



STATISTICAL SOCIETY OF AUSTRALIA INCORPORATED NEWSLETTER

 **September 2011**
Number 136

Young Statisticians' Conference 2011 – Learners Today, Leaders Tomorrow!

IN THIS ISSUE

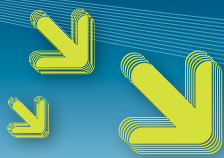
Young Statisticians' Conference 2011—Learners Today, Leaders Tomorrow!.....	1
Editorial.....	2
Events.....	2
From the President.....	3
Young Statisticians' Conference 2011 story cont.....	6
Young Statisticians' Conference 2011—A Reflection.....	7
Thanks from the Program Chair...8	
International Engagement.....	9
Adaptive Designs For Clinical Trials Workshop.....	10
Australian Statistical Conference 2012.....	11
Science Meets Parliament.....	12
Branch News.....	13
Three Statistics Courses.....	28



Young Statisticians at the Conference Dinner

The YSC2011 was held at St. John's College, University of Queensland St Lucia campus in Brisbane. It was well attended by more than 50 young statisticians from around Australia and New Zealand. Besides 32 contributed talks, we had four keynote addresses: Professor Kerrie Mengerson from Queensland University of Technology, Dr. Ross Darnell from CSIRO, Professor Mike Daniels from University of Florida and Associate Professor Dongseok Choi from Oregon Health and Science University, as well as a careers session jointly presented by Dr Ross Darnell and Dr Sanjoy Paul from Queensland Clinical Trials & Biostatistics Centre in University of Queensland.

All four keynote talks were inspiring and informative. Kerrie, on behalf of BRAG (Bayesian Research and Applications Group) co-authors, spoke about Bayesian modelling of epidemics. She illustrated the infectiousness of epidemics by inviting delegates to join her in an "infection game". Ross talked about sampling issues in measuring biodiversity, demonstrating substantial sampling issues in the context of marine surveys. Mike gave a brief introduction of model-based approaches to handle missing data in longitudinal studies. Dongseok shared his experience of using clustering methods for clinical trials and genetics data. >> pg6



September



PO Box 213, Belconnen ACT 2616
We are located on the ground floor of
ABS House, room GN 311.
Phone 02 6251 3647
Fax 02 6251 0204
Email eo@statsoc.org.au
Website www.statsoc.org.au

EDITORS

Alice Richardson
School of ISE,
University of Canberra

Michael Adena
Datalytics

CORRESPONDENCE

Please direct all editorial
correspondence to Alice Richardson
Email eo@statsoc.org.au

DISCLAIMER

The views of contributors to this
Newsletter should not be attributed to
the Statistical Society of Australia, Inc.

SUBSCRIPTIONS

The Newsletter of the Statistical
Society of Australia is supplied free to
all members of the society. Any others
wishing to subscribe to the newsletter
may do so at an annual cost of A\$30.00
for an issue of four numbers.

ADVERTISING

Advertising will be carried in the
Newsletter on any matters which
the Editors feel are of interest to the
members of the Society. For details of
advertising rates, etc.

Contact the SSAI Executive Officer at
eo@statsoc.org.au

**DEADLINE FOR NEXT ISSUE:
10 November 2011**

Editorial



Alice Richardson.

This issue of the newsletter reports on many branch events held over the winter months of May to July. Whilst for some branches the change of seasons does not bring a huge drop in temperature, in the southern states it certainly does and some branches choose not to hold mid-winter meetings. The editors hope that you don't let the break in routine break your patterns of attendance at meetings. Maybe a "winter warmer" meeting with soup and some "cool" data analysis will keep networks alive when the days are short! Please feel free to communicate your ideas for keeping branches active over winter to your Branch Council, or through the newsletter.

The International Engagement Section has been active over the last few months, and a report appears in the issue. On the broadest international note of all, Stefan Schweinfest, Chief of the Statistical Service Branch of the UN Statistics Division, has contacted the SSAI with the following information.

"The United Nations has recently announced the 2011 National Competitive Examinations in Statistics. As you may know, it is through this examination process that the UN System, including our office, recruits young professional statisticians. I am pleased to let you know that qualified candidates from your country, 32 years or younger, are eligible to participate in the examinations. The deadline for applications is 10 September 2011.

Detailed information about the examinations and the application and employment conditions can be found at <http://careers.un.org/lbw/home.aspx?viewtype=NCE>."

Do any of you have experience sitting these exams, or working for the UN or one of its agencies? The Editors are always keen to print stories reflecting the diverse experiences of our members.

Alice Richardson

Alice Richardson
Editor

Michael Adena

Michael Adena
Editor

EVENTS

Optimal Design of Experiments—Theory and Application

International Conference in Honor of the late Jagdish Srivastava
25–29 September 2011, Vienna

Biometrics Society Australasian Region meeting: Biometrics by the Kiama Blowhole

4–8 December 2011

"The Sebel Harbourside" in Kiama, NSW

Pre-conference workshops Dec 2, 2011(Sydney),
December 3, 2011 (Wollongong)

<http://www.biometrics.org.au/conferences/kiama2011/index.html>

Statistics for Strategies in Development

11th Islamic Countries Conference on Statistical Sciences (ICCS-11)
18–21 December 2011, Amman, Jordan

8th World Congress in Probability and Statistics (jointly organised by the Bernoulli Society and IMS)

9–14 July 2012, Istanbul, Turkey

Australian Statistical Conference 2012

"Celebrating 50 years of the society"

9–12 July 2012, Adelaide, SA

<http://www.sapmea.asn.au/conventions/asc2012>

From the President

Dear Members

It is a pleasure to contribute to our SSAI Newsletter in my capacity as the new Society President.

First, I wish to acknowledge the contribution of the many members who have served in leadership roles in the Society in the past year. This includes service on Branch and National Executive Committees, Chairs of Sections and Young Statisticians Network, members of the Accreditation and Awards Committees, and the Editors and Associate Editors of our Journal. Thank you for your commitment of time and energy to these roles. Similarly, I wish to acknowledge those who have been willing to take up these roles in this new term; I look forward to working with you all.

Second, I wish to outline the five main issues that I would like to address this year:

1. Promote our Profession.

We have one of the best professions in the world! It encompasses such a broad range of theory, methods, computation and applications. Ours is an area of national and international priority, and as a Society we can promote our many success stories at all levels of academic, industry, government and community. We also have challenges at all of these levels; again, as a Society, we are better placed to address these together.

2. Review and revise the SSAI Strategic Plan and Financial Plan.

The Strategic Plan provides a blueprint for our Society. I encourage you to visit our website and read the Plan, and I welcome your feedback. Our Society is now in relatively good financial shape thanks to the hard work and generosity of many members and associates, so it is also timely to review our Financial Plan.

3. Support and promote the Society's groups.

The Branches, Sections, Young Statisticians Network and committees are the lifeblood of our Society. In addition, the Australian and New Zealand Journal of Statistics is a great asset to the Society and our

profession in general. I will endeavour to work with these groups to support their plans and activities throughout the year.

4. Expand and consolidate linkages between SSAI and other professional societies within Australia and internationally.

There are many Societies within Australia that encompass statisticians and allied professionals as teachers, researchers, practitioners and managers. It is my belief that SSAI can benefit from greater understanding of, and closer interaction with these groups. Similarly, we can benefit strongly from closer ties with our sister Societies such as RSS and ASA, and our geographic neighbours.

5. Review and revise what we deliver as a Society to our members and how members engage with the Society.

I welcome the participation of all members in the Society. As with all such groups, SSAI is not made up of 'them', but of 'us'. It does not exist for 'them'; it exists for 'us'. It will not grow and deliver value to its members and the wider community if we leave it to 'them'; this is up to 'us'. If you would like to discuss any issue pertaining to the Society, offer your services for any role, or suggest activities in which the Society could engage, please feel free to email or phone me (k.mengersen@qut.edu.au; 0408627312) or contact our Executive Officer, Jenny Rintala (eo@statsoc.org). As a corollary to this, we will soon be canvassing members about your areas of expertise and how you might like to be involved in the Society. As part of this, I encourage all of you to spread the word about the Society. Encourage your colleagues to join; inform your employers about the Society and your involvement in it; talk to the wider professional and lay community about us. In this way we can contribute to a strong, large Society of which we are all proud.

Kerrie Mengersen ■

SSAI EXECUTIVE COMMITTEE

Central Council:

President: Kerrie Mengersen
Secretary: Dr Doug Shaw
secretary@statsoc.org.au

BRANCH PRESIDENTS AND BRANCH SECRETARIES

Canberra

President: Bill Gross
Secretary: Dr Ray Lindsay
secretary.actbranch@statsoc.org.au

New South Wales

President: Richard Gerlach
Secretary: Arthur Hung
Arthur.HUNG@cancerinstitute.org.au

Queensland

President: Adrian Barnett
Secretary: Helen Thompson
helen.thompson@qut.edu.au

South Australia

President: Alan Branford
Secretary: Paul Sutcliffe
sutters@bigpond.net.au

Victoria

President: Michael Phillips
Secretary: Sandy Clarke
sfinch@unimelb.edu.au

Western Australia

President: Berwyn Turlach
Secretary: Ryan Admiraal
R.Admiraal@murdoch.edu.au

SECTION CHAIRS

Bayesian Statistics

Scott Sisson
Scott.sisson@unsw.edu.au

Environmental Statistics

Bronwyn Harch
Bronwyn.Harch@csiro.au

Social Sciences

Michele Haynes
m.haynes@uq.edu.au

Statistical Education (co-chairs)

Michael Martin
Michael.martin@anu.edu.au

Peter Howley
Peter.howley@newcastle.edu.au

Surveys and Management

Veronica Rodriguez
veronica.rodriguez@abare.gov.au

Biostatistics (co-chairs)

Mark Griffin
m.griffin@adasis-oz.com
Ian Marschner
ian.marschner@defs.mq.edu.au

Section for International Engagement

Mark Griffin
m.griffin@adasis-oz.com

Young Statisticians' Network

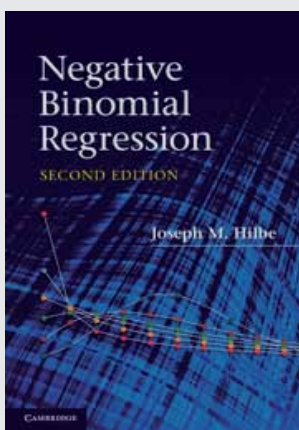
Susanna Cramb
susannacramb@cancerqld.org.au

Further contact details for Society Secretaries and Section Chairs can be obtained by contacting the Society on (02) 6251 3647

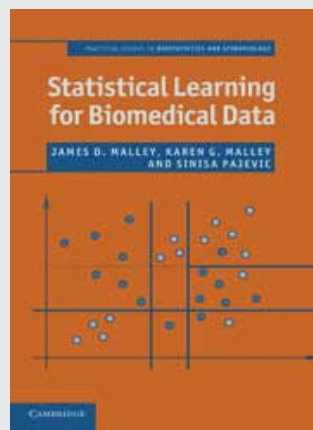
Statistical Society Special Offer

Did you know that Statistical Society of Australia members receive 20% OFF Cambridge statistics books? Take a look at our newest titles!

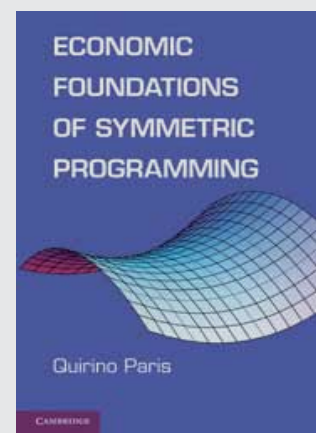
Use code **STATS11** when shopping online at www.cambridge.edu.au



NOW \$78.36 RRP: \$97.95



NOW \$45.56 RRP: \$56.95



NOW \$71.96 RRP: \$89.95

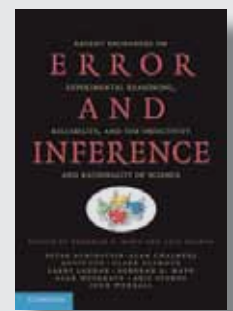
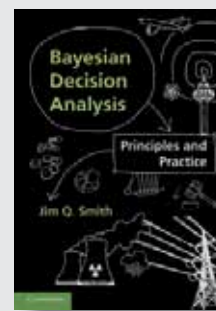
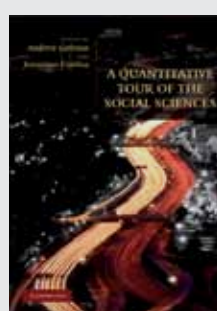
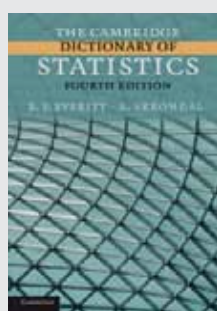
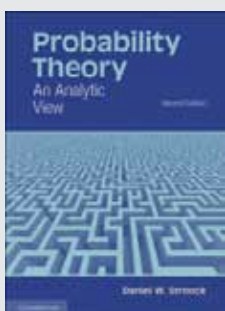
How to Order;

1. ORDER SECURELY ONLINE by going to www.cambridge.edu.au/STATS11

Enter the promotion code **STATS11** when prompted at the checkout, and discounted prices will load automatically

OR 2. Contact us either by email or phone: Rose Johnstone (03) 8671 1405 rjohnstone@cambridge.org

Go to www.cambridge.edu.au/stats11 for a full list of discounted titles



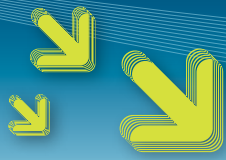
Post Graduate Internships

Take flight

Explore the stimulating and rewarding paid career enriching industry projects on offer at www.amsi.org.au/internships.php



An Innovative University and Industry Collaboration



Zhangxin Frank Liu

Young Statisticians' Conference 2011 cont.



Young Statisticians at the Conference Dinner — a show of treats.

The contributed talks ranged from theoretical to applied statistics. It is worth mentioning that we even had undergraduate students presenting their research work. The careers session discussed different fields that statistics can lead one into.

Congratulations to Julia Polak, Kevin Murray and Davis McCarthy for receiving the Young Statisticians' Presentation Awards. They not only demonstrated their interesting research work, but also inspired us with their excellent communication skills. I would also like to congratulate Susanna Cramb for winning the Best Poster Award.

The conference dinner is an essential component of any young statisticians' conference. In addition to delicious food and wine, there was also a trivia competition. After some intense rivalry between teams, all participants enjoyed

their chocolate prizes.

Overall, based on the feedback we received, it was indeed a great conference. Being the Chair of both Young Statisticians Network and the YSC Organising Committee, I would especially like to acknowledge the help from our members: Dr Mark Griffin, Susanna Cramb, Garth Tarr, Andrey Kostenko, Hong Soo Wong and Steve Quinn for their enormous help in putting this conference together. Kevin Wang, my predecessor who helped organise YSC 2007



LOOKING FOR A JOB?

For a listing of current statistical vacancies in Australia and New Zealand visit:

<http://www.statsci.org/jobs>

Do you have a job to advertise on the website?

Email a position description to jobstatscnc.org.au | Listing is free!

and 2009, provided invaluable organisational guidance. I would also like to thank our society: SSAI, as well as our sponsors: ABS, CSIRO, KPMG and University of Queensland for their phenomenal support in making this conference possible.

I strongly encourage those who have enjoyed YSC2011 to start thinking about taking a more active role within SSAI. I hope YSC2013 will attract more young statisticians from Australia, New Zealand and even from other countries.

Zhangxin Frank Liu

Chair, Young Statisticians' Network and Chair, YSC2011 Organising Committee ■

25% Discount on Wiley Publications!

The SSAI is delighted to announce that we can now offer our members a special discount of 25% on online purchases with Wiley or Wiley-Blackwell (<http://www.wiley.com/statistics>). To receive this discount, please order your books by telephone and quote the code SSA25.

Toll free phone 1800 777 474
(from within Australia only)

Toll free phone 0800 448 200
(from New Zealand only)

Other overseas phone + 61 7 3354 8455

E-mail: custservice@johnwiley.com.au

 WILEY-BLACKWELL

Young Statistician's Conference 2011: Learners Today, Leaders Tomorrow — A reflection

Stephen Horn



This biennial event, organised by and for early career statisticians (at various points in a research higher degree or a few years into employment is typical) under the auspices of the Statistical Society of Australia, took place in Brisbane over two (full) days in July. Of the 55 participants, 31 contributed presentations. Major universities in all mainland states were represented. Key employers also gave plenary presentations. Contributed talks were held to 15 minutes, and the quality of papers and standard of presentation was uniformly high within that constraint. All papers generated discussion, which is rare in professional conferences.

The topic breakdown inevitably reflected the orientation of teaching faculties or employer interest, but juxtaposition in this congenial forum more clearly than usual showed interlocking of theory and application: biological, ecological, medical, geological, infrastructure planning, finance, official. Statistics can be seen as pushing the boundaries of quantitative analysis; not left behind in service courses or reinvented in formulas and automated procedures. For once we seemed to be engaging with the idea of a discipline that is built around a way of thinking about problematic phenomena whose prime characteristic is indeterminate plurality. Well not so far perhaps, but for

this observer it did seem to be heading in that direction; and conversations around and about the formal side did not dispel that impression.

How good is YSC a predictor of the state of statistics in the next ten to 30 years? Our mercurial master of ceremonies showed that the horizons of professional life need not be restricted to the workplace or these shores: '5 days in Suva' should send at least some of this audience home with extra self belief; and likewise stir the Society into standing up for the discipline even as it battles for air in a contested market space.

Stephen Horn ■

Thinking Statistically

Elephants Go to School

A UNIQUE TEXTBOOK

By
Sariinder Singh

Reviews:
Collins Carbone, *Technometrics*,
2007, 49(4), 496
Marcin Kozak, *Statistics in
Transition*, 2006, 7(6), 1407-9.

**A new way to learn
statistics using pictures,
jokes, and tales.**

**A lot of learning with
fun through 651 pages.**

Forewords by
David Robinson
and
Stephen Horn

Good for all ages
+ Good for all libraries
+ Good for all majors
+ Good for all schools
+ Good for you too

PLEASE HAVE A LOOK



**Kendall/Hunt Publishing Company 4050
Westmark Drive, P.O. Box 1840 Dubuque
Iowa 52004-1840, U.S.A.
www.kendallhunt.com**

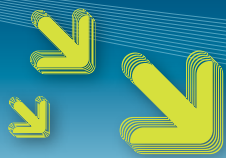
MASA

Model Assisted Statistics and Applications
An International Journal

Editor-in-Chief: Sarjinder Singh
Co-Editor-in-Chief: Stan Lipovetsky
Managing Editor: Stephen Horn
Treasure: Sylvia R. Valdes

Welcomes to:
Sampling, Econometrics, Bayesian
Statistics, Time Series, Design of
Experiments, Multivariate Analysis

www.iospress.nl



ASC2010
Brenton Clarke

Thanks from the Program Chair

I noted with interest the contribution of thanks in the “President’s Column” in the June Newsletter and for his thanks in particular to both Jane Speijers as Chair of the Organizing Committee and myself as Chair of the Program Committee for ASC2010. I was also very pleased to see that all the members of the local organizing committee were duly mentioned for their contribution, which was considerable.

There were of course many others involved one way or another in the running of the conference including, all those who contributed talks and or posters....But I would like to take this opportunity to sing the praises of some who contributed from both near and far to the conference program. We began some time in 2007, three years out from the conference, with our first organizing committee meeting held at Data Analysis Australia. At that meeting we voted Jane to be chair of the organizing committee as she had previous experience in organizing a GENSTAT conference and there was a potential of a GENSTAT meeting alongside ASC2010 (this fell by the wayside but we were nevertheless thankful that Jane took the job to organize ASC2010!). In the initial meeting, as the then current president of the WA Branch and as an academic with some world experience, I put myself forward to seek out a Scientific Program Committee, to report through me to the Organizing Committee. The Organizing Committee agreed! I must say this was my first effort in such matters and I never quite realized how much work that was to entail. Rather than list all who participated in the process I shall recall only those who were continuing members of the program committee. They are Ray Chambers (University of Wollongong), Ross Darnell (CSIRO), John Henstridge

(DAA), Charles Pearce (Adelaide University), Katia Stefanova (Department of Agriculture and Food (WA), Paul Sutcliffe (Australian Bureau of Statistics), Alan Welsh (Australian National University), and Frank Yu (Australian Bureau of Statistics).

John Henstridge acted as a Deputy Chair of both the Organizing Committee and the Scientific Program Committee. He is to be particularly commended for his many contributions to both committees in addition to allowing us to use his rooms for meetings and telephone conferencing. He was ever insightful on how we should approach things from the point of view engaging appropriate speakers who could attract a wide audience across the spectrum of the society of statisticians both from inside the SSAI and outside the SSAI, across the country, and outside the country. I particularly valued his listening ear and responsive insights as a sounding board on how things should proceed.

In addition we had contributions from Jane Waslin and Marie-Louise Rankin, and also Neville Bartlett. Ray Chambers and Frank Yu are also to be commended in persisting with the selection of the Foreman lecturer as part of the ABS Sponsorship. We were very fortunate to have through the auspices of the CSIRO visitors such as Jerry Friedman, arranged with Murray Cameron, and David Balding, arranged with Michael Buckley and Ian Saunders.

From the time of the initial program committee meetings in which we talked about the boom in Western Australia and Charles Pearce uttered the wise words of, “A boom is also followed by a bust!”, the ASC committees faced some serious problems including the Global Financial Crisis and financial concerns following

ASC2008. However, the ASC2010 was in the end a successful conference as measured by both statistics and fiscal measures. The Program Committee had done its work by 2010 and ceased to meet on a formal basis in the run up to the conference. The remaining work of fitting more than 200 presentations into the program appropriately, and how to include social events was managed by Jane and myself with the cheerful help from Anna Munday.

I am particularly thankful to Jane. We eventually had an almost daily phone call arrangement, either from home or from work. She was also an inspiration for getting the associated workshops organized, including saving \$10,000 vis a vis providing laptops to the conference from the various colleagues on the committee, and polishing advertising for the workshops and Persi Diaconis’s public lecture. Her major contribution to financial detail was always welcome. Also we could not have done without PROMACO, but hey that is their business.

Brenton Clarke ■

International Engagement

Section news
Mark Griffin



INTERNATIONAL CONFERENCE FOR HEALTH STATISTICS IN FIJI (JULY 2011)

The First International Conference for Health Statistics in the Pacific Islands was held in Fiji from the 5th to the 8th of July 2011. According to the Dean of the Fiji School of Medicine, statistics is the technical skill needing the second-most attention to development and education within countries such as the Pacific Islands (where the importance of statistics comes second-place only to building technical expertise in human resources). The ICHSPI conference series was established to meet these international needs, while also enabling international researchers the opportunity to enjoy the beauty and culture of the Pacific Islands. The 2011 conference hosted 50 delegates from the Pacific Islands (Fiji, Solomon Islands, Guam, New Caledonia, Marshall Islands, the Northern Mariana Islands, and Hawaii) and 15 delegates from Australia, New Zealand, and the USA. Plans for the 2012 conference (renamed as the Joint Oceanic Conference for Statistics and Information Systems, JOCSIS) are well underway and will be held in Madang, Papua New Guinea in July 2012. This conference series is jointly organised by members of the Statistical Society of Australia (Section for International Engagement), the Australian Development Agency for Statistics and Information Systems, the New Zealand Statistical Association, and the American Statistical Association. For further details about this conference program please email the Conference Chair Mark Griffin at m.griffin@adasis-oz.com.

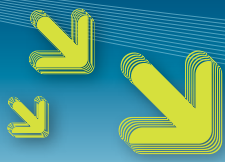
AUSTRALIAN STATISTICIANS AT THE ASA JOINT STATISTICAL MEETINGS 2011

In October 2010 the American Statistical Association (ASA) approved the formation of the Friends of Australasia (a special-interest group within the ASA), and in March 2011 the Statistical Society of Australia approved the formation of the SSAI Section for International Engagement. These two parallel groups promote the well-deserved place of statisticians from Australia and New Zealand in the international statistical community.

At the ASA Joint Statistical Meetings in Miami in August 2011, the Friends of Australasia was formally unveiled to the delegates at the conference. The Friends organised an evening mixer/open business meeting at the conference and set up an information table near the main registration booths. Through these activities the Friends were able to promote the upcoming SSAI conference to be held in Adelaide in July 2012, the upcoming conferences of the New Zealand Statistical Association, and the future conferences in the Pacific Islands. Through the Friends, Louise Ryan (Head of CSIRO Mathematics, Informatics, and Statistics) also led a round-table discussion entitled "Do you want to visit Australasia?" where she further promoted job opportunities and conference in the region, and Justin Fisher (a key speaker at the Pacific Island conferences) led a round-table discussion entitled "Preparing to work in a developing nation".

The ASA Friends of Australasia and the SSAI Section for International Engagement will continue to work towards bridging statisticians across these geographic boundaries. For more information about these activities, please email Mark Griffin (Chair of these two groups) at m.griffin@adasis-oz.com.

Mark Griffin ■



Caro Badcock

Adaptive Designs For Clinical Trials Workshop | 6-7 April 2011

The adaptive design workshop proudly supported by the Australian Pharmaceutical Biostatistics Group (APBG), the Statistical Society of Australia (SSAI) and the George Institute that was held in Sydney in April was a great success. To our knowledge, this was the first workshop in adaptive designs to be held in Australia. We were fortunate enough to have 3 leading international researchers conduct the training, Dr Brenda Gaydos (Eli Lilly, U.S.A), Dr Frank Bretz (Novartis, Switzerland, adjunct professor of the University of Hannover Germany) and Dr Patrick Kelly (Sydney University). There were 35 attendees including people from industry, academia, contract research organisations and research institutes.

Day 1 was focused on a high level overview on adaptive designs including regulatory aspects as well as operational and logistical issues. Day 2 was aimed at the more technically minded. It included the topics; response adaptive dose-finding studies, a Bayesian approach, group sequential designs and adaptive designs for confirmatory trials.

The wealth of experience that was shared over those 2 days was encouraging and we look forward to seeing how the applications of adaptive design are implemented in the industry. Hopefully this will help get more accurate information for patients' prescribers and payers in a timely manner. Stay tuned for more workshops in this area.

Caro Badcock ■



Brenda holding the group enthralled with her experiences.



Brenda taking us through the Bayesian approach



Frank introducing the session on adaptive designs for confirmatory studies



Frank providing a considered response to one of the many questions fired at him and Brenda during the two days



Patrick deep in explaining alpha spending functions.

Australian Statistical Conference 2012

Gary Glonek



Adelaide Convention Centre

The Australian Statistical Conference in 2012 will be held in Adelaide, South Australia, 9-12 July, at the Adelaide Convention Centre. In 2012, the Statistical Society of Australia will celebrate the 50th anniversary of the creation of the national society. The theme for the Conference is "Celebrating 50 years of the society". It provides opportunities for presentations on a wide range of topics and recognises the role that statistics plays in all aspects of modern life. In the lead up to the conference, we will be providing details about the scientific program, social program and the conference location in the SSAI newsletter. In this issue, we are

pleased to announce the diverse range of internationally renowned keynote speakers who have been confirmed for the conference.

Professor Anthony Davison, Ecole Polytechnique Federale de Lausanne

Professor Christl Donnelly, Imperial College London

Professor Peter Donnelly, Oxford University

Professor Christian Robert, Université Paris-Dauphine

Professor Sophia Rabe-Hesketh, University of California Berkeley

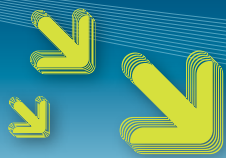
Professor John Storey, Princeton University

Stay tuned for full details of our complete invited program and pre and post conference workshops that will be announced over the coming weeks. For further information or to register your interest, please visit the conference website:

<http://www.sapmea.asn.au/conventions/asc2012>

Gary Glonek

Scientific Program Committee Chair



SMP
Adrian Barnett

Science Meets Parliament

My journey to Science Meets Parliament started with a major wardrobe malfunction. The culprit was my one suit, which I wear once in a blue moon. After its last outing I had diligently taken it to the dry cleaners, who had less than diligently given me the wrong trousers.

Starting with an attention grab is one of the strategies I learned during the two days in Canberra at Science Meets Parliament. The first day was actually Science Meets Journalism, as the two-hundred plus scientists in attendance were gently but repeatedly scolded for not using the political language when talking about their work. Politicians, journalists and the general public have a frightening short attention span. If you don't get their attention quickly, then you've lost them. Start with arresting facts and figures, and finish with how your amazing work will secure jobs, money and votes. Never whinge at a politician, and never ever ask them for money.

We also learned about the vomit principle. That's where you repeat your idea to as many

politicians and influential figures as possible until the point where you're ready to vomit. This is the moment when your idea might finally be taken seriously.

A common theme was that scientists should be more willing to engage with the public and media on science issues. This included speaking to the media if the government did something you did or didn't agree with.

The most common topic was climate change, and how this primarily scientific issue had become a primarily political one. It was also the main issue when we sat in the public gallery during question time, and when we heard Professor Ian Chubb (Chief Scientist of Australia) speak at the National Press Club. Chubb was asked by one journalist if he was sick of being asked about climate change, another example of the vomit principle.

I tried the vomit principle on Kim Carr (Minister for Innovation, Industry, Science and Research) by asking him about his willingness to reduce the complexity of ARC and NHMRC grant applications which are often over 100 pages long with only 10 pages of science. He started with a stern answer to a question I hadn't asked, but then interestingly admitted that he had to sign off on, "decisions that he could publically defend." If next year's ARC applications are half the size then you know who to thank. If they are 10 pages longer then it is time for more of us to speak out.

The organisers of Science Meets Parliament did a great job, especially considering the extra pressures of dealing with the volcanic ash cloud which threw most people's travel plans into chaos. Every scientist got to meet a politician and talk about their work. Together with two other scientists, I met Senator Sue Boyce (Queensland). It was a refreshingly open meeting punctuated by ringing bells which called Senator Boyce to dash off for a division.

Science Meets Parliament provided some valuable lessons, including checking your clothes at the dry cleaners. There is an understandable level of cynicism around current politics, but this should not stop us from speaking our mind on those issues where we are experts and the public and government need guidance.

Adrian Barnett ■

Statistical Estimation of Total Discharge Volumes

Canberra
Ray Lindsay



Dan Pagendam, of CSIRO Mathematics, Informatics and Statistics spoke to the Canberra Branch on 24 May, on “Statistical Estimation of Total Discharge Volumes”, especially using gauging station data, and incorporating uncertainty into the estimates. Total discharge volumes are estimated by integrating the hydrograph, and usually displayed on a log scale. As the data are collected at gauging stations at discrete time points, there is uncertainty about what the reading would have been between observations. Gaps in the series can occur when the station breaks down, or become obstructed.

Stochastic differential equations offer a framework for modelling stream flow as a continuous time stochastic process, in other words an ordinary differential equation plus volatility. A possible model suggests an infinitesimal change in the log of flow as a drift function, plus a volatility function times a Brownian motion factor. Dan relied on a numerical approximation of these equations through a discrete Euler-Maruyama scheme.

Dan explained how one can easily generate trajectories that bridge the discrete time observations, by requiring that there be equivalent backwards and forwards diffusion processes.

Another cause of uncertainty in the flow is that gauging stations measure stream depth, not flow explicitly. Empirical relationships between stream height and stream flow, developed over a long period, are used to calculate the modelled flow, but these curves are inherently uncertain. There is also an error in measuring stream depth, and a small change in this can translate into a large error in stream discharge.

Dan combined these uncertainties (process uncertainty, rating curve uncertainty and measurement uncertainty) by simulating a large number of hydrographs, computing

their integrals and studying the distribution of discharge volume. This method appears to have potential, though the reversibility assumption can sometimes lead to silly predictions when there are long gaps in the series.

In conclusion, Dan has developed a statistical approach to estimating total discharge volumes and gives a range of possible predictions. The method is largely non-parametric and data-driven so could be used by non-modellers, once packaged into an easy to use form. The methods have potential in other datasets of discrete observations of a continuous time process.

After the talk, members of the Branch enjoyed dinner at the Spicy Ginger Café.

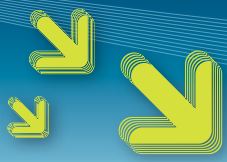
Ray Lindsay ■

20% SSAI Member Discount Promotion!

Cambridge University Press Australia is pleased to offer an exclusive 20% SSAI member discount off selected statistics titles. Please go to <http://www.cambridge.org/aus/catalogue/promotion.asp?nav=view&code=STATS11> to see the available titles.

To apply the discount, simply enter the promotion code STATS11 when prompted at the checkout stage of your order, and the prices will be automatically updated.)

CAMBRIDGE
UNIVERSITY PRESS



AMSI workshop funding rounds *2011 now open*

The Scientific Advisory Committee reviews and approves sponsorship by AMSI of a diverse range of symposia, workshops, theme programs and lecture tours.

Prof. Jon Borwein (The University of Newcastle) is Chair of the Scientific Advisory Committee with a membership of eminent national and international mathematical scientists.

Applications for workshop sponsorship up to \$10,000 are open to academics from AMSI member departments.

**2011
rounds close:
2 September 2011
1 December 2011**

See: www.amsi.org.au/proposals.php



National
Collaboration
in the Mathematical
Sciences

Queensland Branch News

Queensland
Adrian Barnett



MAY

The May seminar for the Queensland branch was presented by Associate Professor Noor Azina Ismail, from the University of Malaya. Noor is a regular visitor to Australia, having received a Masters degree from the University of New South Wales, and a PhD from Queensland University of Technology.

Noor spoke about identifying and predicting students' attitudes toward mathematics using latent class analysis. Latent class analysis (LCA) is a popular tool in identifying latent groups of related cases using a set of categorical or continuous observed variables. Since the data structure in this study includes students nested in schools, a traditional LCA that assumes observations are independent of one another is not suitable to analyse this type of data. Noor illustrated the use of a multilevel latent class analysis (MLCA) as an alternative. Latent classes of students' attitude toward mathematics among 4466 Malaysian students in the 8th grade were considered (this data is available freely on-line). The best model comprised of three student-level latent attitudes. Students with higher achievement and more books in the home had a better attitude towards mathematics. The number of books in the home may be a proxy for socio-economic status.

After the talk there was a lively discussion about the use of school league tables in Australia, and entrenched negative attitudes towards mathematics and statistics in some children and parents.

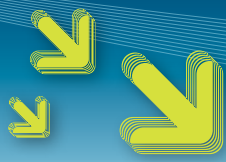
JUNE

Queensland's June meeting welcomed Professor Joe Whittaker from the Department of Mathematics and Statistics, Lancaster University, England. Joe spoke about the analysis of survey data, involving data collected on a set of response variables defined over a finite population. He demonstrated a graphical method that gives a bird's eye view of variable inter-relationships. Each variable is represented as a node and the line thickness represents the strength of the association. The plots are useful for highlighting: i) those variables that strongly modify the conditional distribution of another variable, and ii) those which have little influence. He used an interesting example detailing the association between smoking, income and education. A strong dependence for smoking was clear between individuals and their partners. Another example on job and salary satisfaction demonstrated how it was possible to partition the variables so that associations on flowed in one direction. Joe discussed how the graphs are useful for non-specialists, particularly for those people who typically only use multiple cross-tabulations for investigating associations. The graphs can also be viewed as a useful exploratory analysis. The group had an interesting discussion about incorporating variables with missing data. An R package is in development that will draw these graphical models.

JULY

The Queensland Branch awarded 6 scholarships for members to present at YSC2011 in July. As a stipulation of the award, recipients were asked to provide a brief summary of their presentation and also their conference highlights. The award recipients were: Earl Duncan, Jeff Hsieh, Jegar Pitchforth, Matt Moores, Susanna Cramb and Tuti Thamrin. All applications received were from, and hence awarded to, honours and PhD students from Queensland University of Technology (QUT). Submissions from the students are given below.

Earl Duncan, Honours Student, QUT: My presentation was about common dose-finding methodologies in phase I clinical trials with particular focus on pharmaceutical drugs, such as those used in the treatment of cancer. Some common experimental designs such as the non-parametric 3+3 design and the continual reassessment method were briefly described followed by a brief look at some common design criteria. My presentation concluded with a worked example illustrating the design algorithm process for a D-optimal design for a simple quadratic regression model. The number one highlight of the conference for me was the high quality and variety of presentations. It was really great to hear how statistics was being applied to a host of real-world problems, from sampling marine life on the Great Barrier Reef to the role of statistics in developing nations. It was also a fantastic opportunity to network and meet new friends. The conference exceeded my expectations and I would encourage anyone who has the opportunity to attend a conference in the future to seize the opportunity.



Queensland cont.

Jegar Pitchforth, PhD Student, QUT: My presentation was an overview of my work around modelling the incoming passenger process for Brisbane International Airport. In the presentation we defined the research as discovering an applied method for combining and visualising multiple complex systems models in order to make use of and describe all the data relevant to system behaviour. The final goal of this research is to produce a range of software applications for airport stakeholders. As this was my first YSC, the main highlight of the conference was to meet and be inspired by all the other Young Statisticians in attendance. In particular Julia Polak's practical demonstration of how distributions move when dividing a sample into sub-populations, and Celestine Bond's exploration into political influence on the stock market sparked my imagination and re-inspired me to learn more about the issues of dealing with real data in statistical analysis. While all of the talks were interesting and informative, I also found the opportunity to network with fellow students as well as experienced statisticians invaluable to gaining a better understanding of the practical side of working as a statistician, and a first chance to forge new friendships. A big thank you to all those involved in the organisation!



Matt Moores

Matt Moores, PhD Student, QUT: I presented a poster on my PhD project in medical image segmentation: using prior information in a Bayesian model to infer the location of boundaries in cone-beam computed tomography. YSC provides a great environment for first-time presenters, with short talks (10-12min plus questions) and a receptive and welcoming audience. There was a wide variety of topics, including geostatistics, biostatistics, nonparametric and Bayesian statistics. The social events were also great, especially trivia night (and I'm not just saying that because we won). I made friends with statisticians from all over Australia and New Zealand. I also had the opportunity to discuss my research with high profile invited speakers.

Susanna Cramb, PhD Student, QUT: At YSC2011 I presented a poster examining the influence of advanced stage of breast cancer at diagnosis on geographic survival differentials in Queensland. Since it was examining small areas I used Bayesian hierarchical models to conduct these analyses. This work was done in collaboration between QUT and Cancer Council Queensland, and my co-authors were Kerrie Mengersen and Peter Baade. I was humbled and thrilled to receive the best poster prize, and this was a definite conference highlight! Other highlights included hearing the distinguished keynote presenters and all the other excellent presentations, learning new things, meeting new people and catching up with past conference friends, as well as the great food and freebies! I thoroughly enjoyed YSC2011 and I'm looking forward to YSC2013.



Tuti Thamrin

Tuti Thamrin, PhD Student, QUT: My presentation was about modelling of lifetimes of gene expression survival data via Bayesian analysis to predict patients' survival using the Weibull model and the Weibull mixture model. It also addressed methods for model selection such as the Aitkin Posterior Criterion (APC), Deviance Information Criterion (DIC) and Bayesian Information Criterion (BIC). We could benefit from this conference in many ways such as networking among Young Statistician and other relevant parties, developing statistical knowledge and sharing information on the current statistical issues. This conference also has provided insights for future career development for Young Statisticians.

Adrian Barnett ■

BioInfoSummer 2011

AMSI Summer Symposium in Bioinformatics

Date: 5–9 December

Location: Gene Technology
Access Centre,
Melbourne

BioInfoSummer provides an introduction to the broad research area of Bioinformatics for students and early career researchers from the mathematical and biological sciences.

This year BioInfoSummer will include workshops for school teachers.

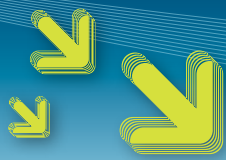
Sponsors

EMBL
Australia



Walter+Eliza Hall
Institute of Medical Research





South Australia
Paul Sutcliffe

Variable Selection in Linear Mixed Models with application to QTL analysis

Dr Julian Taylor, a postdoctoral fellow at the Mathematics, Informatics and Statistics division of CSIRO was the speaker at the May South Australian Branch meeting. Julian gave an entertaining talk on variable selection in linear mixed models for modern plant breeding experiments using the analysis of Quantitative Trait Loci (QTL).

The motivating example was from a multi-phase wheat quality experiment in which there are field experiments, milling experiments and baking experiments. Each phase has different traits. For example, the field phase provides yield and seed size. The milling yield and the volume of the loaf are recorded from the other phases. There are non-genetic sources such as milling days, positions and operators and the usual field (row/column) factors. In wheat quality experiments it is also often of interest to include genetic components that may involve many breeding lines with an associated large set of markers spanning their whole genome. In this example given there were 410 markers genotyped from 180 lines of a double haploid population. For these experiments the models become complex as they are required to capture all extraneous non-genetic variation, the genetic variation between replicated individuals as well as estimate possible QTL's of interest across the whole genome. The first two suggest linear mixed models but the high dimensional nature of the third requires more complex techniques such as variable selection to be integrated with mixed models.

To achieve this integration an extended linear mixed model is proposed where the QTL effects are penalized using a re-parameterized, well known class of penalties. The re-parameterization allows one of the penalty parameters to be estimated using current mixed model techniques. The second parameter of the penalty is determined using a modified form of the Bayesian Information Criterion (BIC). To ensure efficiency throughout the model fitting process, the Mixed Model Variable Selection (MMVS) method uses the flexible



Itr	Int	Log		BIC		L_{-1}		L_{-2}		L_{-3}	
		Size	Z-stat	Size	Z-stat	Size	Z-stat	Size	Z-stat	Size	Z-stat
A	18	0.20	2.25								
IA	1	0.37	3.95	0.34	3.33	0.36	3.46	0.38	3.46		
IB	18	-0.35	-3.94	-0.28	-2.98	-0.28	-2.94	-0.34	-3.35		
ID	2	0.55	5.50	0.46	4.25	0.46	4.18	0.44	3.75	0.42	3.37
ID	18	0.30	2.73								
IA	5	0.21	2.34								
IA	12	0.23	2.48								
IB	21	0.19	1.96	0.38	3.74	0.34	3.42	0.42	3.91	0.37	3.31
IB	14	-0.24	-2.71	-0.30	-3.09	-0.32	-3.28				
IA	10	-0.21	-2.22	-0.24	-2.32						
IA	13	-0.27	-2.87	-0.25	-2.52	-0.26	-2.59			-0.39	-3.50
IB	30	0.20	2.21								
IA	17	-0.32	-3.56	-0.28	-2.94	-0.33	-3.44	-0.35	-3.35		
IB	1	-0.21	-2.30								
IB	15	0.34	3.72	0.36	3.66	0.38	3.86	0.37	3.51		
IA	2	-0.27	-2.76								
IA	8	0.27	2.79								
IA	13	-0.20	-1.94	-0.28	-2.56						
ID	5	0.26	2.72	0.28	2.76	0.30	2.90			0.34	3.06



software package ASReml-R (Butler, et al. 2009), as its core linear mixed model fitting routine.

A general simulation study revealed the extended class of penalties achieves varying degrees of estimator shrinkage depending on one of its parameters. The simulations also show the link between the number of false positives and the number of true coefficients using the same penalty. The

MMVS method was applied to the wheat quality data set from the Food Futures Flagship, CSIRO where the focus is the analysis and interpretation of QTL. Overall, the simulations reveal that the BIC is not optimal, but works well in practice. This approach was shown to be effective in whole genome analysis of QTL.

Paul Sutcliffe ■

Census 2011

South Australia
Paul Sutcliffe



The speaker at the April South Australian Branch meeting was Caroline Deans, manager of the South Australian Census Management Unit, at the Australian Bureau of Statistics (ABS). The Australian Population and Housing Census will be conducted on Tuesday 9th August 2011 and was at the time of the talk 111 days away. The Census is often described as Australia's largest peace time operation, costing \$440 million over its five year cycle and employing 43,000 people.

In opening her talk Caroline noted that this year's Census will be the 16th National Census and a hundred years since the first Census. The Census is a massive logistical management. In the field phase there are 2,300 collectors and 300 area supervisors just in South Australia.

Historically the Census enumeration methodology has been a method which requires the collector to deliver to every household a questionnaire and to return to collect completed forms (drop-off and pick-up). However, more recently in 2006

and again in 2011 development of an eCensus enables households to 'return' their questionnaire by completing a web-based version. This has an advantage over the physical collection method since no agreed return time needs to be made, resulting in less intrusion for the householders and reduced cost of collection.

There are procedures in place to avoid over counting (e.g. duplication etc) in the Census. Under counting in the Census is measured in a follow-up survey (post-enumeration survey) which 'checks' the number of people at households at the time of the Census. Discrepancies can be analysed and an assessment of the overall undercount is published after each Census.

A key Council of Australian Government (COAG) requirement is to get the indigenous count right. In response, the ABS has developed a northern Australian collection strategy to co-ordinate the collection of forms from indigenous communities.

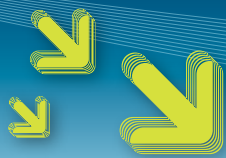
From the number and type of questions after the talk it is obvious that the Census is an interesting topic. For researchers some key points were raised:

- Meshblocks are small geographical areas which are combined to form collector workloads, and although limited data will be released at the meshblock level, customised aggregations won't be available.
- There is a change in standard geography which analysts should be aware. The Census will be output at the SA1 level (approximately equivalent to CD level).

- Questions in the Census generally remain the same, and some, such as the question on religion are voluntary.
- The ABS encourages statistical users to use the table builder functionality for basic tables.
- For more in-depth analysis the 1% sample unit record file when it becomes available can be purchased. For researchers at universities there is an agreement which arranges organisational approval so access may be available through this means.

A very interesting and informative talk was followed by a meal at the Union Hotel.

Paul Sutcliffe ■



South Australia
Richard Woodman

Analysis of Perinatal trials

Dr Lisa Yelland, a senior statistician at the Australian Research Centre for Health of Women and Babies (ARCH), within the University of Adelaide, discussed the appropriateness of different approaches to account for the clustered nature of data in perinatal trials, in which multiple birth infants (e.g. twins) are required to be treated as non-independent observations due to their sharing of both genetic and environmental factors. The motivating example was data from the DINO trial, in which infants were assessed for the odds of mild delay in mental development at 18 months after receiving either high or standard doses of fish oil, in their infant-milk formulas. Recommended approaches to analysis for similar trials have varied from ignoring of the clustering completely, particularly when there are a relatively small proportion of multiple births, to always accounting for the clustering, regardless of the multiple birth rate.

In her presentation, Dr Yelland outlined the assumptions and model characteristics of the two most commonly used approaches used to account for clustering – generalized estimating equations (GEE's) and mixed-effects models (MEM's). Importantly, the audience was also reminded of the different parameter interpretations for

GEE's (population averaged effects) and MEM's (average cluster-specific effects). In this particular setting, the estimated beta coefficients for the GEE are the log odds ratio that compares the average probability of a mild delay in mental development in birth clusters receiving different treatments (averaged across all infants or birth clusters), and the coefficients for the MEM represent the average within infant (or birth cluster) log odds ratio of a mild delay in mental development when each infant (or birth cluster) receives one treatment versus the other. As such, this highlighted the point that the strict interpretation of the effects for MEM's is not possible in this scenario (since all infants within each birth cluster received the same treatment).

A simulation study was then described in which the author had examined the combined effects of the extent of clustering (i.e. the size of the intra-class correlation coefficient (ICC)) and multiple birth rate, on the Type-1 error rate and the coverage rate for 95% confidence intervals. For low levels of clustering (ICC=0.1), both GEE's and MEM's performed well in terms of both the Type-1 error rate and coverage rates, regardless of the multiple birth rate. However, at high levels of clustering (ICC=0.9), when using MEM's there was a steady and pronounced increase in both Type-1 error rates and

a similar reduction in 95% CI coverage rates, even at relatively moderate multiple birth rates (around 5% of subjects). GEE's, however still performed well in terms of Type-1 error rates at high ICC's, although there was a slight drop in the 95% CI coverage rate at very high multiple birth rates.

In her conclusion, Dr Yelland advised that in contrast with previous recommendations, clustering should not be ignored in perinatal trials, particularly if the ICC is high. In such trials, GEE's appear to perform fairly well regardless of the degree of clustering or multiple birth rate, whereas MEM's vary in performance depending on both the ICC and multiple birth rate and should therefore be used with caution.

Richard Woodman ■

Comparative Risk Assessment: what is it and how can it be useful

Victoria 
Steven vanderHoorn



Steven vanderHoorn

Our April seminar was deferred by a few days to 3 May, and was presented by Steve Vander Hoorn from the Statistical Consulting Centre at the University of Melbourne. Steve became involved with the Comparative Risk

Assessment (CRA) project in 2000 when he joined a team of researchers led by the World Health Organisation and Harvard School of Public Health. Initially, he was one of the statisticians working within the core development group of the CRA project and later contributed as an advisor on statistical questions arising across different areas of the Global Burden of Disease (GBD) project..

Steve began his talk by providing an introduction to the CRA framework in the GBD project. A Comparative Risk Assessment is the procedures and methods used to estimate disease and injury burden

arising from population exposure to a given hazard using comparable definitions, frameworks and outcome measures, and is used to examine such issues as the impact of high blood pressure on heart disease, which involve a population health based examination of risk.

CRAs allow for the size of the (current and future) burden of disease due to exposure to be quantified – this is of value for policy purposes, and allows for this burden to be compared with other diseases/injuries. It also allows the benefits of interventions to be quantified.

As part of the risk assessment, a method is required for determining the number of deaths/ cases (from each disease and injury) that would be avoided if exposure to a risk factor, or group of risk factors, were at a different level. This can sometimes be done by direct epidemiological studies, but what can be done for populations other than the epidemiological cohort?

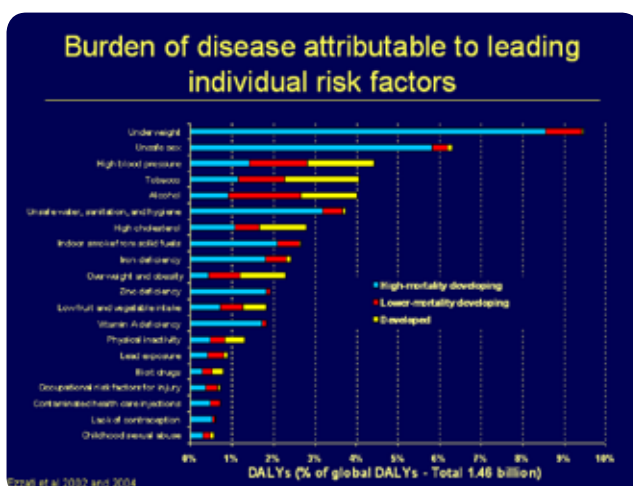
Steve briefly referred to a number of risk assessments with a population perspective that had used novel approaches before

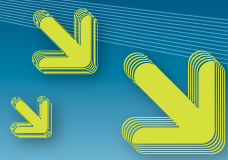
outlining some key features of risk factors used in a risk assessment exercise. These should ideally be potentially modifiable, assessed irrespective of their place in a causal chain, defined and measured with comparable analytics and methods, and cover protective as well as hazardous factors.

Steve then presented some information on individual risk factors that contributed most to the burden of disease. The highest risk factor was childhood and maternal underweight, which had a considerably higher risk than unsafe sex, which rate second. Both of these were considerably higher than the next set of factors, which had similar risks: high blood pressure, smoking tobacco, alcohol use, and unsafe water, sanitation and hygiene. Working groups, which include statisticians, epidemiologists, and occupational experts, have been established for each of the risk factors to locate all the published information about their risk factor.

An alternative question for the risk assessment is to work out how to determine

Several example illustrations presented in Steve’s talk:





Victoria cont.
Steven vanderHoorn

the fraction of deaths/ cases (from each disease and injury) that would be avoided if exposure to a risk factor, or group of risk factors, were at a different level. This involves looking at the determinants of proportional effects such as exposure distribution, relative risk, and baseline or counterfactual exposure, and the determinants of absolute effects, such as the total disease-specific mortality/ burden.

Biomarkers (such as blood pressure, blood glucose) can be used as exposure variables, as these can be standardised, are single variables, and can be defined as continuous variables. But many of these are costly, difficult or intrusive and hence there is limited data available.

Other options for exposure variables include those that measure behaviour, technology or the environment, and many can be obtained via household surveys and also measured longitudinally. However, these tend to be harder to standardise across surveys and many exposures are "multidimensional" and harder to measure as continuous variables.

Steve then worked through an example of measuring exposure to alcohol use, noting that there is not just one measure of alcohol consumption. Patterns of drinking could be defined as: the proportion of adults who abstain from alcohol; the proportion who have heavy drinking occasions; the proportion who drinking daily or nearly daily;

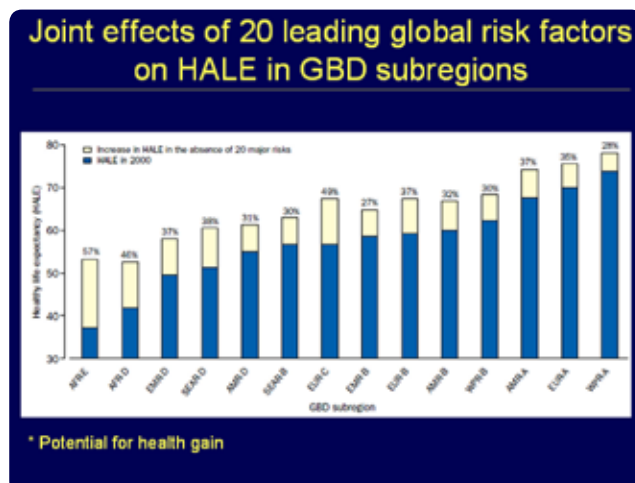
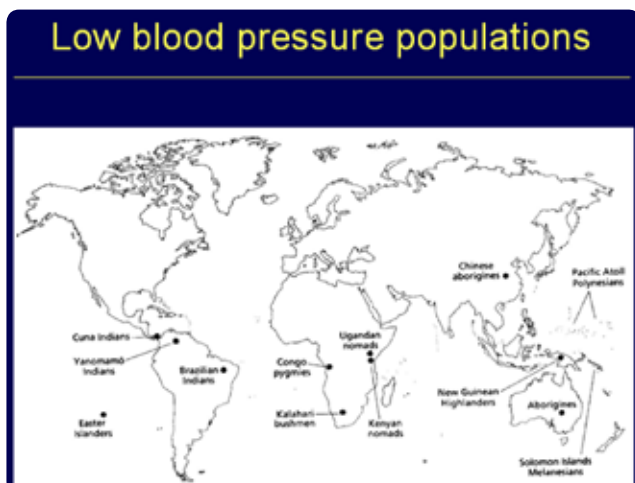
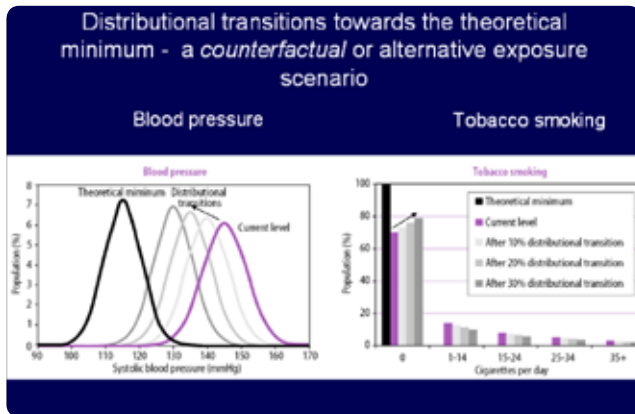
drinking that is linked to violence; and so on. There is a need to come up with better ways of describing problem drinking that is linked to injury and disease.

Choosing an appropriate effect size also has its challenges. The CRA framework requires knowledge of the relative risk by age and sex for the population, and the need to determine how much the proportion risk varies across populations, even individuals. Potential sources of bias also need to be avoided or at least assessed.

The theoretical minimum risk exposure distribution also needs to be determined – for some risk factors "zero" is meaningful (e.g. smoking), but for some others (eg BMI, blood pressure) zero may be undefined or physically unattainable, or zero may not be the lowest risk (e.g. alcohol consumption). And what are the optimal exposures for protective risks such as physical activity, fruit, vegetable and fish intake?

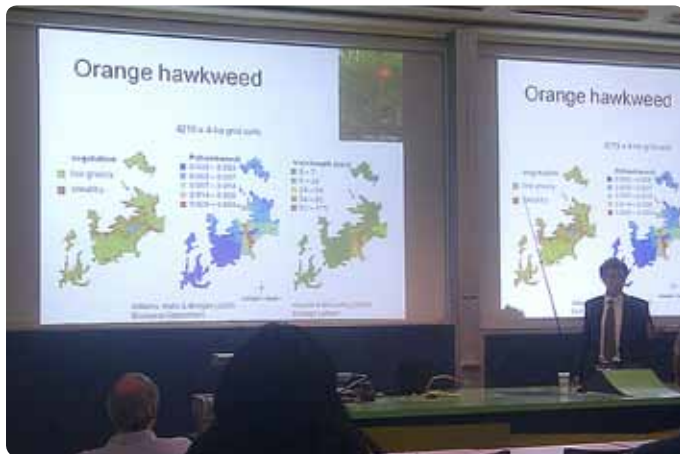
Steve then finished his presentation by providing some examples of continuous exposure distributions and attributable fractions versus attributable deaths

Steven vanderHoorn ■



Applications and misapplications of statistics in ecology

Victoria cont.
Steven vanderHoorn



Professor Michael McCarthy delivering his presentation



The May seminar was presented by Associate Professor Michael McCarthy, from the School of Botany at the University of Melbourne and also deputy director for the ARC Centre of Excellence for Environmental Decisions. His research focuses on developing minimally simple models for ecology, with a particular interest in developing approaches that can assist rationale environmental management.

In his presentation on the applications and misapplications of statistics in ecology, Michael noted that ecology and related disciplines have a strong statistical tradition but the quality of modern applications of statistics in ecology is mixed.

Michael used the “toe clipping of frogs” in one of his examples on the misapplication of null hypothesis significance testing. Toe clipping, as a method of marking and identifying frogs in the wild, has long been used as a fast and cheap method. For a frog with 4 toes on each of the front feet, and 5 toes on each of the back feet, there are about 1000 combinations where 3 toes are removed. (I think some of us were feeling a little sorry for the frogs!)

It did not seem surprising that there was a correlation between the proportion of marked frogs who were found again, and the number

of toes that had been removed (up to 9 toes!). However, we were reassured that this may not be because the frogs had died due to the physical disadvantage – it may have been that the more toes they had removed, the more they had learnt to be “trap shy”.

Michael reported that he had examined many papers that investigated the return rate for frogs and other species and found that the importance of statistical power was often ignored or misunderstood by ecologists. For example, one paper reported that the 27 of the 808 marked frogs that were recaptured were not significantly different from unmarked animals in length or weight, and hence concluded that there was no negative effect of toe clipping! In general, the experiments tended to have very low return rates, making it impossible to test the null hypothesis that the removal of toes has no impact on survival.

The next example examined the use of hierarchical models to analyse bird dispersal and the response of birds to vegetation clearance. Hierarchical models are an example of a good application because they deal appropriately with the different sources of variability. Hierarchical modelling has become more popular with ecologists, with at least two books on this issue written

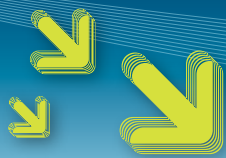
especially for this profession.

The distance birds travel from their nest to where they first breed is difficult to estimate but useful to know, in terms of gene flow and response to fragmentation of vegetation. It is of course easier to find birds that nest close to their birth nest, than those who have travelled further, and sample sizes are usually small as well.

Statistical models have been built that take into consideration body mass, sex, feeding guild (method) and wing morphology (shape) to predict the responses to fragmentation of vegetation. The researchers looked at 24 different landscapes in northern Victoria and how the amount of tree cover related to the appearance of the birds. The results were as expected, showing that the birds that travelled greater distances were less affected by tree cover.

Michael concluded his presentation with thanks to the various researchers who had worked on each of the studies, and noted that a balance is needed between using funds on doing research as opposed to using funds to more directly manage the ecological problems that are being faced.

Steven vanderHoorn ■



Victoria cont.
Carol Soloff

Statistical Methods for Differential Gene Expression Analysis Using High-Throughput Sequencing and The Central Limit Theorem

The seminar this month featured presentations from two of our young statisticians, Davis McCarthy, and Han Gan.

Davis, from the Walter and Eliza Hall Institute of Medical Research and the Department of Medical Biology at the University of Melbourne, spoke first on *Statistical Methods for Differential Gene Expression Analysis Using High-Throughput Sequencing*.

Cheap, high-throughput DNA sequencing has transformed contemporary biology. RNA sequencing (RNA-Seq) experiments represent an exciting approach to investigate gene expression, in particular for finding genes that have different levels of expression under different experimental conditions (differential expression). However, data analysis currently presents a bottleneck in RNA-Seq studies and the high-dimensional count data (thousands of genes, a handful of samples) generated raises many statistical challenges.

Davis described a negative binomial (NB) model for differential expression in RNA-Seq data. The NB model can account for biological variability in the data, of great importance for the reliable assessment of differential expression.

Lessons learned from the analysis of microarrays guide their approach.

In particular, the idea of sharing information between genes to improve inference is applied to RNA-Seq data. Weighted and approximate conditional likelihood methods allow information sharing similar to an empirical Bayes approach. Whereas standard statistical methods would fail with such small sample sizes, empirical Bayesian approaches yield excellent results on RNA-Seq datasets.

Davis also presented NB generalized linear model (GLM) methods for the assessment of evidence for differential expression in experiments with complicated designs. The R/Bioconductor package "edgeR" provides a highly efficient software implementation of these novel methods. The package is open-source and freely available and has already proven a useful analytical tool for biological and bioinformatics researchers in Australia and overseas. The sophistication of the analysis using GLM methods offers greater insight into biological questions than existing methods allowed.

Han Gan, who is a first year PhD student in Probability at the University of Melbourne and the student representative on the Victorian branch of the SSAI, then spoke about the *Central Limit Theorem*, which is a fundamental result in statistics that many people use without a second thought. Given a data set, people will often simply calculate the mean, and derive a confidence interval, without thinking about the validity of the normal approximation. Using Stein's method, it is possible to not only prove the central limit theorem,

but also to calculate bounds for the convergence.

The talk began with the classical proof to the central limit theorem with generating functions, and then moved onto some of the motivations for the advantages of Stein's method. Starting with Stein's lemma, it was shown how to derive bounds on the distance between the distribution of one random variable to the normal distribution using the Kolmogorov metric.

Eventually, not only was the classical theorem proven, but a stronger form, Lindeberg's central limit theorem, was derived.

The talk was concluded with a cute open problem that Han was working on, a combinatorial question with applications in card counting.

Carol Soloff ■

Annual Young Statisticians' meeting

Western Australia

Peter Wallis, Ryan
Admiraal, & Rebecca
O'Leary



The May meeting of the WA branch was the annual Young Statisticians' meeting and consisted of two presentations. The first talk on "Environmental modelling of demersal fish species abundances off South Western Australia", was given by Ben Fitzpatrick, the 2010 recipient of the WA branch honours scholarship.. This study examined the distribution of three species of demersal (seabed-dwelling) fish from the *Wrasse* family off the southern coast of Western Australia. The aim of the study was to explore models that provide accurate predictions of the abundance of these fishes using environmental conditions. The data for the study came from the 2007 Marine Futures Project data.

Fish abundance was measured using a baited video system at 58 locations around Cape Naturaliste and 56 locations around Injidup (19km south of Cape Naturalist). To avoid double-counting, only the maximum number of fish of some species appearing on screen at any one time was recorded in each one-hour video. Two explanatory variables were chosen and measured at each site: habitat complexity (a categorical variable with 5 states) and depth. The particular species were "site attached" with a maximum reported home range corresponding to a 23m radius, so measurement sites were separated by at least 250m to ensure independent observations across sites.

Using the Cape Naturaliste data, two generalised linear models were considered for each species to predict fish abundance using habitat depth and complexity. The first model considered was a Poisson GLM. Because this model assumes equidispersion of fish within each habitat, a negative binomial GLM was also considered, as it relaxes this constraint. The models were compared



In foreground, Peter Wallis, Ryan Admiraal, Rebecca O'leary and friend.

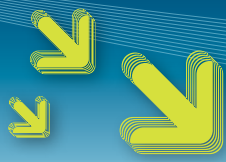
using Pearson and deviance goodness-of-fit tests. Overdispersion was also tested using the overdispersion parameter test (Lawless, 1987). The predictive accuracy of the negative binomial GLM was assessed by testing the model on the Injidup data.

Both goodness-of-fit tests confirmed the Poisson GLM to be a poor fit to the data for each species. The overdispersion parameter test confirmed this result, indicating that overdispersion was present in the data (i.e. the fish tend to aggregate). The negative binomial GLM produced a minor improvement but still provided a poor fit to the data. Depth was found to have a significant negative correlation to fish abundance for *O. lineolatus* and *P. biserialis* in the negative binomial GLM, indicating these species prefer to live in shallower water. The expected abundance of *C. auricularis* was found to be significantly higher in low profile reef and sand inundated reef habitats versus high profile reef habitats. This was in contrast

to the speaker's expectation that fish abundance would increase with habitat complexity. The predictive accuracy of the negative binomial GLM was poor, with high predictive sum of squares for the Injidup data relative to the residual sum of squares for the Cape Naturaliste data.

The second talk, given by Yuichi Yano, was on "Generalised Linear Modelling of the Asthma Hospitalisation Risk and Air Pollutant Co Concentration in Perth". The study examined the relationship between four particular air pollutants and asthma hospitalization rates. The study was the first to consider the spatial distribution of air pollutant concentration in Perth. Due to the small number of pollution monitoring stations in Perth, estimation of this spatial distribution of air pollutant concentration was required.

The air pollutant data for this study came from the Bureau of Meteorology. Concentrations of carbon monoxide (CO), nitrogen dioxide (NO₂) and



Western Australia cont.

particulate matter of two sizes were measured at five monitoring stations around Perth. 2006 emission inventory data was available for four locations in addition to the five monitoring stations, and was thus also recorded to provide better spatial resolution. Wind data was recorded at twelve locations. The asthma hospitalisation data consisted of monthly counts of asthma-related hospital admissions for each postcode over the period 2000-2007.

The statistical analysis had three main components. First, the spatial distribution of each air pollutant was estimated using Gaussian plume modelling (GPM) applied to the measurement stations data and the output smoothed using ordinary lognormal Kriging (LK) of the emission inventory data. Wind, horizontal and vertical distance from measured observations, and solar radiation strength were used to predict the concentration of each pollutant in postcodes for which emissions data was not available in 2006. Second, the monthly asthma hospitalisation risk for each postcode was estimated for the period 2000-2007 from the count data using a Bayesian Hierarchical Model (BHM). Finally, CO concentration was used to predict local asthma hospitalisation risk using a GLM.

CO was most heavily concentrated around the central metropolitan area in 2006, but the distribution varied over the year. On average, concentration was lowest in months where solar radiation was "strong" (December-February), and highest in months where solar radiation was "slight" (June-August). The asthma BHM indicated that hospitalisation risk varies across postcodes, and within each postcode the risk varies over time. The lowest risk was observed in the central Perth metropolitan area, and on average

the risk appeared to be lower in hotter months (December-January). The GLM indicated that there is a significant relationship between CO concentration and asthma hospitalisation in some postcodes. The strength of this influence was found to vary over postcodes.

The June meeting featured a talk entitled "A non-parametric blended satellite-gauge precipitation estimation" by Ming Li of CSIRO. Precipitation estimation is important for both water forecasting and water accounting, and different technologies are available to assist in estimation. These technologies include rain gauges, rainfall radar, and satellite data. Each technology is different in terms of frequency of reports and precision. For instance, rainfall radar has only about a one kilometre resolution and reports measurements every 10 minutes, and satellite-based estimates have an even lower resolution at anywhere from 6 to 25 kilometres with reports being transmitted every half hour.

The goal of Ming's research was to see whether or not blending satellite-based rainfall estimates with rain gauge data could improve the accuracy of daily rainfall estimates, particularly in areas where rain gauge stations are sparse. Two forms of blending currently in use include physical-based blending (which uses radar and rain gauge data and assumes that the radar reflectivity and rainfall intensity follow some physical rule) and kriging (which is a geostatistical-based method that assumes spatial stationarity, i.e. the correlation between two observations is based only on the distance between two observations). Ming wanted to determine whether a non-parametric approach blending satellite and rain gauge data could provide a viable alternative, as radar is typically found only

in metropolitan areas and rain gauges are sparsely distributed through rural regions. In particular, he used single and double smoothing kernel estimation (which generates pseudodata and uses both the original data and pseudo data). Using several different measures of performance (such as cross-validation, performance indicators such as root mean squared error and minimum absolute error, and real-time and post-real-time dataset validation), he found double smoothing kernel estimation to significantly improve estimates of daily rainfall over other methods when there are big rainfall events and to provide worse estimates when there is no rainfall.

Clair Alston from the Queensland University of Technology (Brisbane) gave an awesome talk in the July meeting of the Western Australia branch. This presentation was about the new Python package that Clair and her colleagues Chris Strickland, Robert Denham and Kerrie Mengersen are writing. Bayesian analysis uses Markov chain Monte Carlo (MCMC) estimation as a solution to complex integration problems. Typically, however, implementation of MCMC is code intensive and time consuming. PyMCMC is awesome as it solves these problems. This function applies Bayesian analysis to model applied data from experimental studies such as in medicine, environmental, ecological and agricultural research. In particular, it is suitable for large datasets such as CT scans, remote sensed imagery, etc. This package is awesome as the type of models it includes are linear regression, generalised linear regression, linear mixed, generalised linear mixed regression, non-linear mixed, and time series. It contains three types of Metropolis Hastings: Metropolis Hastings,

independent Metropolis Hastings and random walk Metropolis Hastings.

PyMCMC is 4 to 10 times faster than R library MCMCglmm, particularly for more complicated models, e.g. mixture models. This function can also be easily extended to other models. Python is FREE open source software. Another awesome feature is that the software identifies realistic initial values, unlike WinBUGS which typically gives unrealistic initial values.

Therefore this function is AWESOME. This function will be released later this year. I know I can't wait. If you are interested in PyMCMC then email Clair.

Peter Wallis, Ryan Admiraal ■

“BAYES FOR BEGINNERS”

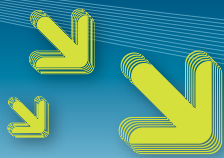
Workshop by Rebecca O'Leary

The Western Australian branch together with The University of Western Australia Centre for Applied Statistics offered two mini-workshops each of two days on 19th-22nd July 2011: Introductory Bayesian analysis using WinBugs and Bayesian analysis using R. This course was given by Professor Kerrie Mengersen and Clair Alston from Queensland University of Technology.

Topics covered included an introduction to Bayesian models, Bayesian computation via MCMC, and Bayesian networks. Practical sessions will include the use of WinBugs for Bayesian analysis, Bayesian analysis in R and GeNie for Bayesian networks. The course was deliberately aimed at a beginner level and was strongly practical in nature.

15 people attended the first course and 21 people the second. The course was well received. Even had one person, who was only going to attend the first course, after attending the first two days enrolled for second course.

Rebecca O'Leary ■



THREE STATISTICS COURSES

with a mix of lecture presentations and practical work



Design & Analysis of Experiments

22 - 27 September 2011 (4 days)

presented by
Dr. Graham Hepworth

Learn about the principles and practice of designing experiments, and the analysis of experimental data. Applications include randomised trials in medicine or the social sciences, designed experiments in the biological sciences, studies of processes in engineering, as well as many other possibilities in other disciplines.



Outline and applications:
www.scc.ms.unimelb.edu.au/dae.html
Venue: The University of Melbourne
\$990

Statistics for Research Workers

9 - 16 November 2011 (6 days)

presented by
**Associate Professor Ian Gordon
and Dr. Sue Finch**

This very popular course gives a basic understanding of statistical ideas and methods involved in carrying out research.

Statistical topics covered will include estimation, hypothesis testing, regression and inference based on the normal distribution.



Outline and applications:
www.scc.ms.unimelb.edu.au/srw.html
Venue: The University of Melbourne
\$1210

Statistical Consulting Centre

Contact Deborah Maillardet: dmai@unimelb.edu.au Phone: 03 8344 6995

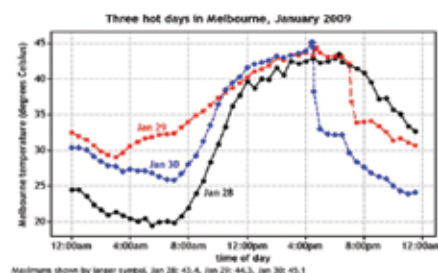
Producing Excellent Graphics Simply

23 November 2011 (1 day)

presented by
Dr. Sue Finch

Workshop designed to show participants the principles and practice of effective graphics.

The course covers how to produce graphs of data and graphs of statistical inferences, simply, quickly and effectively.



Outline and applications:
www.scc.ms.unimelb.edu.au/pegs.html
Venue: The University of Melbourne
\$330