

the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 65

January 2012



International Statistical Institute



Institut International de Statistique

the Survey Statistician

The Newsletter of the International Association of Survey Statisticians
No. 65 January 2012



Editors

Natalie Shlomo and Frank Yu

Section Editors

Pierre Lavallée — Country Reports
John Eltinge — Book/Software Review
Robert Clark — Ask the Experts
Leyla Mohadjer — New and Emerging Methods

Circulation/Production

Jennifer Marley **Hans Lucas**
Henry Chiem

The Survey Statistician is published twice a year in English and French 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

No. 65, January 2012

- 3 Letter from the President
- 5 Letter from the Editors
- 7 Report from the Scientific Secretary
- 9 Contact Information for IASS Country Representatives
- 11 News and Announcements
- 14 Ask the Experts:
Feedback of survey information for updating of sampling frames
- 17 Country Reports
 - ◆ Canada
 - ◆ France
 - ◆ India
 - ◆ New Zealand
 - ◆ Poland
 - ◆ United States of America
- 24 New and Emerging Methods:
Evaluating, comparing, monitoring and improving representative awareness of Survey Response
- 31 Book and Software Review:
Study of Record Linkage Software for the 2010 Brazilian Census Post Enumeration Survey
- 40 Upcoming Conferences and Workshops
- 50 In Other Journals
- 63 Welcome New Members
- 65 IASS Officers and Council Members
- 66 Institutional Members
- 67 Change of Address Form

Letter from the President

Dear Colleagues

This is my first 'Letter from the President' and I must admit to being quite nervous as I write it. I took over from Susan Linacre at the IASS General Assembly at the WSC in Dublin in late August, and I am still very much learning on the job. That I have the luxury of still being in a 'learning' phase at this point is in no small part due to the excellent job that Susan did in her term as President of the IASS. Without her wise counsel, and the projects that she set in train, I would be a rather more 'unrelaxed' person than I am at the moment.

The year 2012 is set to be one of change for IASS. The most important new initiative, and the one that was discussed the most in Dublin, is the move of the Association's administration from the office of INSEE in Libourne, France, to the ISI office in The Hague, Holland. This move is necessitated by the decision of INSEE that they can no longer afford the resources to support the IASS administration. This is quite understandable given that INSEE has provided these resources, free of charge, to IASS since its inception, and all good things come to an end. On behalf of the members of IASS, I would therefore like to again express our gratitude to INSEE for its support over the years. Over the last couple of months, Catherine Meunier has been working hard to transfer the membership database over to the ISI office, and members will shortly receive their renewal notices via the ISI rather than from Catherine as in the past. There are quite a few hurdles, however, that still need to be overcome to complete this move. Steve Heeringa is currently working on a redraft of the IASS statutes that will specifically identify the Association as being based in the ISI office. These revised statutes will be voted on at the IASS General Assembly at the World Statistics Congress in Hong Kong in 2013. There are still legal issues that need to be sorted out however, since the Association is registered in France and therefore needs to be wound up in France before it can be 'reconstituted' in Holland. We have asked the ISI for advice on the best way this can be done and will keep you informed as things progress. The key person 'on the ground' arranging these changes is the IASS Executive Director, Catherine Meunier, and I would like to thank her on behalf of the IASS for her hard work on this issue over the last few months. It is very nice to know that Catherine will be keeping an eye on this transition over the next two years!

Of course, the most immediate impact of the move is that we now have to pay ISI for the administrative services that were previously provided free of charge by INSEE. This has meant that we have had to increase IASS membership fees from 2012. The increase is the minimum possible that can cover the anticipated increase in our costs, and we will need to see to keep a strict eye on our finances over the next two years leading into the next General Assembly. At this stage we intend to revisit the issue of fees then. In the meantime however, we have identified that an increase in the number of institutional members of the Association will go a long way towards offsetting our increased administration costs. Mick Couper has kindly volunteered to lead an Institutional Member Drive that will look at how this increase can be achieved. I am sure that Mick will be very grateful to receive any suggestions from IASS members of institutions involved in survey sampling that would benefit from membership of the Association. Offers of help in approaching them will be most appreciated, since it should come as no surprise that it is our members 'on the ground' that have the connections necessary to open the doors of these institutions!

The IASS website is also about to be refurbished. Our new webmaster, Justin Lokhorst, has been developing the new site over the last few months in concert

with Olivier Dupriez and Catherine Meunier. Hopefully, the new site will come 'on line' early in the New Year and will offer a considerable increase in functionality over the current site which has remained fairly static for a few years now. Watch this space.

One of the key activities of the IASS is to provide support to international statistical meetings that have a sample survey focus. Responsibility for coordinating requests for assistance and disbursement of funds for this purpose used to rest with Peter Lynn, and now rests with Steve Heeringa. In this context, IASS has recently provided financial support to:

2011

1. Italian Conference on Survey Methodology ITACOSM11
2. Second International Workshop in Applied Statistics, Bogota
3. Workshop on Using Multi-Level Data from Sample Frames, NORC
4. Third Baltic-Nordic Conference on Survey Statistics
5. International Workshop, 3rd Brazilian School on Sampling

2012

1. International Conference on Establishment Surveys ICES-IV, Montreal
2. ASA Conference on Sampling Hard to Reach Populations
3. 7th Francophone Conference on Sampling

Another important activity of the IASS is the award of the Cochran-Hansen prize at the meetings of the ISI every two years. This prize is extremely important for an organisation like IASS because it is one of the few ways that we can reach out to the many bright young people in the developing world who are moving into our profession, and to honour the best of them. I would like to think that all members of IASS who are in a position to encourage nominations for this prize will do so. The Cochran-Hansen Prize for 2011 was won by Dr Solange Correa of the Brazilian Institute of Geography and Statistics, and I would like to thank Denise Britz for her efforts to encourage nominations from the Latin American countries. Yves Tille is Chair of the Cochran-Hansen Prize Jury for the 2013 award, and I am sure that he will be delighted to receive nominations for this award from all over the developing world.

The other important IASS activity tied to the biennial ISI meetings are the short courses that we run in order to build up the methodological skills of sample survey professionals attending the meeting. The IASS short courses at the Dublin World Statistics Congress were superbly organised and coordinated by our previous Scientific Secretary, Leyla Mohadjer. At this stage we are still waiting to hear what format the short course programme at the 2013 Hong Kong WSC will take. However, we do know that the focus of the meeting will be Youth and a key aspect of the short course programme will be resource-building, so it seems clear that there will be significant changes from the way short courses have been traditionally run at ISI meetings. Ineke Stoop, our new Scientific Secretary will be coordinating the IASS component of the short course programme in Hong Kong and I am sure that she will keep us up to speed on developments on this front.

It only remains for me to thank all the new and continuing members of the IASS Council, as well as the new members of the IASS Executive for all the help and support that they have provided over 2011.

A Happy Christmas and a Peaceful New Year to all.

Ray Chambers
IASS President



Letter from the Editors

The January 2012 issue of the IASS newsletter *The Survey Statistician* contains articles of interest and important information regarding upcoming conferences, journal contents, updates from the IASS Executive and more. We would like to ask all IASS members to please take an active role in supporting the 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. Your active role in contributing to the IASS newsletter will ensure that it remains relevant to our membership and is interesting and enjoyable to read.

So what is in the January 2012 issue of the IASS newsletter?

In the New and Emerging Methods Section (edited by Leyla Mohadjer and Andrea Piesse), Barry Schouten and his co-authors have contributed an article on 'Evaluating, comparing, monitoring and improving representativeness of survey response through R-indicators and partial R-indicators', where they provide a conceptual framework for representative survey response which can be used for survey data collection monitoring and tailoring to reduce non-response error.

In the Ask the Experts section, Robert Clark discusses the issue of whether or not "dead units" discovered in the survey process could be fed back to the sampling frame to improve the quality of the frame for future surveys. The discussion includes sections from Hedlin and Wang (2004) with permission by Hedlin.

We are also very pleased to have a contributing article for the Book and Software Review Section by Andrea Diniz da Silva and co-authors on the 'Study of Record Linkage Software for the 2010 Brazilian Census Post Enumeration Survey'. The team at IBGE Brazil present their experiences evaluating different record linkage software packages to be used in the data linkage between the 2010 Brazilian Census and the Post-Enumeration Survey.

We wish to thank all of the authors and editors of these sections for their important contributions on behalf of the IASS membership. Please let Ineke Stoop (i.stoop@scp.nl) 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 your questions to Robert Clark (rclark@uow.edu.au). If you are interested in writing a book review or software review, please get in touch with the editor of the section, John Eltinge (eltinge.john@bls.gov).

As in the past, this issue of *The Survey Statistician* includes a letter from the current President, Ray Chambers and updates from the Scientific Secretary, Ineke Stoop and the Chair of the Hong Kong 2013 Programme Committee, Éric Rancourt.

The Country Report section has always been a central feature of the IASS newsletter and we thank the editor of the section, Pierre Lavalée

(pierre.lavallee@statcan.gc.ca) for his continuing efforts to get timely reports from the different countries. We ask all country representatives to please submit articles and share information on current activities, research and applications in survey methods in your countries. In particular, for the next issue of *The Survey Statistician* that will come out in July 2012, we would like to ask country representatives to also contribute a report specifically related to the recent round of Census. The topic can include any methodological aspect of the Census, for example, population questionnaire design, data capture, adjustments for census undercount, statistical disclosure control, etc. All articles should be sent to Pierre.

The IASS Vice-president, Denise Silva has taken on the responsibility of updating the list of country representatives. Please get in touch with her at denisebritz@gmail.com if you are currently serving as your country representative or alternatively can suggest a new representative for your country.

This issue of *The Survey Statistician* also includes the Tables of Contents from recent journals and advertisements for upcoming conferences and workshops. If you would like to advertise in *The Survey Statistician*, please send your adverts to the editors: frank.yu@abs.gov.au and n.shlomo@soton.ac.uk.

Finally, we wish to send our warmest congratulations to the 2011 Waksberg Award winner, Danny Pfeffermann, who gave a keynote lecture at the recent Statistics Canada Methodology Symposium. In addition, we wish to congratulate the future Waksberg Award winner for 2012, Lars Lyberg.

As always, we have many thanks for everyone working hard to put this newsletter together and in particular Jennifer Marley and Henry Chiem of the Australian Bureau of Statistics for their invaluable assistance. We wish to thank the Section Editors, Ray Chambers for his continuing support, the IASS Executive Director, Catherine Meunier and Rolande Charette for translations into French.

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

We hope you enjoy this January 2012 issue of *The Survey Statistician*. Please send us your feedback and comments on how we can make improvements.



Report from the Scientific Secretary

More than Surveys and Guinness: The IASS at the ISI in Dublin

Surveys statisticians had a busy time at the World Statistics Conference in Dublin. Firstly they had to walk along the Liffey, enjoy Irish song and dance, try some Guinness and admire the impressive Convention Centre. Secondly, they could participate in one or more of the IASS short courses preceding the conference. Leyla Mohadjer, the scientific secretary of the IASS, provided a detailed report. Thirdly, they could attend the IASS General Assembly meeting during which long-standing council members (among whom Leyla) stepped down and new ones were introduced. It is clear that it will be difficult to succeed a team that was so successful. At the General Assembly meeting a number of important topics were introduced and discussed, among which were the steps being taken to modernize the IASS website, the short courses programme in Hong Kong in 2013, and the move of the IASS office from the INSEE office at Libourne in France to the ISI office in The Hague. Luckily, Cathérine Meunier will continue to support the IASS as Executive Director while this important move takes place, with the most immediate consequence being that the ISI office will be dealing with IASS subscriptions from 2012.

But the main activity at the ISI in Dublin was of course the invited and contributed paper sessions. IASS organized or co-organized 13 Invited Paper Sessions. Some of these focused on specific sectors (healthcare, agriculture and environment). Others were devoted to methodological issues such as complex analysis, inference and estimation. There were sessions on the Census, small-area estimation, confidentiality, data mining and paradata. But here too the focus was on more than surveys. In particular, linking administrative files to survey data is becoming increasingly important, as was shown by sessions on “Legal, ethical and privacy issues arise out of increasing use of linked administrative files”, and “Record linking and imputation in administrative data”. Finally, there was the session where the winner of the 2011 Cochran-Hansen prize, Dr. Solange Correa Onel of the Brazilian Institute of Geography and Statistics, presented the ground-breaking research that won her the award. Many congratulations to Solange!

So we had a busy time, but we learned a lot, enjoyed ourselves a great deal, went home tired and inspired and looking forward to see each other again in Hong Kong!

Sponsorship Conferences

Overview of conferences sponsored by the IASS in 2011 and 2012

The objective of the IASS is to promote the study and development of the theory and practice of statistical censuses and surveys and associated subjects throughout the world. To pursue this objective, the IASS is able to provide limited financial support to relevant conferences and workshops. In addition the IASS is keen to promote such activities and related outputs.

IASS has supported the following conferences:

2011

- Second Italian Conference on Survey Methodology ITACOSM11 (Italy)
- Second International Workshop on Applied Statistics with Special Topics in Survey Sampling and Census Data (Columbia)
- International Workshop on Using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys (US)
- Third Baltic-Nordic Conference on Survey Statistics BaNoCoSS-2011 (Sweden)
- International Workshop, 3rd Brazilian School on Sampling III ESAMP (Brazil)

2012

- Fourth International Conference on Establishment Surveys (ICES IV) (Canada)
- International Conference for Surveying and Enumerating Hard to Reach Populations (H2R) (US)
- Seventh Francophone Conference on Sampling (sondages2012) (France)

Preparing for Hong Kong 2013

With Dublin 2011 just behind us the preparations for the World Statistics Conference in Hong Kong are already in full swing. In addition to preparing invited paper sessions, the IASS Board is responsible for short courses that meet the training and continuing education needs of the many young statisticians in the world. This will be done in cooperation with Hong Kong University. In the beginning of 2012, the solicitation process of initial proposals for short course topics will start.

Update on the IASS programme for the ISI World Statistics Congress in 2013

The IASS scientific program for the 2013 World Statistics Congress in Hong Kong is under progress. We have received more than 25 proposals for invited paper sessions. The ISI has given the association executives until the end of 2011 to come up with its final list of proposals. Then, in the first months of 2012, the ISI will work on the overall program and finalize it for April 2012.

I would like to thank all those who proposed sessions and in particular the members of the organizing committee: Martin Balepa (Mali/Afristat), Yves Berger (UK), Christine Bycroft (New Zealand), Takis Merkouris (Greece), Jean Opsomer (US), and Xiuhua Tian (China). Once our list of sessions is final, potential presenters will be contacted to solicit or confirm their participation and to submit titles and start working on their papers.

Should you have any question about the program, please do not hesitate to contact me (eric.rancourt@statcan.gc.ca).



Contact Information for IASS Country Representatives

<u>Name</u>	<u>Country</u>
Mr. Nacer-eddine Hammouda	Algeria
Ms. Alicia Masautis	Argentina
Mr. Paul Sutcliffe	Australia
Dr. Camille Vanderhoeft	Belgium
Mr. Walter Castillo Guerra	Bolivia
Mr. Moffat Malepa	Botswana
Mr. Cristiano Ferraz	Brazil
Mr. François Ilboudo	Burkina Faso
Mr. John Kovar	Canada
Ms. Maria de Lurdes Lopes	Cape Verde
Mr. Juan Eduardo Munoz	Chile
Mr. Huang Langhui	China
Mr. Mhadji Nailane	Comoros
Mr. Cakpo Benjamin Zanou	Côte d'Ivoire
Mr. Luis Carlos Silva	Cuba
Mr. Vaclav Cermak	Czech Republic
Mr. Peter Linde	Denmark
Dr. Imbi Traat	Estonia
Mr. Jean-Marc Museux	Eurostat (Europe)
Mr. Paavo Väisänen	Finland
Mr. Benoît Riandey	France
Mr. Jean-Pierre Zima Mefe	Gabon
Dr. Ralf Münnich	Germany
Mr. Anastassios Iliakopoulos	Greece
Mr. David Fitch	Guatemala
Ms. Fatoumata Danfaca	Guinea
Mr. Emmanuel Charles	Haiti
Mr. Zoltán Csereháti	Hungary
Dr. Gayatri V. Singh	India
Ms. Luisa Kadun Burck	Israel
Prof. Claudio Quintano	Italy
Dr. Ryozo Yoshino	Japan
Mr. Martins Liberts	Latvia
Prof. Bechara Hanna	Lebanon
Dr. Danutė Krapavickaitė	Lithuania
Mr. Antonio Baigorri Matamala	Luxembourg
Ms. Julia Rachel Ravelosoa	Madagascar
Dr. Suresh Chandra Babu	Malawi
Mr. Aziz Mohammad	Malaysia
Mr. Lamine Diop	Mali
Mr. Sidna Ould N'dah	Mauritania
Mr. Miguel Cervera	Mexico
Dr. Rudra Suwal	Nepal
Mr. Shyam Upadhyaya	Nepal
Mr. Jos de Ree	Netherlands
Ms. Diane Ramsey	New Zealand
Mr. John Créquer	New Zealand
Mr. Adetoun Aribiki Imolehin	Nigeria

Dr. Mark Griffin	Pacific Islands
Mr. Faisal Awartani	Palestine
Ms. Leonara Laguna	Peru
Mr. Gervacio G. Selda, Jr.	Philippines
Dr. Tomasz Zadlo	Poland
Mr. Paulo Jorge Gomes	Portugal
Mr. Alexis Lukaku Nzinga	D. R. Congo
Dr. Inho Park	Republic of Korea
Innocent Ngombe Bibemo	Central African Republic
Mr. Matar Gueye	Senegal
Ms. Dolores Lorca Lopez	Spain
Mrs. Anjeles Iztueta Azkue	Spain-Basque Region
Mr. Christina Prado Valle	Spain-Basque Region
Mr. Philippe Eichenberger	Switzerland
Mr. Peter Lundquist	Sweden
Dr. Ibrahim Ali	Syria
Mr. Vitalis Eustach Muba	Tanzania
M. Ouagadjio Bandoumal	Chad
Prof. Oztas Ayhan	Turkey
Dr. Olga Vasylyk	Ukraine
Prof. Peter Lynn	United Kingdom
Mr. Howard Hogan	United States
Mr. Federico Segui	Uruguay
Dr. Nguyen Quoc Anh	Vietnam

If you are interested in becoming your country representative or have any updates, amendments or changes to this list, please contact Denise Silva at denise.silva@ibge.gov.br.

News and Announcements

Waksberg Award

The journal *Survey Methodology* has established an annual invited paper series in honour of Joe Waksberg to recognize his contributions to survey methodology. Each year a prominent Survey Statistician is chosen to write a paper that reviews the development and current state of an important topic in the field of survey methodology. The paper reflects the mixture of theory and practice that characterized Joe Waksberg's work.

This year the Waksberg Award was given to Prof. Danny Pfeffermann of the Hebrew University and University of Southampton for his outstanding contributions to *Survey Methodology*. The topic of his paper is 'Weighting of Complex Survey Data: Why model? Why is it a problem? How can we handle it?' and will be published in *Survey Methodology*. Prof. Pfeffermann also gave the keynote Waksberg Invited Speaker Address at the Statistics Canada Methodology Symposium in November 2011.

In addition, we are happy to announce that the next recipient of the Waksberg Award for 2012 is Prof. Lars Lyberg of Statistics Sweden and Stockholm University. His paper will be on 'The Evolution of the Survey Quality Concept'.

Congratulations to both Danny and Lars on a well - deserved recognition of their contribution to Survey Methodology.

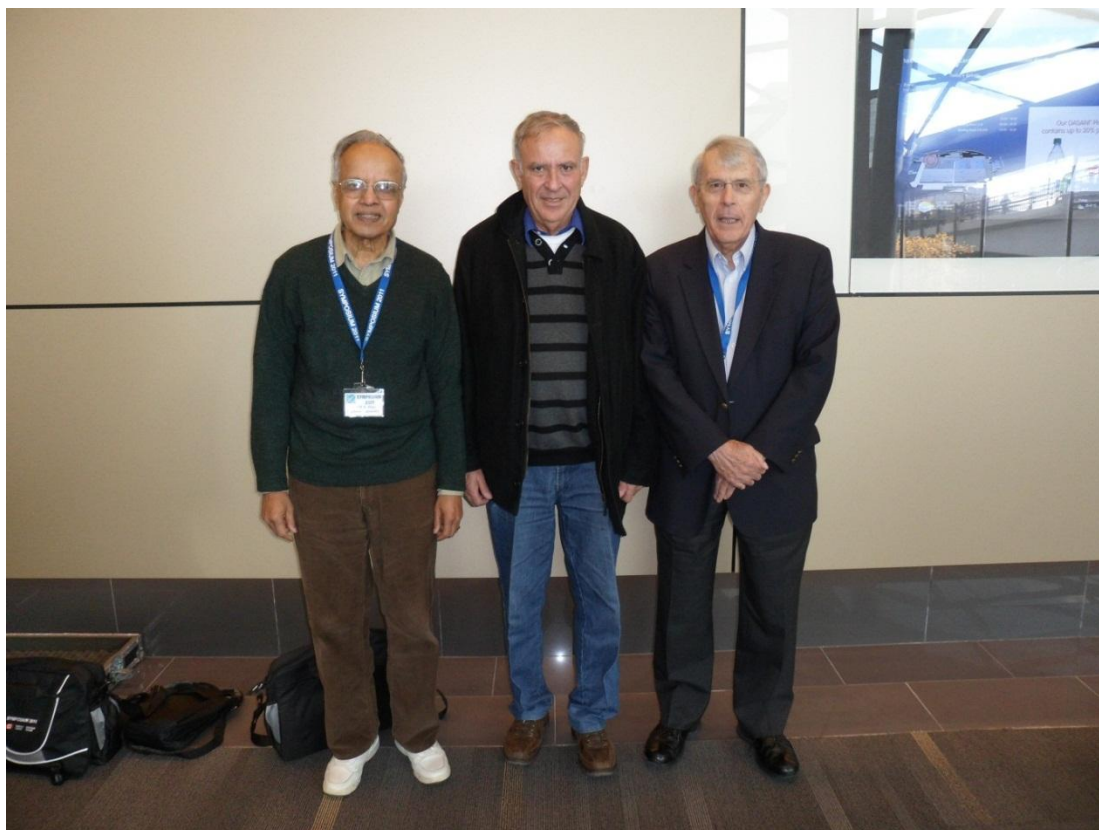
Nominations are now sought for the 2013 Waksberg Award. The author of the 2013 Waksberg paper will be selected by a four-person committee appointed by *Survey Methodology* and the American Statistical Association. **Nomination of individuals to be considered as authors or suggestions for topics should be sent before February 28, 2012 to the chair of the committee, Mary Thompson (methomps@uwaterloo.ca).**

Previous Waksberg Award Honorees and their invited papers are:

- 2001 Gad Nathan, "Telesurvey Methodologies for Household Surveys-A Review and Some Thoughts for the Future." *Survey Methodology*, vol. 27, no. 1, pp.7-31.
- 2002 Wayne Fuller, "Regression Estimation for Survey Samples." *Survey Methodology*, vol. 28, no. 1, pp. 5-23.
- 2003 David Holt, "Methodological Issues in the Development and Use of Statistical Indicators for International Comparisons." *Survey Methodology*, vol. 29, no. 1, pp. 5-17.
- 2004 Norman Bradburn, "Understanding the Question-Answer Process." *Survey Methodology*, vol. 30, no. 1, pp. 5-15.
- 2005 J.N.K Rao, "Interplay Between Sample Survey Theory and Practice: An Appraisal."

- Survey Methodology*, vol. 31, no. 2, pp. 117-138.
- 2006 Alastair Scott, "Population-Based Case Control Studies." *Survey Methodology*, vol. 32, no. 2, pp. 123-132.
- 2007 Carl-Erik Särndal, "The Calibration Approach in Survey Theory and Practice." *Survey Methodology*, vol. 33, no. 2, pp. 99-119.
- 2008 Mary Thompson, "International surveys: Motives and Methodologies." *Survey Methodology*, vol. 34, no. 2, pp. 131-141.
- 2009 Graham Kalton, "Methods for Oversampling Rare Subpopulations in Social Surveys." *Survey Methodology*, vol. 35, No. 2, pp. 125-141.
- 2010 Ivan Fellegi, "The organisation of statistical methodology and methodological research in national statistical offices." *Survey Methodology* vol. 36 no. 2., pp. 123-130.
- 2011 Danny Pfeffermann. "Modelling of complex survey data: Why is it a problem? How can we approach it?" *Survey Methodology*, to be published.
- 2012 Lars Lyberg, "The Evolution of the Survey Quality Concept." *Survey Methodology*, to be published.

**Photograph of Waksberg Award Honorees:
Jon Rao, Danny Pfeffermann and Graham Kalton**



Morris-Hansen Lecture Series

The Morris Hansen Lecture series was established by the Washington Statistical Society (WSS) in 1990 with a financial grant from Westat, Inc. It was set up to honor Morris Hansen, whose pioneering contributions to survey sampling and related statistical methods during his long and distinguished career at the U.S. Census Bureau and at Westat established many standards and methods, mostly still in use, for the conduct of surveys. In 1996 Westat made an additional financial contribution to WSS so that the Hansen Lecture series could continue.

This year, the 21st Annual Morris Hansen Lecture was held on Tuesday, October 18th, 2011 in the Jefferson Auditorium of the U.S. Department of Agriculture's (USDA's) South Building in Washington, DC. The distinguished speakers were Robert M. Groves, Ph.D., Director, U.S. Census Bureau, and Roderick J. Little, Ph.D., Associate Director for Statistical Methodology and Standards, U.S. Census Bureau. The lecture was titled *Total Survey Error: Missing Conceptual Components, and Design-Based/Model-Based Viewpoints*. The discussant was Natalie Shlomo, Ph.D., University of Southampton, United Kingdom.

The National Academy of Sciences has included a biography of Morris Hansen in its 1996 volume entitled "Biographical Memoirs" (vol. 70) published by the National Academy Press. The memoir of Morris Howard Hansen was co-authored by Joseph Waksberg and Edwin D. Goldfield and is available at: stills.nap.edu/html/biomems/mhansen.htm.



Ask the Experts

If information on dead units is obtained in a survey, should it be fed back to the sampling frame?

The answer below is contributed by the column editor, Robert Clark. Part of this article draws on information from Hedlin and Wang (2004), as suggested and permitted by Dan Hedlin.

When surveys (particularly of businesses) are conducted, it is often found that some of the sample units have ceased to exist. These are often referred to, somewhat over-dramatically, as “deaths”. This phenomenon usually occurs because businesses have merged or are no longer operating, even though they remain on the sampling frame. This is a common situation in business surveys operated by National Statistical Institutes, because frames are normally updated using information that is incomplete and out of date, and units are generally not removed from frames unless it is very clear that they are no longer operating.

The correct way of handling deaths amongst sample units is to retain these units in the calculation of estimation weights, but to assign them a zero value for all variables of interest. If simple random sampling without replacement is used, this would mean assigning a weight of (N/n) to all units in the sample, where N is the number of population units, and n is the number of sample units, with both N and n including deaths. This gives unbiased estimates of population totals over repeated sampling.

It may seem reasonable to remove the defunct units from the frame, and to update both N and n . After all, we know the units are no longer part of the target population, so why not use this information? If there are d dead units in sample, this would result in applying weights of $(N-d)/(n-d)$. The problem with this idea is that we have removed all of the dead sample units from the sample count, but we have not removed all of the dead population units from the population count. Ideally, we should weight using $(N-D)/(n-d)$, where D is the number of dead population units, but unfortunately we don't know D , otherwise we would have removed these units already. If we incorrectly weight using $(N-d)/(n-d)$, the relative bias will be positive and approximately equal to (d/n) , the sample death rate (this approximation based on $n \ll N$ and $d \ll n$). The death rate can be 10% in practice, or even higher, so a bias of this magnitude would not usually be acceptable.

A more subtle question is whether defunct units can be removed from the frame for the purpose of running future surveys. In many National Statistical Institutes surveys share information on deaths they obtain through their samples with other surveys to speed up the information process. It would seem natural that this new information should be made available to other sample surveys, which otherwise may include the dead units in their samples and thereby lose precision. However, as pointed out by Srinath (1987) among others, such a procedure may cause bias, if the samples are

dependent on each other, for example in a rotating panel survey or if overlap between different surveys is used.

To see this, consider a situation where all dead units are identified as dead and deleted at the first wave of a panel survey. If no further deaths have occurred up to the second-wave observation of the panel units, the second-wave sample contains only live units. Suppose that no information has been kept about the number of dead units that have been deleted from the frame. While more information about the population has been gathered when the deaths were recorded at the first wave, there is actually less information in the second wave-sample on the proportion of live units in the population. This lack of information may cause bias.

A safe recommendation would be that no information on deaths from sample surveys, other than from completely enumerated strata, may be used to update the frame when samples are co-ordinated over time (cf. Ohlsson 1995, p. 168, and Colledge 1989, p. 103). However, to prohibit feeding back seems to deny oneself the use of all available information. Lavalley (1996) and Hedlin and Wang (2004) give expressions for the feedback bias. Other references include Hidioglou and Srinath (1993) and Hidioglou and Laniel (2001). Hedlin and Wang (2004) assume that the total is estimated with the common expansion estimator. They show that the feedback bias can be large. With as little as 5% ineligible units on the frame, feeding back information of these from sample surveys can result in about 2-3% bias. However, a small-scale simulation study indicates that if the proportion of ineligible units is 3% or less, the feedback strategy does not seem to create problems in terms of bias and variance.

One way that information on deaths can be legitimately used is to retain these units on the frame, but to flag units found to be defunct. If these units are selected in sample, they would be retained for the purpose of calculating weights and assigned zero values for all variables of interest, but would not actually be interviewed or sent a form. This approach can be used in all cases, even when the proportion of ineligible units is more than 3%, provided a consistent definition of defunctness is used.

References

Colledge, M.J. (1989). Coverage and Classification Maintenance Issues in Economic Surveys. In *Panel Surveys*. (Eds. Kasprzyk, D., Duncan, G. J., Kalton, G., and Singh, M. P.). New York: Wiley, 80-107.

Hedlin, D. and Wang, S. (2004). Feeding Back Information on Ineligibility from Sample Surveys to the Frame. *Survey Methodology*, 30, 167-174.

Hidioglou, M.A. and Laniel, N. (2001). Sampling and Estimation Issues for Annual and Sub-Annual Canadian Business Surveys. *International Statistical Review*, 69, 487-504.

Hidioglou, M.A. and Srinath, K.P. (1993). Problems Associated with Designing Subannual Business Surveys. *Journal of Business and Economic Statistics*, 11, 397-405.

Lavallée, P. (1996). Frame Update Problems with Panel Surveys. *Proceedings of Statistical Days '96*, Statistical Society of Slovenia, 252-261.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In *Business Survey Methods*. (Eds. Cox, B., Binder, D., Chinnappa, N., Christianson, A., Colledge, M. and Kott, P). New York: Wiley, 153-169.

Srinath, K.P. (1987). Methodological Problems in Designing Continuous Business Surveys: Some Canadian Experiences. *Journal of Official Statistics*, 3, 283-288.

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.



Canada

Survey of Young Canadians

The Survey of Young Canadians (SYC) is a new cross-sectional survey that aims to provide information on child development and on the prevalence of various risks and protective factors for children in order to develop policies and programs that will help children and youth. The SYC replaces the National Longitudinal Survey of Children and Youth (NLSCY) as the main source of information for children living in Canada. The NLSCY which started in 1994 and was administered every two years until 2008 served both longitudinal and cross-sectional purposes. However, one of the main issues of the NLSCY's cross-sectional sample was that the desired quality for estimates in some of the smaller provinces was difficult to achieve due to lack of sample. The NLSCY used rotation groups from the Labour Force Survey (LFS) to get its sample and since the LFS was not specifically designed to target households with children, smaller provinces with less LFS sample could not provide the desired SYC targets. A number of alternative frames better suited to target children were considered in the design of the SYC and the one that proved the most promising was the Canadian Child Tax Benefit (CCTB) file.

The CCTB is a non-taxable amount paid every month to help eligible families with the cost of raising children under 18 years of age. The Canada Revenue Agency (CRA) manages the program and maintains the list of applicants to the CCTB. Under the authority and conditions of the Statistics Act, CRA has had a Memorandum of Understanding to share the data with Statistics Canada for a number of years, for example to measure intra-provincial migration. However, the objective had never been to use the CCTB as a sampling frame, so a thorough evaluation of the CCTB, with focus on coverage and quality of contact information was undertaken. By 2009, the estimated coverage of children on the CCTB exceeded 95% for children up to the age of 18, with slightly higher rates for younger children and no evidence of undercoverage bias for key socio-demographic variables. The quality of contact information was evaluated through a field test of 1,000 children that was conducted in February 2010 as well as with linkages of the CCTB to current Statistics Canada survey data. All showed the contacted information to be of high quality.

As a result of our evaluations, the CCTB file was used as the sampling frame for SYC in the fall of 2010 and the data is currently being processed for April 2012 dissemination. The SYC was one of the first large scale Canadian household surveys to use an administrative database not maintained internally as its sampling frame.

For more information, contact Geoff Bowlby (613-951-3325 or geoff.bowlby@statcan.gc.ca), Special Surveys Division or Pierre Caron (613-951-

1459 or pierre.caron@statcan.gc.ca), Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

Real Time Remote Access

The Real Time Remote Access (RTRA) project is a recent initiative undertaken by Statistics Canada (StatCan) to increase access to microdata. In recent years, Statistics Canada has invested much time and effort into examining ways to meet the demands of researchers while, at the same time, ensuring that the legislative requirement under the Statistics Act to protect the confidentiality of respondent data is met. A key option that Statistics Canada has decided to explore is the development of a RTRA application. This application is essentially an on-line remote access facility that would allow researchers to run—more or less in real time—data analyses on microdata or lightly masked microdata sets kept in a central and secure location under the control and care of StatCan. When fully operational, the RTRA application would provide faster access to more data and give more flexibility to researchers by providing a system that potentially gives them desktop access to data 24 hours a day, seven days a week.

In past years, research and development was done on different aspects of RTRA such as the security, legal, system and functionality requirements along with the basic features, procedures and governance of the program. The key steps were: i) identifying the scope: descriptive statistics and modelling for household surveys, census and administrative data; ii) defining an approach, i.e., a gradual implementation with the construction of a prototype using SAS software that will be tested in conjunction with limited partners; expansion would follow after a careful evaluation of risks; iii) developing the process model; examination of the informatics as well as the contractual and legal aspects; iv) developing the governance model for the application; and v) ensuring proper organizational implementation: privacy impact assessment and risk and threat assessment, communication and marketing plans for the launch of the system. A prototype was launched in Spring 2010. The system was functional with no manual intervention and limited to seven household surveys and to only a few partners.

Research and development was also completed on the methodological aspects of RTRA, especially for the Statistical Disclosure Control (SDC) methods to secure the outputs. For the first output, tabular frequencies, an additive controlled rounding (ACROUND) method was chosen for its simplicity of application and programming and its ability to protect against potential links to Public Use Microdata Files (PUMFs) and multiple query submissions. In early 2011, SDC methods were developed for means, percentiles and proportions. Future developments include SDC methods for ratios and totals, modelling, censuses and administrative data. The possibility of introducing minimal noise into the microdata file will also be considered.

For more information, contact Heather Dryburgh (613-951-0501 or heather.dryburgh@statcan.gc.ca), Microdata Access Division or Michelle Simard (613-951-6910 or michelle.simard@statcan.gc.ca), Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

France

The End of Life Survey in France merits our attention for the sensitive nature of the questions it deals with and the protection techniques it uses for the extremely confidential data it collects. The survey studies the circumstances surrounding deaths, treatments in place at the end of life and final medical decisions made. Medical decisions at the end of life are the focus of an important debate in Europe and France, having very different legislation in different countries, which are sometimes subject to revision. The Eureld (European End of Life Decisions) survey showed the diversity of practices among Northern Europe and Southern Europe (Van der Heide, *Lancet* no. 362, 2003). Since 2005, French legislation prohibits prolonging life by technological means and inevitably implements palliative care with respect to the will of the patient, without authorizing euthanasia. In 2010, the *Institut national d'études démographiques* decided to conduct a survey on this subject in line with European surveys, bearing in mind the context of French legislation.

The methodology used consists of pulling a representative sample of deaths, the most recent possible, and contacting the medical signatory of the death certificate. This method retrospectively assures, in effect, the best representative of the deaths. The questionnaire asks about the final interventions applied to dying person, actions to diminish pain or not, legal or not. The collection is done by self-coded questionnaire via either the Internet or mail, whichever the doctor in question chose. In the Eureld survey, anonymity was guaranteed by a third party. A similar plan was put in place to assure the anonymity of paper questionnaires addressed by mail, whereas a completely automated procedure protected the electronic questionnaires, thanks to irreversible encrypting techniques of identifiers, the hash function. To the surprise of the team, three-quarters of medical professionals chose to send their responses by mail. Those who chose to respond by Internet were most often male, mostly young doctors and mostly doctors practicing in hospitals. Surveys about doctors are very difficult to conduct, at least in France. Thus, the response rate is affected: the rate of 43% of participant doctors obliged to take a serious examination of the representativeness of the sample. This control was conducted by comparing the distributions of the available data from death certificates on people who died and the distribution of the same variables in the questionnaire (a pairing between the two bases is impossible because of the protocol used to guarantee the anonymity of the responses). A survey of doctors who didn't respond to the survey, despite the different follow up procedures that were attempted (by telephone or postal mail), showed that the survey non-response was essentially linked to a lack of time and neither to the subject of the survey nor the content of the questionnaire.

This survey covers a panorama of how end of life is dealt with in France and brings objective elements that stimulate debate on this important question for society. The first results of the survey are available at <http://fdv.site.ined.fr> or from Sophie Pennec (penne@ined.fr) and will be the topic of scientific publications.

India

Dr. Gayatri V. Singh
Assistant Professor, AMITY University,
India

In continuation to the last report focused on Indian Statistics, this report is focused on the Statistics Journals publishes in India.

Statistical journals published in India

The oldest and premium journal of India is SANKHYA also known as The Indian Journal of Statistics. This journal is published by Indian Statistical Institute (ISI), Kolkata. The Indian Journal of Statistics was founded by **Prasanta Chandra Mahalanobis** in 1933. He became the first editor of this journal and continued until his death. The research articles of this journal covers the broad areas of Theoretical Statistics, Probability and Applied Statistics. The journal is now published in two series: A and B. Series A issues come out in February and August while Series B issues come out in May and November. Each volume has four parts - two in Series A and two in Series B.

Journal of the Indian Statistical Association (JISA) is published by the Indian Statistical Association (ISA), Maharashtra. The ISA publishes two issues of the journal every year. It is published from Department of Statistics, University of Pune, Maharashtra.

The Indian Society of Agricultural Statistics, Indian Agricultural Statistical Research Institute (IASRI), New Delhi publishes an International Peer Reviewed Journal called “**Journal of the Indian Society of Agricultural Statistics**”. Three issues of the Journal (April, August and December) are published annually. The first volume of the Journal was released in 1948. The Journal devoted to the publication of original research papers on all aspects of Statistics and Computer Applications preferably with innovative applications in Agricultural Sciences or that have a potential application in Agricultural Sciences.

International Journal of Statistics and Systems (IJSS) publishes top-level work to covering all disciplines and branches of statistics and management systems. It is published by Research India Publications, Delhi. This journal is published quarterly. The list of statistical journals published in India:

Journal	Frequency	Published by
Advances and Applications in Statistics	Monthly	Pushpa Publishing House, University of Allahabad
Advances in Performance Analysis	Six Monthly	Notable Publications
Aligarh Journal of Statistics	Annual	Aligarh Muslim University, Aligarh
Calcutta Statistical Association Bulletin		Calcutta Statistical Association, Kolkata
Journal of the Indian Association for Productivity, Quality (IAPQR)	Bi-Annual	Calcutta University, Kolkata
Journal of Decision and Mathematical Sciences	Annual	Society of Decision and Mathematical Science, New Delhi,
Journal of Difference Equations and Applications	Monthly	Overseas Publishers Association
Journal of Information & Optimization Sciences	Annual	Anal. Publications

Journal of Statistics & Management Systems	Annual	Academic Forum, India
Journal of the Indian Statistical Association	Six Monthly	Indian Statistical Association
Opsearch	Quarterly	I.S.I., New Delhi
Statistical Methods	Annual	Statistical Methods, India
International Journal of Stochastic Modelling and Applications	Bi – Annual	Calcutta University, Kolkata

Note: This list of Journals is not an exhaustive list.

The Central Statistics Office (CSO), Ministry of Statistics and Programme Implementation has introduced a new series of Consumer Price Indices (CPI) on base 2010=100 for all-India and States/UTs separately for rural, urban and combined with effect from January, 2011. The General Indices for rural, urban and combined are 113.1, 109.8 and 111.7 respectively. State/ UT wise Group level provisional indices are available in the Ministry's website (www.mospi.gov.in).

New Zealand

Significant Investment in Official Statistics in New Zealand.

In the May 2011 Budget, the New Zealand Government increased Statistics New Zealand's funding by \$58 million over four years, plus \$13 million in capital funding. This investment reflects the Government's commitment to official statistics produced and used by government departments, and to the continued supply of the country's vital economic and social statistics.

The funding will support a change programme (see: Statistics NZ Strategic Plan 2010 - 2020, http://www.stats.govt.nz/about_us/our-publications/snz-strategic-plan-2010-20.aspx) that will target efficiency and effectiveness gains across the Official Statistics System (OSS). There will be a significant focus on upgrading and standardising Statistics NZ's IT systems and working on transforming the delivery of statistics. Also, there will be a real focus on the OSS.

Official statistics must provide better value for the taxpayer dollar by finalising an agreed list of New Zealand's Tier 1 statistics, improving access to government held data, and identifying opportunities to increase efficiencies across the OSS. We have to operate smarter across the system, building in common platforms and standard methods as far as possible.

Ultimately great strides have to be made in marrying up users and the data – ease of access is critical.

The Government will continue to monitor Statistics NZ closely to ensure that efficiencies are maximised and that the department delivers on the objectives set out in the above Strategic Plan. Good decision-making is based on good information. The quality of life of all New Zealanders is intrinsically linked to getting things right in this space.

For the Statistical Methods branch, this investment means that we need to lead the organisation through sound statistical methodology and innovative solutions to many problems. We have already started the thinking process, but mindful not to deliver a 2010 statistical system in 2020.

Poland

Tomasz Żądło

7th Conference Survey Sampling in Economic and Social Research (in 60th anniversary of foundation of the Department of Statistics at the University of Economics in Katowice)

The conference took place 18-20 September 2011 in Katowice and was organized by the Department of Statistics of the University of Economics in Katowice with the cooperation of the Department of Statistical Methods of the University of Łódź and Polish Statistical Association.

Forty-two persons from 9 countries, representing universities, statistical agencies and other institutions participated in the conference. The purpose of this conference was to stimulate research both in theoretical and methodological developments in survey sampling and related fields, and practical applications of the methods, including their potential usage in various research areas. Topics discussed during the conference included: estimation of population parameters based on complex samples, statistical inference based on incomplete data, small area estimation, sample size and cost optimization in survey sampling, sampling designs, statistical inference using auxiliary information, model-based estimation, longitudinal surveys, practical implementations of sampling methods, sampling in statistical quality control, and sampling in auditing.

The invited papers were presented by

- Prof. Malay Ghosh (University of Florida, USA) *Estimation of median incomes for small areas : A Bayesian semiparametric approach.*
- Prof. Jean-Claude Deville (Ecole Nationale de la Statistique et de l'Analyse de l'Information, Laboratoire de Statistique d'Enquête, France) *Uses of calibration and balanced sampling for the correction of non-response in surveys.*
- Prof. Parthasarathi Lahiri (University of Maryland, USA) *Robust small area estimation*
- Prof. Nicholas T. Longford (Universitat Pompeu Fabra, Spain) *Policy-related small-area estimation.*

Details including abstracts of all of the talks are available at the conference website <http://web.ue.katowice.pl/metoda>.

The Organizing Committee would like to thank sponsors of the conference: University of Economics in Katowice, Katowice City Hall and SPSS Poland.

National Census in Poland

The census was held 1 IV-30 VI 2011 and it was the first census in the Republic of Poland after joining EU and the first when internet self-enumeration was possible. More technical information is available at <http://www.stat.gov.pl>.

Congress of Polish Statistics

The Congress of Polish Statistics will be organized 18-20 April 2012 in Poznań in Poland to celebrate the 100th anniversary of the Polish Statistical Association. More information is available at: <http://www.stat.gov.pl/pts/kongres2012/english>

United States of America

In June 2011, the Census Bureau announced a restructuring of its Field organization. Beginning in January 2012, the Census Bureau will transition from the current structure of 12 data collection offices located throughout the US, to 6 data collection offices. Further, the management of data collection will be modified to adapt to the new structure. The supervision of field representatives (interviewers) will move from an office-based supervisor in one of the 12 offices to a structure of 600 field-based supervisors that work from home. The restructuring effort introduces an expanded level of technical and data analysis of survey data collection activities at the remaining 6 offices. The day-to-day activities of the field representatives (interviewers) remain unchanged. The restructuring decision was made based on a number of factors, but the primary driver was cost-savings. As a result, the Census Bureau is projecting savings of as much as \$15 Million / year in infrastructure costs alone. This process is expected to be fully implemented by January 2013.

For more information, contact Howard R. Hogan (howard.r.hogan@census.gov).



New and Emerging Methods

Evaluating, comparing, monitoring and improving representativeness of survey response through R-indicators and partial R-indicators

Barry Schouten, Jelke Bethlehem, Koen Beullens, Øyvind Kleven, Geert Loosveldt, Annemieke Luiten, Katja Rutar, Natalie Shlomo and Chris Skinner

Abstract: Recently, a strong focus on methods for survey data collection monitoring and tailoring has emerged as a new paradigm to efficiently reduce nonresponse error. Paradata and adaptive survey designs are key words in these new developments. Prerequisites to evaluating, comparing, monitoring and improving quality of survey response are a conceptual framework for representative survey response, indicators to measure deviations thereof, and indicators to identify subpopulations that need increased effort. We describe a set of indicators, called R-indicators, that we believe are fit for these purposes.

1. Introduction

Since surveys, especially those administered by interviewers, are expensive tools, survey institutes constantly seek a balance between reducing unit-nonresponse and costs of survey data collection. In recent literature, a shift of focus has taken place from nonresponse analysis and nonresponse adjustment to nonresponse monitoring and tailoring of data collection (Groves and Heeringa 2006, Kreuter et al. 2010). The new paradigm of survey data collection monitoring and tailoring has led to a growing interest in quality indicators for evaluating and monitoring response and in sources of data for constructing such indicators.

Indicators for evaluating nonresponse error serve four purposes:

1. To compare response between different surveys that share the same target population, e.g. households or businesses.
2. To compare the response to a survey longitudinally, e.g. monthly, quarterly or annually.
3. To monitor the response to a survey during data collection, e.g. after various days, weeks or months of fieldwork.
4. To adapt the data collection design by tailoring based on historic data, and available frame data and paradata.

We provide an overview of R-indicators and partial R-indicators that were designed to be fit for these purposes. They were introduced by Schouten, Cobben and Bethlehem (2009) and Schouten, Shlomo and Skinner (2011). The indicators are based on measures of variability in response propensities.

Alternative indicators exist. The most well-established response quality indicator is the response rate. This has the advantage of simplicity and ease of calculation, but suffers from often having only a limited relation to nonresponse bias. The latter might be taken to be the ideal measure of nonresponse error. However, it is rarely measurable directly and, moreover, most surveys are designed to produce a large number of survey estimates and the correspondingly large number of nonresponse

biases might be too great to serve many needs of quality indicators e.g. between-survey comparisons. Nonetheless, indicators that focus on the impact of nonresponse on specific survey statistics have been proposed recently, see Wagner (2010) and Andridge and Little (2011). These indicators are especially useful when surveys have only a few key substantive variables.

2. R-indicators and partial R-indicators

The indicators that we propose build on two definitions: representative response and conditionally representative response. The first definition is introduced by Schouten, Cobben and Bethlehem (2009); the second is discussed by Schouten, Shlomo and Skinner (2011). Let X be a vector of auxiliary variables, e.g. a vector consisting of age, gender and degree of urbanization of residence area for a household survey or the type of business and the number of employees for a business survey. Response is called *representative with respect to X* when the response propensities of all subpopulations formed by the auxiliary variables are constant (and, hence, equal to the overall response rate).

Let Z be one of the variables in the vector X , e.g. age or number of employees, and let X^- represent vector X where Z is deleted. Response is called *conditionally representative for Z given X^-* , when response propensities for the subpopulations formed by Z are constant within the subpopulations formed by the other auxiliary variables in X . Hence, age is conditionally representative with respect to gender and urbanization when the propensities for different age categories are constant when the gender and degree of urbanization are fixed.

In other words, representative response can be viewed as arising when the respondents form a random subsample of the survey sample itself and conditionally representative response arises when they form a stratified random subsample.

Representativeness indicators or R-indicators (Schouten, Cobben and Bethlehem 2009) measure the extent to which survey response deviates from representative response. The response to a survey is less representative with respect to X when the response propensities exhibit more deviation around the overall response rate. The *R-indicator* is a simple measure and is based on the standard deviation of response propensities.

$$R(\rho_X) = 1 - 2S(\rho_X), \quad (1)$$

where ρ_X indicates the individual response propensities given the auxiliary variables X and $S(\rho_X)$ is the standard deviation of these propensities. The transformation in (1) assures values between 0 and 1, where 1 is representative response. As we will show in the following sections, R-indicators can be used to compare representativeness of surveys, but they cannot be used to identify subgroups in monitoring and improving representativeness of response. For this reason, R-indicators were supplemented by partial R-indicators, see Schouten, Shlomo and Skinner (2011). Partial R-indicators are available at the variable level and at the level of categories of variables in order to allow for detailed levels of analysis. Furthermore, partial R-indicators can be computed unconditionally and conditionally, thereby relating to representative and conditionally representative response.

Essentially, partial R-indicators decompose the variation in response propensities according to different values of the auxiliary variables. The standard deviation $S(\rho_X)$ in (1) can be rewritten as

$$S^2(\rho_X) = S_W^2(\rho_X | H) + S_B^2(\rho_X | H), \quad (2)$$

where H is some stratification of the population into subgroups and $S_W^2(\rho_X)$ and $S_B^2(\rho_X)$ are, respectively, the within and between variances of the response propensities. Partial R-indicators ideally have values equal to zero.

The *unconditional partial R-indicator* for a variable Z in vector X is based on the between variance for a stratification of the population using Z , e.g. the between variance for different age groups or different numbers of business employees. It measures the extent to which the response is representative with respect to Z . It is defined as the variation of response propensities estimated using X between the strata of Z

$$P_u(Z, \rho_X) = S_B(\rho_X | Z).$$

The between variance in (3) only depends on the other components in X , when response propensities are estimated using main effects models. In practice, the variances $S^2(\rho_Z)$ and $S_B(\rho_X | Z)$ are often very close to each other.

The *conditional partial R-indicator* for Z given X^- is based on the within variance for a stratification of the population using X^- , e.g. the remaining variance due to age within subgroups formed by gender and urbanization or due to number of employees within groups formed by business type. The conditional partial R-indicator measures the extent to which response is conditionally representative. It is defined as

$$P_c(Z, \rho_X) = S_W(\rho_X | X^-). \quad (4)$$

The within variance in (4) is the remaining variation in response propensities estimated using X within the strata formed by all variables but Z . Hence, it isolates the variation that can be attributed to variable Z adjusting for multicollinearity.

Partial R-indicators for categories of variables are composed by isolating the contribution to between and within variances attributable to those categories. The unconditional category-level partial R-indicators have a positive or negative sign in order to indicate the direction of the representativeness of the category, e.g. a negative sign for males implies that males responded less to the survey than females. The unconditional partial R-indicator for a category k of variable Z is defined as

$$P_u(Z, k, \rho_X) = \sqrt{\frac{N_k}{N}} (\bar{\rho}_k - \bar{\rho}), \quad (5)$$

where N_k is the number of population units in category k , $\bar{\rho}_k$ is the average response propensity in the stratum formed by $Z = k$, and $\bar{\rho}$ is the overall average response propensity. We have that $P_u(Z, k) \in [-0.5, 0.5]$.

The conditional category-level partial R-indicators do not have a sign and are always positive. Conditionally it does not have a meaning to attach a direction to the indicator as units in a particular category may respond better in one subpopulation but less well in another. For example, young males may respond better than young females while older males may respond worse than older females. Conditional partial R-indicators for categories of variables are constructed by distributing (4) over the categories of Z . We refer to Schouten, Shlomo and Skinner (2011) for detailed descriptions of the various indicators.

One important remark needs to be made concerning the estimation of the R-indicators and partial R-indicators. Since individual response propensities are unknown, they need to be estimated from the survey data. The indicators themselves are subject to the variation in samples. This variation needs to be accounted for. We refer to Shlomo, Skinner and Schouten (2011) for a detailed description and discussion. The website www.risq-project.eu contains open source code in SAS and R that computes the various indicators and accounts for the sampling variation.

3. Example: monitoring contact and participation in a household survey

The use of indicators for monitoring survey data collection is illustrated through the 2005 Dutch Survey of Consumer Confidence. The Survey of Consumer Confidence (SCC) is a monthly cross-sectional telephone survey among 1500 households. In the following we focus on the two subprocesses that have the strongest impact on response rates and response representativeness: contact and cooperation. We combine the less influential causes of nonresponse (language problems and not able during fieldwork) with refusal into the subprocess of obtaining cooperation. Hence, when monitoring cooperation conditional on contact, we analyze contacted cases only.

We selected the variables type of household, age, ethnicity, urban density of residence area, gender and average house value at zip code. From literature, these variables are known to relate to contact and cooperation. In this particular study we did not have paradata observations, but they can be added for the two subprocesses in a straightforward way.

In Table 3.1 the indicators are given for contact and cooperation and for the overall response. Cooperation is evaluated conditionally on contact and eligibility, and contact is evaluated conditionally on eligibility. Somewhat surprisingly, in obtaining contact all variables play almost an equal role except for average house value. However, after conditioning on the other variables it is age that contributes most to the variance in response propensities. The partial R-indicators for cooperation are larger than for contact. Unconditionally, the strongest variables are age, type of household and gender, but after conditioning it is age that contributes most. Remarkable also is the large value for ethnicity. It turns out that ethnicity has an impact on cooperation that is almost orthogonal to the impact of the other variables.

Summarizing, in improving representativeness for contact we should look at age and for cooperation we should consider age and ethnicity. The other variables, however, also play a role in each of the subprocesses but not as distinct. In almost all cases, the partial R-indicators for cooperation are larger than those for contact, i.e. potentially the biggest increase in representativeness comes from reducing refusal.

Table 3.1: Variable-level partial R-indicators for six auxiliary variables in the SCC 2005.
The variables are type of household, age, ethnicity, urban density, gender and average house value at zip code.

	<i>Unconditional</i>			<i>Conditional</i>		
	<i>Contact</i>	<i>Cooperation</i>	<i>Response</i>	<i>Contact</i>	<i>Cooperation</i>	<i>Response</i>
Household	0.0018	0.0054	0.0063	0.0008	0.0012	0.0015
Age	0.0017	0.0069	0.0058	0.0021	0.0049	0.0036
Ethnicity	0.0014	0.0031	0.0037	0.0010	0.0025	0.0029
Urban	0.0014	0.0013	0.0019	0.0009	0.0010	0.0010
Gender	0.0021	0.0049	0.0058	0.0008	0.0006	0.0010
House value	0.0007	0.0026	0.0027	0.0008	0.0007	0.0008

Table 3.2 shows the category-level partial R-indicators for age for contact, cooperation and overall response. A closer look at the contact category-level indicators for age reveals that persons younger than 45 affect representativeness in a negative way. Especially persons between 25 and 40 show low contact propensities and, as a consequence, increase the variation in contact propensities. Conditioning on other variables has a minor impact. Hence, when improving contact representativeness, the focus should be on persons between 25 and 40 years of age.

For obtaining cooperation one age group immediately jumps out, persons of 70 years and older. This relatively small group has a large negative value. Conditional on the other variables the value remains large. Therefore, we conclude that measures to improve cooperation representativeness could aim at elderly persons.

Table 3.2: Category-level partial R-indicators for age in the SCC 2005.

	<i>Unconditional</i>			<i>Conditional</i>		
	<i>Contact</i>	<i>Cooperation</i>	<i>Response</i>	<i>Contact</i>	<i>Cooperation</i>	<i>Response</i>
< 25	-0.0027	0.0025	0.0003	0.0020	0.0057	0.0044
25 -29	-0.0084	0.0088	0.0020	0.0082	0.0101	0.0048
30-34	-0.0088	0.0139	0.0065	0.0092	0.0126	0.0064
35-39	-0.0065	0.0168	0.0109	0.0088	0.0109	0.0057
40-44	-0.0014	0.0184	0.0163	0.0048	0.0118	0.0085
45-49	0.0016	0.0108	0.0114	0.0031	0.0080	0.0055
50-54	0.0041	0.0049	0.0077	0.0039	0.0090	0.0057
55-59	0.0029	0.0108	0.0124	0.0038	0.0129	0.0118
60-64	0.0008	0.0063	0.0066	0.0026	0.0107	0.0087
65-69	0.0004	0.0024	0.0025	0.0024	0.0087	0.0070
>69	0.0083	-0.0595	-0.0508	0.0116	0.0366	0.0278

4. Discussion

R-indicators and partial R-indicators do not solve the nonresponse problem. The indicators rely on information that is auxiliary to a survey or register in order to evaluate the representativeness. As a consequence, the indicators depend strongly on the auxiliary variables that are selected. The choice of variables depends on the intended use of the indicators and obviously on the availability of variables. When comparing different surveys, auxiliary variables need to be available in all surveys and be of general interest to all surveys. When the quality of a single survey is assessed, the scope can and must be extended to variables that relate also to the key survey topics. The dependence on auxiliary variables is a weakness of the indicators, but no indicator can be constructed that is independent of external information and, hence, of auxiliary variables.

The strength of the indicators lies in its mathematical basis and its structured use of auxiliary information in evaluating representativeness of response. Like the R-indicator can be viewed as supplemental to the response rate, so can partial R-indicators be viewed as supplemental to subgroup response rates. We believe that partial R-indicators are more relevant than subgroup response rates. Also, R-indicators, contrary to subgroup response rates, allow for different levels of analysis. The main differences between subgroup response rates and partial R-indicators are:

1. Partial R-indicators are linked to R-indicators, i.e. they represent the contribution of variables to the lack of representativeness, while subgroup response rates are linked to response rates. In other words they are conceptually different, although they have response propensities as basic ingredients.
2. Partial R-indicators are available at the variable level.
3. Partial R-indicators are computed both unconditionally and conditionally.
4. Partial R-indicators are weighted differences between subgroup response rates and overall response rates. The weights are equal to the distribution of the population over the subgroups. Subgroup response rates do not account for the size and, hence, impact of a subgroup.

The second difference makes R-indicators and partial R-indicators fit for detailing the level of analysis. The third and fourth differences make partial R-indicators relevant in searching for groups to target that help improve representativeness of survey response.

A criticism often put forward is that any traces of nonrepresentative response found by the indicators can be corrected for using standard nonresponse adjustment methods. It is true that candidate variables for evaluation of representativeness are also candidates for nonresponse adjustment procedures. There are, however, two important differences between evaluation of representativeness as presented in this article and nonresponse adjustment.

The evaluation of representativeness is not directed at a particular survey variable, a particular population parameter or a particular estimator. Representativeness is assessed in terms of unit-level nonresponse. For this reason, R-indicators and partial R-indicators provide an incomplete picture when the purpose is adjustment to reduce nonresponse bias of a certain population estimator using a specified estimator. Nonresponse adjustment requires a detailed look at the relation between a specific survey variable and nonresponse. Clearly, when a survey contains a wide range of key variables, then nonresponse adjustment tends to focus on the explanation of nonresponse behaviour rather than on the explanation of a survey variable. In these settings the indicators may provide input to nonresponse adjustment.

The second difference lies in the assumptions underlying nonresponse adjustment. Most nonresponse adjustment methods assume a missing-at-random mechanism; the probability to respond does not depend on the survey variable within strata formed by auxiliary variables. In general, this assumption fails to hold, but any weaker assumption cannot be tested without additional information or intrinsic knowledge of the nature of the nonresponse behaviour. The indicators in this article measure the extent to which a survey researcher's reliance on missing-at-random assumptions may be misplaced. Although variation of response propensities within strata could still be independent of the value of the survey variable, we believe that the larger the value of the R-indicator, the more questionable is the validity of the missing-at-random assumption.

More empirical evidence and experience are needed to get a full insight into the potential role of R-indicators and partial R-indicators in data collection. Although the values of the indicators are especially interesting in a relative sense, i.e. from one design to another, it is paramount to get a better feeling about acceptable levels of the indicators. What levels would mean that immediate action is needed? Finally, the nonresponse error is one component of the total survey error. A blind focus on a single error source is too simplistic. Care is needed when altering data collection designs in order to improve representativeness. Future research should focus on combining these efforts with assessment of other relevant errors, most notably measurement errors.

Acknowledgements: This research was supported under the research project Representativity Indicators for Survey Quality (RISQ, www.risq-project.eu), funded by the European Commission in the 7th EU Research Framework Programme.

References

- Andridge, R.R. and Little, R.J.A. (2011), Proxy Pattern-Mixture Analysis for Survey Nonresponse. *Journal of Official Statistics*, 27 (2), 153-180.
- Groves, R.M., Heeringa, S.G. (2006), Responsive design for household surveys: tools for actively controlling survey errors and costs, *Journal of the Royal Statistical Society: Series A*, 169, 439-457.
- Kreuter, F., Olsen, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casa-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M. and Raghunathan, T.E. (2010), Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society: Series A*, 173, 389-407.
- Schouten, B., Cobben, F. and Bethlehem, J. (2009), Indicators for the representativeness of survey response, *Survey Methodology*, 35, 101-113.
- Schouten, B., Shlomo, N., Skinner, C. (2011), Indicators for monitoring and improving survey response, *Journal of Official Statistics*, 27 (2), 231-253.
- Shlomo, N., Skinner, C.J., Schouten, B. (2012), Estimation of an indicator of the representativeness of survey response, *Journal of Statistical Planning and Inference*, 142, 201-211.
- Wagner, J. (2008), Adaptive survey design to reduce nonresponse bias, PhD thesis, University of Michigan, USA

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 Ineke Stoop at i.stoop@scp.nl.



Book and Software Review

Study of Record Linkage Software for the 2010 Brazilian Census Post Enumeration Survey

Andrea Diniz da Silva (adiniz@ibge.gov.br);
Otavio SantAna Martins Romeo;
Thiago Silva Soares; Vinicius Layter Xavier
Brazilian Institute of Geography and Statistics, Brazil

1. Introduction

One of the biggest improvements of the 2010 Brazilian Census Post Enumeration Survey (PES) conducted by the Brazilian Institute of Geography and Statistics (IBGE) is the incorporation of new methodologies and technologies developed or improved throughout the last decade. The use of handheld devices for the data collection, already experienced in the 2007 population count, was one of the successful innovations in the Census project, allowing improvement of quality and timeliness in the data collection process.

The use of new technologies had an effect on the structure and format of the data collected as it was delivered immediately after its collection. This change facilitates the need to perform automatic matching of census to the post enumeration survey data. In addition, the development of software packages that implements computational models for record linkage and matching has increased over the last 10 years, making it possible to find several software packages that perform similar tasks.

In order to choose the best software for linking records from the Brazilian Census and the PES, IBGE concentrated efforts in studying software over the last two years. The studies involved assessment of corporate and free (open source) software, taking into account not only operational aspects but also methodological issues of each package. The experiment was based on some documentation and on software trial licenses when available, considering their suitability to the Brazilian Census PES project. The main findings on this study are presented in the next sections.

2. Review of Some Record Linkage Software

Data Quality

The Data Quality software was developed by SAS Institute Inc. Its evaluation was based on the available documentation for the period of May 2009 and December 2010 and on a presentation made by a representative of SAS to staff of IBGE.

Very little literature was found about Data Quality so that the *SAS 9.1.2 Data Quality Server: Reference* was the main source used. Data Quality is a package that enables you to analyse, cleanse, transform and standardise your data. Its documentation offers a comprehensive description on the functions, system options, macros and procedures available, all accompanied by a rich amount of examples on programming.

However, no detailed information on the methodological aspects used for matching or deduplication was found. Some information was also obtained from the Data Quality website, but again no details about the methodological aspects of matching and deduplication were found. Besides, the SAS representative could not clarify issues related to methodological aspects during his presentation, so that no information on the available methodologies and models for deduplication, space reduction or matching could be obtained for further evaluation.

Thus, the literature review and the exchange during the presentation led to the conclusion that the software is well developed for data cleansing, but information to allow evaluation of methodological aspects of the software is still needed. No trial license was available for Data Quality, thus it was not possible to assess the interface, usability and the different features of the software.

QualityStage – IBM

Traded by International Business Machines (IBM), QualityStage is a component of IBM Information Server package integrated to a platform called DataStage and a tool for analysis and data standardization, as well as matching jobs, deduplication and survival processes. For data analysis, the software allows to set type equal to “default” or to a particular character of the variable of analysis. It is useful for checking the composition of the database prior to performing the matching or deduplication. Besides, concatenation of variables can also be done. For standardization, it is possible to build a dictionary to recode values of certain variables by pre-established values in this dictionary.

Matching and deduplication are the most important functions for the 2010 PES, so they were analyzed in more detail. Deduplication is matching done within the same database, so all the functions described for the matching are also available for deduplication.

There are four types of matching available in the software, however only the one-to-one would be useful as a record from the Census can only match with a single record

from the PES and vice versa.

QualityStage also has the option of reducing the comparison space by using blocking. Besides, it has comparison functions called approximate comparison algorithms. There are sixteen different functions but none of them are the comparison functions of Jaro and Jaro-Winkler, which fit best to the methodology chosen for the 2010 Brazilian PES.

The matching results occur through grade cuts and use parameters similar to the Fellegi-Sunter Record-Linkage theory, but no software release shows this theory. They only explain the calculated parameters vaguely.

In general, QualityStage has several functionalities, but many of them are similar and in practice produce the same results. The software is not easily accessed and understood by technicians who are not in the IT area, which makes its use a difficult task. Also, to get the program working efficiently, knowledge and access to the server database is necessary. The comparison functions are not sufficiently clear either, with random scores sometimes obtained without a clear explanation.

RecLink II

This is free and open source software developed by two Brazilian professors aiming at conducting record linkage of health data. RecLink II was developed in C++ language and works with input and output files in XBase standard (extension DBF).

The evaluation of RecLink II was based on the information reported in the literature and also on information available at its website. The terms used in the interface of the program indicate that this software was developed to assist users familiar with computing technical terms. In addition, RecLink II is not intuitive, so that non-experienced users need help from a tutorial.

RecLink II has the advantage of allowing implementation of any algorithm to make comparisons as long as the user feeds the program with an input file containing the necessary algorithms. On the other hand, RecLink II has the disadvantage of not allowing implementation of the EM-algorithm to estimate the M and U probabilities used to calculate the agreement and disagreement values.

Registry PlusTM Link Plus

Link Plus is free and open source software developed to perform probabilistic record linkage to support the National Program of Cancer Registries (NPCR) of the United States Center for Diseases Control (CDC). The evaluation of Link Plus was based on the examination of the *User Guide version 2.0* (available through the “Help” tab or the F1 button) and on tests with simulated and real data.

Link Plus was used to detect duplicates in both Census and PES databases and to link records between the two databases. Although it has been originally developed to

be used in cancer registries, the program can be used with any type of data with fixed width or delimited format, which allows its use in the Brazilian PES.

The Graphical User Interface of Link Plus is very user friendly, showing all setting fields on one screen. Its output files show the matched records together, showing also the scores associated to them, which facilitates the verification of the quality of the matches.

The blocking mechanism of Link Plus has only one exact method and two phonetic methods (Soundex and NYSIIS – New York State Identification and Intelligence System) to make comparison blocks. The program has the advantage to implement the scheme “OR blocking”, which compares pairs formed in blocks based on at least one blocking variable. Most programs use the scheme “AND blocking”.

Since Link Plus was designed to be operated by non-advanced users, who are usually not familiar with the technical terms of matching theories, the field for choice of comparison methods presents illustrative terms instead of metric names. The available matching metrics for comparing alphabetical strings can be summarized in two methods: Jaro-Winkler distance (Winkler, 1990) Last Name and First Name methods and Levenshtein distance (Generic String method).

The most basic comparison method implemented by Link Plus is the Exact, which returns all the agreement weights when the field agree or returns the full disagreement weight when the field does not agree. The Middle Name method only accounts for the occurrence of the middle name initial character versus the full middle name. The Value-Specific method associates lower weights to frequent values and higher weights to rare values in a database. SSN and Zip Code methods are designed to compare the Social Security Number and the Zip Code of the United States. This information is not collected by the Brazilian PES and its usefulness for data from other countries is not known. The Date method is used to verify if the dates compared are the same, checking also for omissions or transpositions.

Link Plus presents two options for obtaining the values of agreement and disagreement, which are used in its linkage process. The first option is related to the “Direct Method”, which uses default or user-defined M-probabilities. The second option calculates the M-probability using the EM-algorithm based on the current data.

Link Plus is distinguished by the speed that it performs the entire process, but the program does not allow setting the parameters of the EM-algorithm. This may be one of the reasons why the values of M and U probabilities calculated by Link Plus do not coincide with those calculated by other programs that also implement the EM-algorithm.

Freely Extensible Biomedical Record Linkage - FEBRL

This is open source software developed in collaboration by the Australian National University in Canberra and the New South Wales Department of Health in Sydney, Australia, for cleaning, standardisation, deduplication and record linkage within the domain of health. The assessment of FEBRL was based on the examination of the Graphical User Interface Manual - Release 0.4.1 and on tests with simulated and real data. The experience of Brazilian researchers and staff from the Australian Bureau of Statistics were also accounted for in the evaluation process.

The interest of the PES team was to assess the deduplication and linkage tools. The standardization tool was not studied because data collection was planned to allocate each variable in its respective field, which has greatly reduced the need for treatment of the data.

The FEBRL Graphical User Interface uses a system of tabs to help the user set rules and parameters to perform the record linkage. These tabs open only after confirmation of the data entered via the “Execute” button, which is tricky and makes the users spend time to get used to it.

Through the “Index” tab, FEBRL can implement six different methods for holding the blocking, also allowing the comparison of records without performing the blocking. Through the “Compare” tab, FEBRL can implement twenty six metrics to perform the comparison of each field in the record. The “Classify” tab can implement six weight vector classification methods.

The main advantage of FEBRL is the implementation of many functions for blocking, comparing and indexing. Thus, the software meets the needs of the Brazilian PES based on its methodological aspects. In addition, FEBRL has three other advantages: the availability of a graphical interface (GUI), the option to import TXT and CSV file formats and the comparison of information from different fields.

The main disadvantage of FEBRL is that it does not implement the EM Algorithm required for the calculation of the M and U parameters, which are used to determine the agreement and disagreement values for each field. Another disadvantage of FEBRL is the need for further processing of the output files for use in the Brazilian PES.

RELAIS – Record Linkage at ISTAT

Developed by the Italian National Statistics Institute (ISTAT), RELAIS is free software that has been examined and tested by the 2010 PES team. This section presents general information about the software, the Brazilian experience with the program and considerations on its usage.

The computer languages used by RELAIS are: Java, which is a language associated with objects, and R, a functional language for computing techniques associated with matching data process. Moreover, RELAIS has an architecture based on MySQL

environment, which allows the use of different databases.

The software matches data from two different databases, so there are several options available for that, such as: reduction of the comparison space, choice of matching methods, comparison functions, among others. There are also three matching methods: deterministic, deterministic rule-based and probabilistic based on the Fellegi and Sunter theory (Fellegi and Sunter, 1969). In addition, there are seven different comparison functions.

The focus of the program is the matching, therefore no other processing functions, such as data standardization, cleansing, deduplication, merge, etc., are available. Any adjustment in the databases used in the matching process should be done through another tool. This requires creating new databases and making another upload of RELAIS.

The matching results are displayed in tables and can be saved in text file format (.txt). These tables show all the variables common to the two databases, even if they were not used in the process of matching. Furthermore, it is possible to generate files with the records that were not matched.

In the Brazilian 2010 PES, RELAIS was tested with real data of two dress rehearsals of the 2010 Census and PES. The entire automatic matching for this exercise was done using RELAIS and its effectiveness was verified.

RELAIS is easy to use, provides good results for matching deterministic models and has a good interface to visualize the results. The software implements the Jaro-Winkler comparison functions and also optimizations for one-to-one match, following the premise of all the Post Enumeration Surveys, *i.e.* a record from the Census refers to only one record from PES and vice versa.

Problems occur when performing probabilistic matching. In this case, RELAIS presents rigid requirements related to the size and quality of the data file and to the quality of the comparison variables used in the matching process. Scannapieco (2008) shows that the parameter estimation used in the probabilistic match is not reliable when the agreement weight equals zero and disagreement weight equals 1 for at least one of the comparison variables. Thus, RELAIS sends an error message, stops the matching processing and does not show the results.

RELAIS presents a clear methodology and has most of the comparison functions necessary for the Brazilian PES built in. Besides, the software allows estimation of the U and M probabilities using the EM algorithm and provides output that can be used in the subsequent stages of the process, such as assisted matching and reconciliation.

3. Some Remarks on the Assessment of the Packages

Most of the software considered here were developed for general purposes of record linkage, e.g. producing a comprehensive data base of clients or users of health services so that not only the requirements of the packages are met but also the profile of the users were taken into account. According to their scopes, some functionality was more or less developed. RELAIS is an exception since it has been developed to be used in a PES project and, for this reason, was the most user friendly software for the Brazilian PES context.

The corporate software (SAS and IBM) provide more information on their own structure and results, whereas non corporate ones (Reclink, Likplus, FEBRL and RELAIS) present a clearer view on the methodology. Thus, users can evaluate which one can meet better the methodological requirements of the project in course. Besides, there are also some differences on the interface, documentation and requirements among the corporate and open and free source software. The non-corporate software present user friendly interface and allow the users to perform a record linkage without the need for extra technical assistance.

Most of the non-corporate software provides good documentation that can guide users not only in the use of the software but also on the choice of the functions and metrics to be used. In addition, all of them are available by download, allowing to test their suitability to the project in course.

Considering the Brazilian PES needs and the profile of its team, a comparison between corporate and non-corporate software was performed, accounting for the following aspects:

- Documentation: assessment of the information available on both the use of the software and the mathematical models implemented. It was specially taken into account the clear explanation about the available functions for blocking, comparing and indexing.
- Interface: evaluation of how user friendly and intuitive the user graphical interface is.
- Features: assessment of functions to perform complete matching considering a project such as a PES, which includes the need for deduplication and matching features.
- Requirements: accounting for the need of additional costs such as other software, training and technical assistance.

Fig 1 – Rating of software studied

Software	Documentation	Interface	Features	Requirements
Data Quality	☹	☹	☹	☹
Quality Stage	☺	☺	☺	☹
RecLink	☺	☺	☺	☺
LinkPlus	☺	☺	☺	☺
FEBRL	☺☺	☺☺	☺☺☺	☺
RELAIS	☺☺☺	☺☺☺	☺☺☺	☺

Positive: ☺ Negative: ☹ Not evaluated: ☹

Among the software studied, FEBRL and RELAIS deserve some highlights. FEBRL makes available a large number of functions for blocking, comparing and indexing. RELAIS allows performing a complete matching, which fits the Brazilian PES since it implements the EM algorithm and provides fitted output for assisted matching and reconciliation. However, in the Brazilian PES project, matching has to be done by enumeration area as soon as the data collection is finished. This is due to the need for immediate reconciliation. Thus, it was necessary to develop a batch program that could be inserted into the production process.

REFERENCES

ALUR, N., et al. IBM WebSphere QualityStage Methodologies, Standardization, and Matching. Armonk, 2008. <http://www.redbooks.ibm.com/redbooks/pdfs/sg247546.pdf>.

ANDREAS Borg and MURAT Sariyar. RecordLinkage: Record Linkage in R. R package version 0.2-2. 2010.

<http://CRAN.R-project.org/package=RecordLinkage>.

CAMARGO Jr, KR and COELI, CM: Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. Cad. Saúde Pública, Rio de Janeiro, 2000.

<http://www.scielo.br/pdf/csp/v16n2/2093.pdf>.

CIBELLA, N., et al. Sharing Solutions for Record Linkage: the RELAIS Software and the Italian and Spanish Experiences. Bruxellas, 2009.

http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S7P2_SHARING_SOLUTIONS_FOR_RECORD_LINKAGE_CIBELLA_TUOTO_.pdf.

CHRISTEN, P. Febrl – Freely Extensible Biomedical Record Linkage. Release 0.4.1. Canberra, Australia. 2008.

CHRISTEN, P. Febrl – An Open Source Data Cleaning, Deduplication and Record Linkage System with a Graphical User Interface. Canberra, Australia.

DINIZ DA SILVA, A., et al. Inovações no Sistema de Pareamento de Domicílios e Pessoas para a Pesquisa de Avaliação da Cobertura da Coleta do Censo 2010. Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2010.

FELLEGI, Ivan and SUNTER, Alan. A Theory for Record Linkage. Journal of the American Statistical Association. 1969.

INTERNATIONAL BUSSINESS MACHINES. InfoSphere QualityStage for z/OS.
<http://www-142.ibm.com/software/products/br/pt/ibminfoqualforzos/> .

LINK PLUS User's Guide. Version 2.0. Center for Diseases Control. Atlanta, 2007.

RECLINK Downloadable Files, Publications Related to RecLink II and Educational Material.
<http://www.iesc.ufri.br/reclink/>.

ROMEO, O. Situação do Pareamento Automático após a PA Experimental de Rio Claro/SP. Diretoria de Pesquisas, IBGE. Rio de Janeiro, 2010.


SCANNAPIECO, M., et al. Relais user's guide 2.0. Roma, 2008.
http://www.istat.it/strumenti/metodi/software/MTSFload/ALTRIlload/RELAISload/manual_relais_2_0.pdf.

TUOTO, T., et al. RELAIS: Don't Get Lost in a Record Linkage Project. Arlington, 2007.
<http://www.fcsn.gov/07papers/Tuoto.VI-C.pdf>.

WINKLER, W. E. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proceedings of the Section on Survey Research Methods*, American Statistical Association. 1990.

WINKLER, W. E. Overview of Record Linkage and Current Research Directions. *Research Report Series, RRS*, 2006.

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 John Eltinge at eltinge.john@bls.gov.



Upcoming Conferences and Workshops

The 2012 International Total Survey Error Workshop (ITSEW) will be held in the Netherlands, September 2-4, 2012

The theme of the 2012 International Total Survey Error Workshop is “Total Survey Error: Past, Present, and Future”.

The workshop seeks presentations of research that explore or analyze the trade-offs of two or more sources of survey error within the context of relatively recent methodological or statistical approaches. However, abstracts that deal with the general topic of non-sampling error are welcome. The workshop format is informal to encourage open discussion and sharing of ideas. Thus, presentations discussing research still in progress or results of completed, progress or results of completed, unpublished research are quite appropriate.

Schedule:

September 2 (Sunday afternoon/evening): Arrival/Get together

September 3: Presentations; Evening program Reception/Poster session

September 4: Presentations; Evening Open mike discussion; Farewell

For more general information about the ITSEW, visit the website at <http://nisl05.niss.org/ITSEW/index.html>

Professor Dr. Edith D. de Leeuw
Plantage Doklaan 40, NL-1018 CN Amsterdam
Tel + 31 20 622 34 38
Fax + 31 20 330 25 97
E-mail edithl@xs4all.nl

ICESIV



Fourth International Conference on Establishment Surveys Survey Methods for Businesses, Farms, and Institutions

June 11–14, 2012, at the Sheraton Centre Montréal in Québec Canada

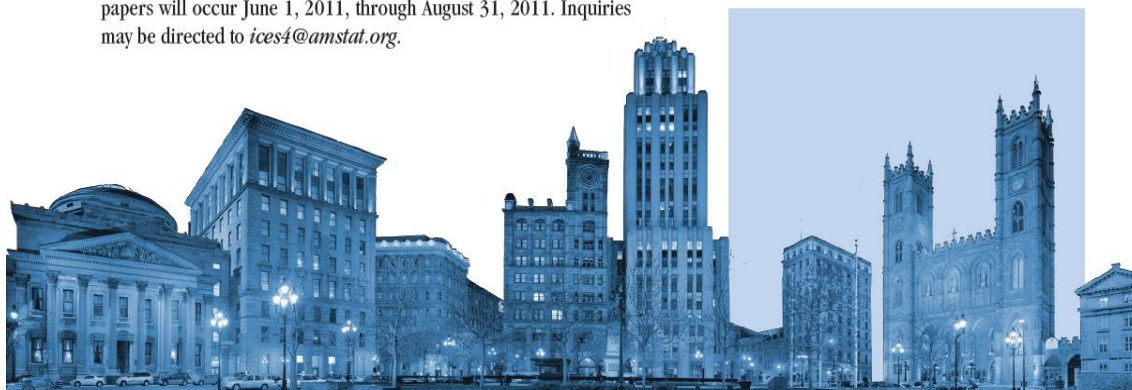
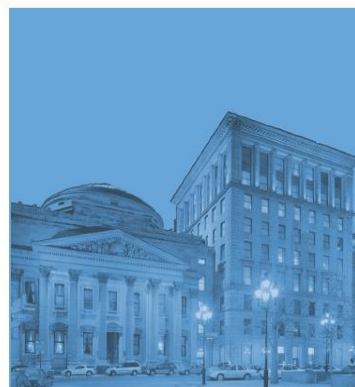
The ICES series of conferences serves an important role in the world of survey methodology. Few other conferences focus on methods and applications for establishment surveys.

The first three conferences—held in 1993, 2000, and 2007—were successful, with more than 400 attendees per conference. The conferences covered a spectrum of survey methods for businesses, farms, and institutions. The fourth conference will continue along this path. Examples of potential topics include the following:



- Updating business registers
- Sample design challenges
- Questionnaire design for establishment surveys
- Collecting data electronically from establishment surveys
- Issues of multi-mode data collection
- Factors that affect establishment survey participation
- Generalized survey processing systems for establishment surveys
- Measuring nonresponse bias
- Variance estimation
- Advances in disclosure protection
- Efficient use of administrative data in establishment surveys
- New directions in establishment surveys

The conference will include short courses, a keynote speaker, poster sessions, software demonstrations, and invited and contributed paper sessions.

Online solicitation for invited papers opens December 1, 2010, and will continue until March 1, 2011. Online solicitation for contributed papers will occur June 1, 2011, through August 31, 2011. Inquiries may be directed to ices4@amstat.org.



Welcome!

	<p>Joint Meeting of</p> <p>y-BIS– International Young Business and Industrial Statisticians</p> <p>and</p> <p>jSPE–Young Portuguese Statisticians</p> <p>Lisbon, Portugal, <u>23 – 26</u> <u>July, 2012</u></p>	
---	--	--

We are pleased to announce that the first meeting of [y-BIS](#) (Young Business and Industrial Statisticians of the [International Society for Business and Industrial Statistics](#)) jointly with the first meeting of [jSPE](#) (Section of Young Statisticians of the [Portuguese Statistical Society](#)), will be held at Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, from **July 23 to 26, 2012**.

The purpose of the **Joint Meeting of y-BIS and jSPE** is to bring together young statisticians and professionals working in Academia and in Industry. The conference will offer opportunities to meet each other, and to share scientific and professional experiences. This meeting will cover many of the research areas in Probability and Statistics, and many areas of application. The session topics include but are not limited to the following areas: Business and Industrial Statistics; Applied Statistics; Computational and Graphical Statistics; Econometrics; Biostatistics; Environmental Statistics and Environmetrics; Bayesian Statistics; Extreme Value Statistics; Multivariate Statistics; Mathematical Statistics; Spatial Statistics; Time Series Analysis; Stochastic Processes; Actuarial Sciences; Probability Theory; Official Statistics; Survey Statistics; Statistical Education.



International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations

October 31–November 3, 2012

Marriott New Orleans at the Convention Center, New Orleans, Louisiana, U.S.A.

Calls for Submissions

February–March 2011:
Invited Call for Submissions

April–May 2011:
Contributed Call for
Submissions

The H2R 2012 conference will bring together survey methodologists, sociologists, statisticians, demographers, ethnographers, and other professionals from around the world to present new and innovative techniques for surveying hard-to-reach populations.

Addressing both the statistical and survey design aspects of including hard-to-reach groups, researchers will report findings from censuses, surveys, and other research related to the identification, definition, measurement, and methodologies for surveying and enumerating undercounted populations.

Identifying, Defining, and Measuring the Hard-to-Reach (HTR)

- Defining HTR populations
- Measuring undercounts for HTR groups
- Improving measurement with administrative records
- Sampling HTR populations

Techniques and Methodologies

- Recruitment methods
- Targeting the HTR
- Use of social marketing and outreach campaigns
- Overcoming language and literacy barriers
- Use of community-based organizations
- Dealing with complex living and housing situations
- Tracking and tracing HTR populations

HTR Subpopulations

- Racial minorities
- Immigrant populations
- Indigenous populations
- Highly mobile and migrant populations
- Homeless and refugee populations
- Sexual minorities
- Populations affected by natural disasters
- Populations in zones of armed conflict
- Stigmatized populations
- Cross-cultural similarities and differences in HTR populations
- Linguistic and cultural minorities

For information, visit www.amstat.org/meetings/h2r/2012 or email H2R2012@amstat.org.

H2R|2012
Survey Methods for Hard-to-Reach Populations

Collection
des actes
des
précédents
colloques
Sondages

Tanger, 2010

Tremblay M.-E., Lavallée P. et El Haj Tiraï M. (dir.)
"Pratique et méthodes de sondage", Dunod. À paraître en 2011.

Marseille, 2007

Guilbert Ph., Haziza D., Ruiz-Gazen A. et Tillé Y. (dir) (2008)
"Méthodes de sondage. Enquêtes électorales, enquêtes dans
le domaine de la santé, enquêtes dans les pays en développement",
Dunod, collection cours et cas pratiques.

Québec, 2005

Lavallée P. et Rivest L.P. (dir) (2006)
"Méthodes d'enquêtes et sondages : Pratiques européenne
et nord-américaine", Dunod, collection cours et cas pratiques.

Autrans, 2002

Ardilly P. (dir) (2004)
"Échantillonnage et méthodes d'enquêtes",
Dunod, collection cours et cas pratiques.

Bruxelles, 2000

Droesbeke J.J. et Lebart L. (dir) (2001)
"Enquêtes, modèles et applications", Dunod.

Rennes, 1997

Brossier G. et Dussaix A.M. (dir) (1999)
"Enquêtes et Sondages. Méthodes, modèles, applications, nouvelles
approches", Dunod.

POSSIBILITES DE BOURSES POUR LES STATISTICIENS DES PAYS EN D

<http://sondages2012.ensai.fr>

Adresse électronique: sondages2012@ensai.fr

Comité scientifique

Olivier Santory (Président - Insee)
Philippe Tassi (Vice-Président - Médiamétrie)

Comité d'organisation

Eric Lesage (Président - Crest - Ensai)
Benoît Riandey (Vice-Président - Ined - SFdS)

Colloque du 05 au 07 novembre 2012

Journées de formation
les 08 et 09 novembre 2012

dans les locaux de l'Ensai,
à Bruz (au sud de Rennes).

<http://sondages2012.ensai.com>

ADRESSE ELECTRONIQUE sondages2012@ensai.fr



POSSIBILITES DE BOURSES POUR
LES STATISTICIENS DES PAYS EN DEVELOPPEMENT



Congress of Polish Statistics To Celebrate the 100th Anniversary of the Polish Statistical Association Poznan, 18-20 April 2012

<http://www.stat.gov.pl/pts/kongres2012/english/index.htm>

Preliminary programme

ANNIVERSARY SESSION

DEVELOPMENT OF POLISH STATISTICAL RESEARCH

EMINENT POLISH STATISTICIANS

- Jerzy Sława Neyman (Prof. Teresa Ledwina, Institute of Mathematics of the Polish Academy of Sciences)
- Jan Czekanowski (Prof. Mirosław Krzyśko, Adam Mickiewicz University)

METHODOLOGY OF STATISTICAL RESEARCH (Prof. Mirosław Krzyśko, Adam Mickiewicz University)

- Development of Statistical Research Methodology
- Small Area Statistics
- Methods of Data Analysis and Classification
- Survey Methodology

REGIONAL STATISTICS (Prof. Tadeusz Borys, University of Economics in Wrocław)

POPULATION STATISTICS (Prof. Janina Józwiak, Warsaw School of Economics)

- New demography of Europe and Poland: challenges for demographic analysis
- Demographic modelling and projecting: factors affecting the demographic reality and future
- Changes in demographic structures: facts and consequences for economy and society
- New standards in population statistics (incl. census, panel studies)
- Historical demography

SOCIAL STATISTICS (Prof. Tomasz Panek, Warsaw School of Economics)

- Poverty, social exclusion and inequality
- Standard and quality of life
- Human capital
- Labour market

ECONOMIC STATISTICS (Prof. Jerzy Wilkin, Warsaw University)

STATISTICAL DATA (Prof. Mirosław Szreder, University of Gdańsk)

HEALTH STATISTICS (Prof. Jerzy Andrzej Moczko, Poznan University of Medical Sciences)

STATISTICS OF SPORTS AND TOURISM (Prof. Bogdan Sojkin, Poznan University of Economics)

DISCUSSION PANELS (Prof. J. Oleński, Warsaw University, Prof. Tadeusz Caliński, Poznan University of Life Sciences)

- Fundamental problems of statistics in the modern world
- The future of statistics

Eighth World Congress in Probability and Statistics

Dear Colleagues,

The eighth World Congress in Probability and Statistics will be in Istanbul from July 9 to 14, 2012. It is jointly organized by the Bernoulli Society and the Institute of Mathematical Statistics. Scheduled every four years, this meeting is a major worldwide event for statistics and probability, covering all its branches, including theoretical, methodological, applied and computational statistics and probability, and stochastic processes. It features the latest scientific developments in these fields.

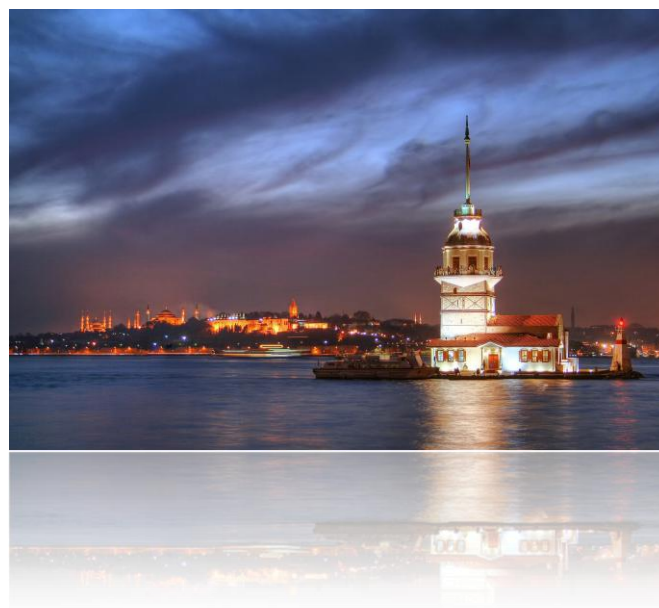
The program will cover a wide range of topics in statistics and probability, presenting recent developments and the state of the art in a variety of modern research topics, with in-depth sessions on applications of these disciplines to other sciences, industrial innovation and society. It will feature several special plenary lectures presented by leading specialists. In addition, there will be many invited sessions highlighting topics of current research interests, as well as a large number of contributed sessions and posters.

The venue of the meeting is Grand Cevahir Hotel & Convention Center located in Istanbul which is a vibrant, multi-cultural and cosmopolitan city bridging Europe and Asia. Istanbul has a unique cultural conglomeration of east and west, offering many cultural and touristic attractions, such as Hagia Sophia, Sultanahmet, Topkapı Palace and Maiden's Tower. On behalf of the Program Committee and the Local Organizing Committee, we invite you to join us in Istanbul for this exciting scientific event. Your participation will ensure that the 2012 World Congress will be a memorable meeting.

Elvan Ceyhan and Mine Caglar,
(Co-chairs of the Local Organizing Committee)

Arnoldo Frigessi,
(Chair of the Program Committee)

<http://www.worldcong2012.org/>



SIS2012

XLVI Scientific Meeting of the ITALIAN STATISTICAL SOCIETY

June 20-22, 2012

Rome, Italy

The Conference will include plenary, specialized, solicited, contributed and poster sessions. There will also be invited sessions from representatives of some European statistical societies.

The Conference Program Committee invites submissions of contributed papers/posters. These can be on any area of interest relevant to theoretical and applied statistics.

1. Submission of a contributed paper

Papers can be submitted on-line. All submissions, except Invited talks, are subject to a blind refereeing process. Information on paper submission, editorial standards and other details will be made available on the conference site.

The deadlines are as follows:

Title submission December 30th, 2011

Paper/Abstract Submission: February 15th, 2012

Author Notification: March 31st, 2012

Final version: May 11st, 2012

2. Submission of a Poster

Researchers who wish to participate without submitting a paper, can send a poster up to the very last minute. The deadline is as follows

Poster	June	8th,	2012.
--------	------	------	-------

Further details will be available on the conference website.

<http://meetings.sis-statistica.org/index.php/sm/sm2012>

Best regards.

The SIS2012 Program Committee



Welcome

Dear Colleague,

The Hellenic Statistical Authority (ELSTAT) and Eurostat are pleased to invite you to the *European Conference on Quality in Official Statistics (Q2012)* which will be held in the "Megaron Athens International Conference Centre" in Athens, Greece, on 30 May - 1 June 2012.

The Quality conferences started in 2001 when the first Conference was held in Stockholm to develop further the conclusions of the Leadership Group on Quality. One of the recommendations of the Conference was to organize Quality Conferences biennially. So far, five Q-conferences have taken place: Q2001 in Stockholm, Q2004 in Mainz, Q2006 in Cardiff, Q2008 in Rome and Q2010 in Helsinki.

Similarly to its predecessors, the Q2012 Conference will offer participants a unique platform for high-level discussions on various types of quality issues and methodological aspects of statistical production. Q2012 will in particular focus on improving quality by the re-engineering of statistical systems and the innovations in the area of quality.

In addition, a series of short training courses will take place on 29 May 2012.

The Conference website offers information on the conference venue, accommodation, short courses and leisure-time activities and deadlines for submission of papers. Registration is now available.

If you have any questions, please contact the Organising Committee by e-mail at q2012@statistics.gr

Looking forward to seeing you in Athens!

ELSTAT

Andreas V. Georgiou
President

Eurostat

Walter Radermacher
Director General

<http://www.q2012.gr/default.asp?p=0>

Joint Statistical Meetings 2012

San Diego

July 28 – August 2 2012

<http://www.amstat.org/meetings/jsm/2012/index.cfm>

Conference Information > Overview

To provide the best and most cohesive program possible, JSM 2012 speakers must commit to their involvement by registering for JSM during the abstract submission process. Details and instructions will be available December 1.

Welcome to the 2012 Joint Statistical Meetings Web site. Follow the tabs at the top of this page to navigate through the site. Comments and suggestions are welcome at meetings@amstat.org. This page is updated frequently, so visit often.

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the [American Statistical Association](#), the [International Biometric Society](#) (ENAR and WNAR), the [Institute of Mathematical Statistics](#), the [Statistical Society of Canada](#), and the [International Chinese Statistical Association](#), and the [International Indian Statistical Association](#). 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.

San Diego, California, the host city for JSM 2012, offers a wide range of options for sharing time with friends and colleagues or sightseeing with family. For information, contact meetings@amstat.org.

The 2012 Joint Statistical Meetings will be held July 28-August 2, 2012, at the [San Diego Convention Center](#), located at 111 West Harbor Drive, San Diego, CA 92101.

ASA Meetings Department • 732 North Washington Street, Alexandria, VA 22314 •
(703) 684-1221 • meetings@amstat.org
Copyright © American Statistical Association.



In Other Journals

Survey Practice

Practical Information for Survey Researcher

[About](#)

[Submit Articles](#)

[Editorial Review](#)

[Guidelines for Authors](#)

[Editorial Board](#)

August 2011

www.surveypractice.org

Contemporary Issues with Public Policy Polls

D. W. Moore

Paradata in Survey Research

B. T. West

A Collection of Caller ID Experiments

D. Dutwin, M. Herrmann, E. B. Porath, S. Sherr

Identifying Non-working Numbers in Cell Phone RDD Samples via HLR-Lookup Technology

B. Struminskaya, L. Kaczmirek, I. Schaurer, W. Bandilla, S. Gabler, S. Haeder

Are Advance Postcards or Letters Better? An Experiment with Advance Mailing Types

A. Richardson



[Home](#) > **Vol 5, No 2 (2011)**

Survey Research Methods

The journal is edited by Jaalo Billiet of K.U. Leuven, Belgium and Rainer Schnell of the University of Duisburg-Essen, Germany.

<http://w4.ub.uni-konstanz.de/srm/>

Estimating underreporting of consumer expenditures using Markov latent class analysis

C. Tucker, P. P. Biemer, B. Meekins

Household survey panels: how much do following rules affect sample size?

M. Schonlau, N. Watson, M. Kroh

Measurement Quality in Indicators of Compositions. A Compositional Multitrait-Multimethod Approach

G. Coenders, V. Hlebec, T. Kogovšek

Reports of relationship timing: Missing data and couple agreement

A. Reimondos, A. Evans, E. Gray

The logo for the Journal of Official Statistics (JOS) features the letters 'JOS' in a stylized, serif font. The 'J' and 'O' are connected, and the 'S' is separate. The letters are a dark red or maroon color.

Journal of Official Statistics

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

Current Issue: September 2011, Vol. 27 No. 3

Published Date : 27-September-2011

A Unit-Error Theory for Register-Based Household Statistics

L. Zhang

Using Statistical Models for Sample Design of a Reinterview Program

J. Li, J. M. Brick, B. Tran, P. Singer

Centre Sampling Technique in Foreign Migration Surveys: A Methodological Note

G. Baio, G. C. Blangiardo, M. Blangiardo

Algorithms for Correcting Sign Errors and Rounding Errors in Business Survey Data

S. Scholtus

Bayesian Estimation of Population Size via Linkage of Multivariate Normal Data Sets

B. Liseo, A. Tancredi

Random Group Variance Estimators for Survey Data with Random Hot Deck Imputation

J. Shao, Q. Tang

Statistical Properties of Multiplicative Noise Masking for Confidentiality Protection

T. K. Nayak, B. Sinha, L. Zayatz

Book Reviews

In Other Journals



Survey Methodology
A Journal Published by Statistics Canada
Volume 37, Number 1, June 2011

<http://www.statcan.gc.ca/pub/12-001-x/12-001-x2011001-eng.htm>

Nonsampling Errors in Dual Frame Telephone Surveys

J.M. Brick, I.F. Cervantes, S. Lee, G. Norman

Maximum Likelihood Estimation for Contingency Tables and Logistic Regression with Incorrectly Linked Data

J.O. Chipperfield, G.R. Bishop, P. Campbell

Hierarchical Bayes Small Area Estimation Under a Spatial Model with Application to Health Survey Data

Y. You, Q.M. Zhou

Small Area Estimation Under Transformation to Linearity

H. Chandra, R. Chambers

The Construction of Stratified Designs in R with the Package Stratification

S. Baillargeon, L.P. Rivest

Replication Variance Estimation Under Two-Phase Sampling

J. K. Kim, C.L. Yu

Cost Efficiency of Repeated Cluster Surveys

S. Kolenikov, G. Angeles

On the Efficiency of Randomized Probability Proportional to Size Sampling

P. Kottnerus

Short Notes:

The Use of Estimating Equations to Perform a Calibration on Complex Parameters

E. Lesage



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

Volume 27, Number 3-4 / 2011

<http://iospress.metapress.com/content/k5pl76h14441/?p=02f85335aed5467d97528fe04e189baa&pi=0>

How statistical literacy, official statistics and self-directed learning shaped social enquiry in the 19th and early 20th centuries

G. A. Lancaster

Official Statistics and statistical literacy: They need each other

S. Forbes, M. Camden, N. Pihama, P. Bucknall, M. Pfannkuch

The national statistical agency as educator

Mary Townsend

Creating statistically literate global citizens: The use of IPUMS-International integrated census microdata in teaching

A. Meier, R. McCaa, D. Lam

Statistical literacy and awareness as strategic success factors of a national statistical office – the case of Statistics Finland

R. Helenius, H. Mikkela

The millennium development goals, national statistical offices, the international statistical literacy project and statistical literacy in schools

J. Sanchez, S. Forbes, P. Campos, P. Giacche, M. Townsend, G. Mooney, R. Helenius

Statistical literacy: A new mission for data producers

M. Schield

Responding to diversity in users' statistical literacy and information needs: Institutional and educational implications

I. Gal, S. T. Murray

Foundations for improving statistical literacy

J. M. Watson

Developments of AtSchool projects for improving collaborative teaching and learning in statistics

N. Davies

Does CensusAtSchool develop statistical literacy?

I. Gal

Discussion

C. K. Wild



ISI - International Statistical Institute

Volume 79, Issue 2, August 2011

<http://onlinelibrary.wiley.com/doi/10.1111/insr.2011.79.issue-2/issuetoc>

Adjusting for Non-Response in Population-Based Case-Control Studies

Y.Jiang, A. J. Scott, C. J. Wild

Risk-Utility Paradigms for Statistical Disclosure Limitation: How to Think, But Not How to Act

L.H. Cox, A.F. Karr, S. K. Kinney

Discussion

J. Domingo-Ferrer

Discussion

G.T. Duncan

Discussion

C.M. O'Keefe

Discussion

N. Shlomo

Rejoinder

L.H. Cox, A.F. Karr, S.K. Kinney

Connections Between Survey Calibration Estimators and Semiparametric Models for Incomplete Data

T. Lumley, P.A. Shaw, J.Y. Dai

Discussion

A.A. Tsiatis, M. Davidian

Discussion

M.S. Handcock

Discussion

J.F. Lawless, J.D. Kalbfleisch

Discussion

A.J. Scott, C.J. Wild

Rejoinder

T. Lumley

Three Factors to Signal Non-Response Bias with Applications to Categorical Auxiliary Variables

C.E. Sarndal

Autocorrelation Functions

R. Finlay, T. Fung, E. Seneta

Short Book Reviews



STATISTICS IN TRANSITION *new series*

An International Journal of the Polish Statistical Association

Volume 12, Number 1, August 2011

http://www.stat.gov.pl/pts/15_ENG_HTML.htm

Subsampling the nonrespondents in cluster sampling on sampling on two successive occasions

H. P. Singh, S. Kumar

Some rotation patterns in two-phase sampling

H. N. Singh, S. Prasad

Estimation of population mean using multi-auxiliary characters with subsampling the nonrespondents

B. B. Khare, R. R. Sinha

Estimation of average income in Cuban municipalities

N. A. Montes De Oca

Bootstrap method with calibration for standard error estimators of income poverty measures

A. Zięba-Pietrzak, J. Kordos, R. Wieczorkowski

The beta Pareto distribution

A. Hurairah

Robustness of the confidence interval for at-risk-of-poverty rate

W. Zieliński

Estimation of domain means on the basis of strategy dependent on depth function of auxiliary variables' distribution

J. L. Wywiał

Remarks about the generalizations of the Fisher index

Bialek J.

Multiple-equation models of ordered dependent variables in exploration of the results of locomotive organ disorders

J. Grosman, M. Kowerski

Identifying an appropriate forecasting model for forecasting total import of Bangladesh

T. Khan

Does a change of occupation lead to higher earnings?

B. Liberda, M. Pęczkowski

Neonatal Mortality by gestational age and birth weight in Italy, 1998 – 2003: a record linkage study

C. Marini, A. Nuccitelli

Book Review: Sampling: Design and Analysis. Sharon L. Lohr. 2nd Edition, International Publication 2010

J. Kordos

ISI Satellite Conference on Improving Statistical Systems Worldwide – Building Capacity Information on the Congress of Polish Statistics to Celebrate the 100th Anniversary of the Polish Statistical Association



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

Current Issue: Volume 3, Issue 1 (2011)

Editorial: In This Issue

Stephen E. Fienberg

An Analysis of Google Logs Retention Policies

V. Toubiana, H. Nissenbaum

Privacy Violations Using Microtargeted Ads: A Case Study

A. Korolova

On Privacy and Public Data: a Study of data.gov.uk

A. C. Simpson

Protecting Consumer Privacy in an Era of Rapid Change—A Proposed Framework for Businesses and Policymakers

FTC Staff

Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework

Department of Commerce Internet Policy Task Force

**TRANSACTIONS ON
DATA PRIVACY**

Foundations and Technologies

<http://www.tdp.cat>

Volume 4, Issue 2, August 2011

<http://www.tdp.cat/issues11/vol04n02.php>

**Forward for the Special Issue of Selected Papers from the 3rd ACM
SIGSPATIAL Workshop on Security and Privacy in GIS and LBS**

Y. Saygin, M.L. Damiani

**Third Party Geo-location Services in LBS: Privacy Requirements and Research
Issues**

M.L. Damiani

C-safety: A Framework for the Anonymization of Semantic Trajectories

A. Monreale, R. Trasarti, D. Pedreschi, C. Renso, V. Bogorny

Show Me How You Move and I Will Tell You Who You Are

S. Gambs, M.O. Killijian, M. Nunez del Prado Cortez

BIOMETRIKA

Biometrika Vol. 98, Issue 3, September 2011

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

Sample size formulae for two-stage randomized trials with survival outcomes

Z. Li, S. A. Murphy

**The covariate-adaptive biased coin design for balancing clinical trials in the
presence of prognostic factors**

A. B. Antognini, M. Zagoraiou

Bayesian isotonic density regression

L. Wang, D. B. Dunson

A class of mixtures of dependent tail-free processes

A. Jara, T. E. Hanson

Semiparametric inference in mixture models with predictive recursion marginal likelihood

R. Martin, S. T. Tokdar

Functional mixed effects spectral analysis

R. T. Krafty, M. Hall, W. Guo

Testing parametric assumptions of trends of a nonstationary time series

T. Zhang, W. B. Wu

Aggregation-cokriging for highly multivariate spatial data

R. Furrer, M. G. Genton

A construction principle for multivariate extreme value distributions

F. Ballani, M. Schlather

Marginal methods for correlated binary data with misclassified responses

Z. Chen, G. Y. Yi, C. Wu

Efficient restricted estimators for conditional mean models with missing data

Z. Tan

Conditional Akaike information under generalized linear and proportional hazards mixed models

M. C. Donohue, R. Overholser, R. Xu, and F. Vaida

Generalized varying coefficient models with unknown link function

C. N. Kuruwita, K. B. Kulasekera, and C. M. Gallagher

Sudoku-based space-filling designs

X. Xu, B. Haaland, and P. Z. G. Qian

Nested orthogonal array-based Latin hypercube designs

X. He and P. Z. G. Qian

Robust designs through partially clear two-factor interactions

R. Lekivetz and B. Tang

Construction of ϕ_p -optimal exact designs with minimum experimental run size for a linear log contrast model in mixture experiments

B. Jin and M.N. Lo Huang

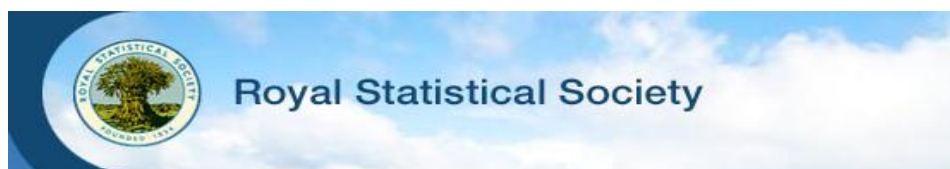
Miscellanea:

On protected estimation of an odds ratio model with missing binary exposure and confounders

E. J. Tchetgen and A. Rotnitzky

Multinomial logit bias reduction via the Poisson log-linear model

I. Kosmidis and D. Firth



Journal of the Royal Statistical Society: Series A

October 2011

Volume 174, Issue 4

<http://onlinelibrary.wiley.com/doi/10.1111/rssa.2011.174.issue-4/issuetoc>

Editorial: The measurement of progress – some achievements and challenges

P. Anand, M. Durand, J. Heckman

A multilevel modeling approach to investigating the predictive validity of editorial decisions: do the editors of a high profile journal select manuscripts that are highly cited after publication?

L. Bornmann, R. Mutz, W. Marx, H. Schier, H.D. Daniel

Design of the 2001 and 2011 Census Coverage Surveys for England and Wales

J. Brown, O. Abbott, P.A. Smith

Surviving slavery: mortality at Mesopotamia, a Jamaican sugar estate, 1762-1832

M. Forster, S.D. Smith

Small area estimation using a reweighting algorithm

R. Tanton, Y. Vidyattama, B. Nepal, J. McNamara

Combining the intensity and sequencing of the poverty experience: a class of longitudinal poverty indices

D. Mendola, A. Busetta, A.M. Milito

A stochastic model for assessing Chlamydia Trachomatis transmission risk by using longitudinal observational data

W. Tu, P. Ghosh, B.P. Katz

The effect of survey design on household reporting of financial difficulty

R. Breunig, R. McKibbin

New Orleans business recovery in the aftermath of Hurricane Katrina

J.P. LeSage, R.K. Pace, N. Lam, R. Campanella, X. Liu

Using paradata to predict best times of contact, conditioning on household and interviewer influences

G.B. Durrant, J. D'Arrigo, F. Steele

Modelling member behavior in on-line user-generated content sites: a semiparametric Bayesian approach

Y.H. Park, C.H. Park, P. Ghosh

Investigating the behavioural characteristics of lottery players by using a combination preference model for conscious selection

R. Baker, I.G. McHale

Book Reviews

Journal of the American Statistical Association



Current Issue

[Volume 106, Number
495](#)

<http://pubs.amstat.org/toc/jasa/106/495>

Applications and Case Studies

Population Value Decomposition, a Framework for the Analysis of Image Populations

C. M. Crainiceanu, B.S. Caffo, S. Luo, V. M. Zipunnikov, N. M. Punjabi

Comment

N. A. Lazar

Comment

K. Shedden

Comment

E. F. Lock, A. B. Nobel, J. S. Marron

Comment

Y. N. Wu

Rejoinder

C. M. Crainiceanu, B.S. Caffo, S. Luo, V. M. Zipunnikov, N. M. Punjabi

Nonparametric Bayes Stochastically Ordered Latent Class Models

H. Yang, S. O'Brien, D. B. Dunson

Robust EM Continual Reassessment Method in Oncology Dose Finding

Y. Yuan, G. Yin

Analysis of Long Period Variable Stars With Nonparametric Tests for Trend Detection

C. Park, J. Ahn, M. Hendry, W. Jang

Fast and Accurate Approximation to Significance Tests in Genome-Wide Association Studies

Y. Zhang, J.S. Liu

Functional Principal Component Analysis of Density Families With Categorical and Continuous Data on Canadian Entrant Manufacturing Firms

K. P. Huynh, D. T. J. Chávez, R. J. Petrunia, M. Voia

Distinct Counting With a Self-Learning Bitmap

A. Chen, J. Cao, L. Shepp, T. Nguyen

A Statistical Framework for the Analysis of ChIP-Seq Data

P. F. Kuan, D. Chung, G. Pan, J. A. Thomson, R. Stewart, S. Keleş

Theory and Methods

Adaptive Confidence Intervals for the Test Error in Classification

E. B. Laber, S.A. Murphy

Comment

R. J. Samworth

Comment

P. Bühlmann

Comment

L. Dümbgen

Comment

Y. Goldberg, M. R. Kosorok

Comment

Y. Yang, G. Cheng

Comment

S. Wei, A. B. Nobel

Comment

W. Polonik

Rejoinder

E. B. Laber, S. A. Murphy

**An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations
With Application to Model Checking**

M. A. Delgado, C. Velasco

**Variational Bayesian Inference for Parametric and Nonparametric Regression
With Missing Data**

C. Faes, J. T. Ormerod, M. P. Wand

Bayesian Inference for the Spatial Random Effects Model

E. L. Kang, N. Cressie

**Posterior Probability of Discovery and Expected Rate of Discovery for Multiple
Hypothesis Testing and High Throughput Assays**

Y. Zhao

Correcting for Population Stratification in Genomewide Association Studies

D. Y. Lin, D. Zeng

**Independent Component Analysis Involving Autocorrelated Sources With an
Application to Functional Magnetic Resonance Imaging**

S. Lee, H. Shen, Y. Truong, M. Lewis, X. Huang

Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches

M. Tao, Y. Wang, Q. Yao, J. Zou

Rao–Blackwellization for Bayesian Variable Selection and Model Averaging in Linear and Binary Regression: A Novel Data Augmentation Approach

J. Ghosh, M. A. Clyde

Optimal Weight Choice for Frequentist Model Average Estimators

H. Liang, G. Zou, A. T. K. Wan, X. Zhang

Permutation Multiple Tests of Binary Features Do Not Uniformly Control Error Rates

E. E. Kaizar, Y. Li, J. C. Hsu

Hierarchical Clustering With Prototypes via Minimax Linkage

J. Bien, R. Tibshirani

Summarizing Insurance Scores Using a Gini Index

E.W. Frees, G. Meyers, A. D Cummings

Linear or Nonlinear? Automatic Structure Discovery for Partially Linear Models

H. H. Zhang, G. Cheng, Y. Liu

A Measure of Stationarity in Locally Stationary Processes With Applications to Testing

H. Dette, P. Preuß, M. Vetter

SparseNet: Coordinate Descent With Nonconvex Penalties

R. Mazumder, J.H. Friedman, T. Hastie

Multi-Layer Designs for Computer Experiments

S. Ba, V. R. Joseph

GLS Estimation of Dynamic Factor Models

J. Breitung, J. Tenhofen

Robust, Adaptive Functional Regression in Functional Mixed Model Framework

H. Zhu, P. J. Brown, J. S. Morris

Density Estimation in Several Populations With Uncertain Population Membership

Y. Ma, J. D. Hart, R.J. Carroll

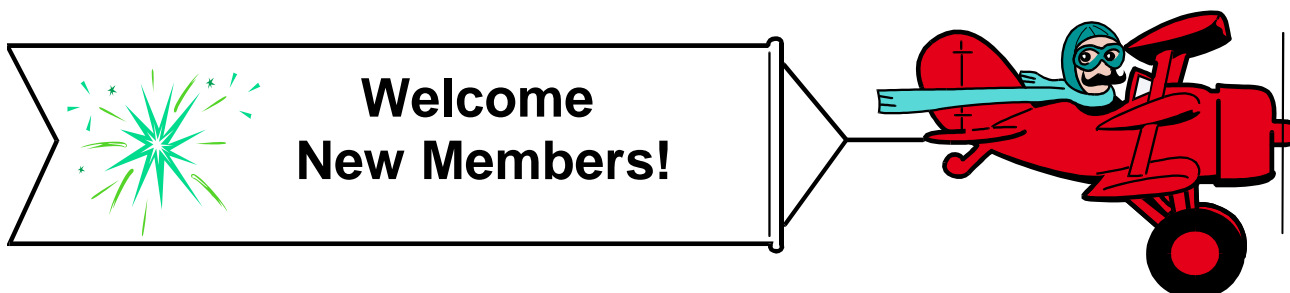
Fusion-Refinement Procedure for Dimension Reduction With Missing Response at Random

X. Ding, Q. Wang

Predicting False Discovery Proportion Under Dependence

S. Ghosal, A. Roy

**General
Book Reviews**



We are very pleased to welcome the following new members and institutional members

Country	Surname	Name
ARMENIA, Republic of	Movsisyan	Vahe
ARMENIA, Republic of	Hayrapetyan	Lusine
AUSTRALIA	Lokhorst	Justin
AUSTRIA	Glaser	Thomas
BULGARIA	Vesselinov	Roumen
CROATIA	Erjavac	Natasa
FRANCE	Brouard	Nicolas
FRANCE	Randrianasolo	Toky
GERMANY	Schmid	Timo
GRENADA	Clarkson	Margaret Beryl-Ann
ITALY	Salvatore	Renato
ITALY	Montinaro	Mario
ITALY	Srivastava	Mukesh Kumar
ITALY	Salvati	Nicola
ITALY	Benedetti	Roberto
ITALY	Coli	Alessandra
ITALY	Pagliarella	Maria Chiara
ITALY	Brunero	Liseo
ITALY	Bianchi	Annamaria
ITALY	De Paola	Rosita
ITALY	Bocci	Chiara
ITALY	Fattorini	Lorenzo
ITALY	L'Abbate	Samuela
ITALY	MartelliBianca	Maria
ITALY	Bramati	Maria Caterina
ITALY	Piersimoni	Federica
KOREA, Republic of	Kim	Hoil
MACEDONIA, Former Yugoslav Republic of	Novkovska	Blagica
NEW ZEALAND	Gosse	Michelle
NORWAY	Wold	Bjørn Kjetil G.
ROMANIA	Ionescu	Alina Mariuca
SWEDEN	Holmberg	John Seved Anders
TANZANIA, United Republic of	Maro	Radegunda Honest
UNITED ARAB EMIRATES	Alhosani	Juma Abdulla
UNITED KINGDOM	Sadig	Husam
UNITED STATES	Kuczynski	Margaret
UNITED STATES	Datta	Gauri Sankar
UNITED STATES	Himelein	Kristen
UNITED STATES	Phipps	Polly

Country	Department	Institution
UNITED STATES	National Agricultural Statistics Service	U.S. Department of Agriculture
UNITED STATES	Survey Research Center	University of Michigan

IASS Officers and Council Members

Executive Officers

President (2011-2013):	Ray Chambers (Australia)	ray.chambers@internode.on.net
President-elect:	Danny Pfeffermann (Israel)	d.pfeffermann@soton.ac.uk
Vice-Presidents (2011-2013):	Denise Britz do N. Silva (Brazil)	denise.silva@ibge.gov.br
	Steven Heeringa (USA)	sheering@isr.umich.edu
Scientific Secretary (2011-2013):	Ineke Stoop (The Netherlands)	i.stoop@scp.nl
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)	n.shlomo@soton.ac.uk
	Marcel de Toledo Vieira (Brazil)	marcel.vieira@ufjf.edu.br
	Alvaro Gonzalez Villalobos (Argentina)	alvarun@fibertel.com.ar
Council Members (2009-2013):	Mike Hidioglou (Canada)	mike.hidioglou@statcan.gc.ca
	Edith D. de Leeuw (The Netherlands)	edithl@xs4all.nl
	Monica Pratesi (Italy)	m.pratesi@ec.unipi.it
	Mick P. Couper (USA)	mcouper@umich.edu
	Eva Elvers (Sweden)	eva.elvers@scb.se
	Yves Tille (Belgium)	yves.tille@unine.ch

Committee Chairs

Hong Kong 2013 Programme Committee	Eric Rancourt (Canada)	eric.rancourt@statcan.gc.ca
Chair of the jury of the 2013 Cochran-Hansen Prize	Yves Tille (Belgium)	yves.tille@unine.ch

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 Organizations

AFRISTAT
EUROSTAT

16 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
CANADA – STATISTICS CANADA
CHINA – DIRECCAO DOS SERVICOS DE ESTATISTICA
DENMARK – DANMARKS STATISTIK
FINLAND – STATISTICS FINLAND
GERMANY – STATISTISCHE BUNDESAMT
ITALY – INSTITUTO CENTRALE DI STATISTICA
KOREA, REPUBLIC OF – KOREA NATIONAL STATISTICAL OFFICE
MEXICO – INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA
MAURITIUS – CENTRAL STATISTICAL OFFICE
NEW ZEALAND – STATISTICS NEW ZEALAND
NORWAY – STATISTICS NORWAY
PORTUGAL – INSTITUTO NACIONAL DE ESTADÍSTICA
SWEDEN – STATISTISKA CENTRALBYRÅN
USA – NATIONAL AGRICULTURAL STATISTICAL SERVICE

4 Universities, Research Centers, Private Statistics Firms

USA – RESEARCH TRIANGLE INSTITUTE
USA – SURVEY RESEARCH CENTER, UNIVERSITY OF MICHIGAN
USA – U.S. DEPARTMENT OF AGRICULTURE
USA – WESTAT

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

CHANGE OF ADDRESS FORM



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

**IASS Secretariat Membership Officer
Margaret de 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>