchmarked small area estimators under the Fay-Herriot area level model
Y.You, J.N.K. Rao and M. Hidiroglou

Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities
P. M. Aronow and C. Samii

Journal of Official Statistics

Dec 2013, Vol. 29 No. 4
http://www.jos.nu/entry.asp

Selective Editing: A Quest for Efficiency and Data Quality
T. de Waal

An Optimization Approach to Selective Editing
I. Arbués, P. Revilla and D. Salgado

Automated and Manual Data Editing: A View on Process Design and Methodology
J. Pannekoek, S. Scholtus and M. Van der Loo

A Contamination Model for Selective Editing
M. Di Zio and U. Guarnera

Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey
P. Lundquist and C.-E. Särndal

Utilising Expert Opinion to Improve the Measurement of International Migration in Europe
A. Wiśniowski, J. Bijak, S. Christiansen, J. J. Forster, N. Keilman, J. Raymer and P. W. F. Smith

Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time
A. Mohorko, E. de Leeuw and J. Hox

Survey Practice

Practical Information for Survey Researcher

VOL 6, NO 4 (2013)
www.surveypractice.org

The ABS Frame: Quality and Considerations
S. B. Roth, D. Han and J.M. Montaquila

Is everyone able to use a smartphone in survey research?
H. Fernee and N.Sonck

Measuring Compliance in Mobile Longitudinal Repeated-Measures Design Study
M. Link

Can Mobile Web Surveys Be Taken on Computers? A Discussion on a Multi-Device Survey Design
M. de Bruijne and A. Wijnant

The Efficiency of Conducting Concurrent Cognitive Interviewing and Usability Testing on an Interviewer-Administered Survey
J. R. Bergstrom, J. H. Childs, E. Olmsted-Hawala and N. Jurgenson

Adapting and Improving Methods to Manage Cognitive Pretesting of Multilingual Survey Instruments
M. Sha and Y. Pan

SRM Survey Research Methods

Vol 7, No 3 (2013)
http://w4.ub.uni-konstanz.de/srm/issue/current

Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys
M. P. Couper

Relative Mode Effects on Data Quality in Mixed-Mode Surveys by an Instrumental Variable
J. T.A. Vannieuwenhuyze and Melanie Revilla

Employed or inactive? Cross-national differences in coding parental leave beneficiaries in European Labour Force Survey data
M. Mikucka and M. Valentova

Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability
N. Kamoen, B. Holleman, H. van den Bergh and T. Sanders

Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference?
A. Mavletova and M. P. Couper

An Evaluation of Incentive Experiments in a Two-Phase Address-Based Sample Mail Survey
D.Han, J. M. Montaquila and J. M. Brick

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

VOL 29, NO 4 (2013)
http://iospress.metapress.com/content/mq540l587097/?p=4fbb2636830946728e5b0a963350fc6d&pi=0

Editorial

IAOS General Assembly August 2013 Report of the President for 2011 to 2013

IAOS Presidential Address by Shigeru Kawasaki

A historical appraisal of Greek survey sampling practices
C. Michalopoulou

Communicating central banking statistics: Making useful sense of statistics in a dynamic world
P.Nymand-Andersen

Spatial harmonization of economic cycles: Statistical confirmation of European-Russian interaction in real sectors of the economy
E. Zarova

International access to restricted data: A principles-based standards approach
F. Ritchie

How to fulfil user needs – from industrial production of statistics to production of knowledge
L. Thygesen and M. Grosen Nielsen

Fuzzy database queries in official statistics: Perspective of using linguistic terms in query conditions
M. Hudec

International Statistical Review
Revue Internationale de Statistique

DECEMBER 2013, VOL 81, ISSUE 3
http://onlinelibrary.wiley.com/doi/10.1111/insr.v81.3/issuetoc

New Book Review Editor for the International Statistical Review
M. Hallin and V.Nair

A Conversation with Colin L. Mallows
L. Denby and J. Landwehr

A Survey of L_1 Regression
D. Vidaurre, C. Bielza and P. Larrañaga

Proportional Odds Models with High-Dimensional Data Structure
F. Maqbool Zahid and G. Tutz

The Modern Census: Evolution, Examples and Evaluation
B. Baffour, T. King and P. Valente

A Summary of Attack Methods and Confidentiality Protection Measures for Fully Automated Remote Analysis Systems
C. M. O'Keefe and J. O. Chipperfield

STATISTICS IN TRANSITION
new series

An International Journal of the Polish Statistical Association

---

VOL 13, NO 3
http://www.stat.gov.pl/pts/15_ENG_HTML.htm

A class of product-type exponential estimators of the population mean in simple random sampling scheme
A. C. Onyeka

The class of estimators of finite population mean using incomplete multi-auxiliary information
M. Srivastava and N. Garg

Probability sample selection method in household surveys when current data on regional population is unavailable
T. Ünalan and H. O. Ayhan

Sampling designs proportionate to sum of two order statistics of auxiliary variable
J. L. Wywiał

The effect of unemployment benefits on labour market behaviour in Luxembourg
N. T. Longford and I. C. Salagean

Households' saving mobility in Poland
B. Liberda and M. Pęczkowski

Coherence and comparability as criteria of quality assessment in business statistics
A. Młodak

Statistics as a profession – statistician as an occupation: observations and comments from a panel of experts
W. Okrasa and B. Witek

The regional statistics – current situation and fundamental challenges
T. Borys

The 22nd Didactic Conference on Teaching Quality Evaluation Methods
A. Kupis-Fijałkowska

Summer School of Baltic-Nordic-Ukrainian Network on Survey Statistics 2013
M. Liberts

The International Year of Statistics/Statistics 2013
J. Witkowski

VOL 5, ISSUE 1 (2013)
http://repository.cmu.edu/jpc/

Statistical Dèjá Vu: The National Data Center Proposal of 1965 and Its Descendants
R. Kraus

Privacy via the Johnson-Lindenstrauss Transform
K. Kenthapadi, A. Korolova, I. Mironov, and N. Mishra

Introduction to Special Section
D. Kifer

Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation
J.M. Abowd, M. J. Schneider, and L. Vilhuber

On Regression-Tree-Based Synthetic Data Methods for Business Data
J.H. Lee, I. Y. Kim, and C. M. O'Keefe

Privacy-Preserving Data Sharing for Genome-Wide Association Studies
C. Uhler, A. B. Slavkovic, and S. E. Fienberg

TRANSACTIONS ON DATA PRIVACY

Foundations and Technologies
http://www.tdp.cat

August 2013, VOL 6, ISSUE 2
http://www.tdp.cat/issues11/vol06n02.php

Preserving Differential Privacy in Degree-Correlation based Graph Generation
Y. Wang and X. Wu

Analysis of Commercial and Free and Open Source Solvers for the Cell Suppression Problem
B. Meindl and M. Templ

On Syntactic Anonymity and Differential Privacy
C. Clifton and T. Tassa

Journal of the Royal Statistical Society SERIES A Statistics in Society

October 2013, VOL 176, ISSUE 4
http://onlinelibrary.wiley.com/doi/10.1111/rssa.2013.176.issue-4/issuetoc

Editorial: Counting the dead properly and promptly
S. M. Bird

Statistics making an impact
J. Pullinger

Handedness, health and cognitive development: evidence from children in the National Longitudinal Survey of Youth
D. W. Johnston, M. E. R. Nicholls, M. Shah and M. A. Shields

Does high involvement management lead to higher pay?
P. Böckerman, A. Bryson and P. Ilmakunnas

Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes
J. Jerrim and A. Vignoles

Assortative mating and divorce: evidence from Austrian register data
W. Frimmel, M. Halla and R. Winter-Ebmer

Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence
D. J. Graham, E. J. McCoy and D. A. Stephens

Scarring effects of remaining unemployed for long-term unemployed school-leavers
B. Cockx and M. Picchio

Journal of the American Statistical Association

VOL 108, ISSUE 504 (2013)
http://amstat.tandfonline.com/toc/uasa20/current

The International Year of Statistics: A Celebration and A Call to Action
M. Davidian

Statistical Learning With Time Series Dependence: An Application to Scoring Sleep in Mice
B. B. McShane, S. T. Jensen, A. I. Pack and A. J. Wyner

Comment
K. Shedden
D. Zeng and Y. Wang

Rejoinder
B. B. McShane, S. T. Jensen, A. I. Pack and A. J. Wyner

Optimal Allocation of Gold Standard Testing Under Constrained Availability: A pplication to Assessment of HIV Treatment Failure
T. Liu, J. W. Hogan, L. Wang, S. Zhang and R. Kantor

Semiparametric Bayesian Estimation for Marginal Parametric Potential Outcome Modeling: Application to Causal Inference
T. Hoshino

Frailty Models for Familial Risk With Application to Breast Cancer
M. Gorfine , L Hsu and G. Parmigiani

A Marginal Approach to Reduced-Rank Penalized Spline Smoothing With Application to Multilevel Functional Data
H. Chen, Y. Wang, M. C. Paik and H. A. Choi

Bayesian Hierarchical Modeling of the HIV Evolutionary Response to Therapy
S. T. Jensen, J. Park, A. F. Braunstein and J. Mcauliffe

Generalized Jackknife Estimators of Weighted Average Derivatives
M. D. Cattaneo, R. K. Crump and M. Jansson

Comment
D. Zeng
H. Dette
E. Mammen
X. Chen

Rejoinder
M. D. Cattaneo, R. K. Crump and M. Jansson

Classification Using Censored Functional Data
A. Delaigle and P. Hall

Selecting the Number of Principal Components in Functional Data
Y. Li, N. Wang and R. J. Carroll

Sampling for Conditional Inference on Network Data
J. Zhang and Y. Chen

Ensemble Subsampling for Imbalanced Multivariate Two-Sample Tests
L. Chen, W. W. Dou and Z. Qiao

Bayesian Modeling of Temporal Dependence in Large Sparse Contingency Tables
T. Kunihama and D.B. Dunson

Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables
N. G. Polson, J. G. Scott and J. Windle

Two-Stage Importance Sampling With Mixture Proposals
W. Li, Z. Tan and R. Chen

Epistatic Clustering: A Model-Based Approach for Identifying Links Between Clusters
J. Zhang

Multiple Testing in a Two-Stage Adaptive Design With Combination Tests Controlling FDR
S. K. Sarkar, J. Chen and W. Guo

Multivariate Density Estimation by Bayesian Sequential Partitioning
L. Lu, H. Jiang and W.H. Wong

On Optimal Designs for Nonlinear Models: A General and Efficient Algorithm
M. Yang, S. Biedermann and E. Tang

Local Linear Regression on Manifolds and Its Geometric Interpretation
M.-Y. Cheng and H.-T. Wu

Generalized Residuals for General Models for Contingency Tables With Application to Item Response Theory
S. J. Haberman and S. Sinharay

Locally Adaptive Bayes Nonparametric Regression via Nested Gaussian Processes
B. Zhu and D. B. Dunson

Case Definition and Design Sensitivity
D. S. Small, J. Cheng, M. E. Halloran and P. R. Rosenbaum

Hidden Markov Models With Applications in Cell Adhesion Experiments
Y. Hung, Yijie Wang, V. Zarnitsyna, C. Zhu and C. F. Jeff Wu

Discovering Findings That Replicate From a Primary Study of High Dimension to a Follow-Up Study
M. Bogomolov and R. Heller

Adaptive Testing of Conditional Association Through Recursive Mixture Modeling
Li Ma

A Progressive Block Empirical Likelihood Method for Time Series
Y. M. Kim, S. N. Lahiri and D. J. Nordman

Cure Rate Quantile Regression for Censored Data With a Survival Fraction
Y. Wu and G. Yin

Resampling Procedures for Making Inference Under Nested Case–Control Studies
T. Cai and Y. Zheng

BIOMETRIKA

June 2013, VOL 100, ISSUE 4
http://biomet.oxfordjournals.org/content/current

Bridging the ensemble Kalman and particle filters
M. Frei and H. R. Künsch

Nonparametric Bayes modelling of count processes
A. Canale and D. B. Dunson

Practical perfect sampling using composite bounding chains: the Dirichlet-multinomial model
N. M. Stein and X.-L. Meng

Two-sample rank tests under complex sampling
T. Lumley and A. J. Scott

Controlling the bias of robust small-area estimators
V. Dongmo Jiongo, D. Haziza, and P. Duchesne

Variable selection in semiparametric transformation models for right-censored data
X. Liu and D. Zeng

Semiparametric estimation for the additive hazards model with left-truncated and right-censored data
C.-Y. Huang and J. Qin

Composite quantile regression for the receiver operating characteristic curve
X. Duan and X.-H. Zhou

Reduced rank regression via adaptive nuclear norm penalization
K. Chen, H. Dong, and K.-S. Chan

High-dimensional volatility matrix estimation via wavelets and thresholding
P. Fryzlewicz

Scaled envelopes: scale-invariant and efficient estimation in multivariate linear regression
R. Dennis Cook and Z. Su

Smoothing splines with varying smoothing parameter
X. Wang, P.Du, and J. Shen

Saddlepoint approximations for the normalizing constant of Fisher–Bingham distributions on products of spheres and Stiefel manifolds
A. Kume, S. P. Preston, and A. T. A. Wood

Simultaneous confidence intervals that are compatible with closed testing in adaptive designs
D. Magirr, T. Jaki, M. Posch, and F. Klinglmueller

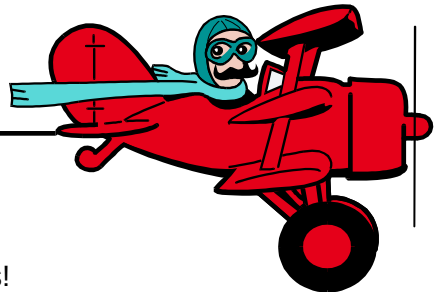
Designs for crossvalidating approximation models
Q. Zhang and P. Z. G. Qian

A simple test for random effects in regression models
S. N. Wood

Posterior consistency in linear models under shrinkage priors
A. Armagan, D. B. Dunson, J. Lee, W. U. Bajwa, and N. Strawn

Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom
E. Susko

Welcome New Members!



We are very pleased to welcome the following new members!

Country	Name	Surname
Ethiopia	Aberash Tariku	Abaye
Nigeria	Boniface O.	Amobi
United Kingdom	Yves	Berger
Israel	Tom	Caplan
Nigeria	Tukur	Dahiru
South Africa	Reza C.	Daniels
Russian Federation	Oxana	Kuchmaeva
Russian Federation	Natalia	Markovich
South Africa	Besa	Muwele
Nigeria	Isiaka	Olarewaju
Nigeria	Olugbemi Alabi	Olujimi
United States of America	Beth-Ellen	Pennell

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

Committee Chairs

Chair of the Rio 2015 Programme Committee	Christine Bycroft (New Zealand)	christine.bycroft@stats.govt.nz

The Secretariat

Executive Director:	Catherine Meunier (France)	catherine.meunier@insee.fr
Treasurer:	Ada van Krimpen (The Netherlands)	an.vankrimpen@cbs.nl
Webmaster:	Justin Lokhorst (Australia)	justin.lokhorst@abs.gov.au
IASS Secretariat Membership Officer	Margaret de Ruiter-Molloy (The Netherlands)	m.deruitermolloy@cbs.nl

Institutional Members

2 International Organisations

AFRISTAT
EUROSTAT

15 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
CANADA – STATISTICS CANADA
CHINA – DIREÇCÃO DOS SERVIÇOS DE ESTATÍSTICA E CENSOS
DENMARK –DANMARKS STATISTIK
FINLAND – STATISTICS FINLAND
GERMANY –STATISTICHE BUNDESAMT
ITALY –INSTITUTO NAZIONALE DI STATISTICA
KOREA, REPUBLIC OF – STATISTICS KOREA
MEXICO –INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI)
MAURITIUS – STATISTICS MAURITIUS
NEW ZEALAND – STATISTICS NEW ZEALAND
NORWAY – STATISTICS NORWAY
PORTUGAL –INSTITUTO NACIONAL DE ESTADÍSTICA (INE)
SWEDEN – STATISTISKA CENTRALBYRÅN

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 INC,

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 Ruiter-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.cbs.nl/iass/index.htm