

Summary of the International Congress on Peer Review and Biomedical Publication, 10-12 September 2009

by Irene Hames

Managing Editor

The Plant Journal, Wiley-Blackwell

imb5@york.ac.uk

The Sixth International Congress on Peer Review and Biomedical Publication was held at the beginning of September in Vancouver, and more than 400 delegates from 34 countries gathered to hear and discuss 49 presentations and view more than 60 posters over 3 days. There was a buzz in the air and expectations were high.

The Congresses were introduced to provide a forum for the presentation of research on peer review and are organised by *JAMA* (*Journal of the American Medical Association*) and the *BMJ* Group. The first Congress, 'Guarding the Guardians', took place in Chicago in 1989, and the meetings have been held every four years since. Only new research can be presented, and equal time is given to discussion. In his summing-up session at the end of the meeting, Drummond Rennie, the Congress Director, gave an overview of the previous Congresses, highlighting a number of important studies that have come out of them. He stressed that the Congresses exist to promote research, and in 2005 he did what he said he would do when the research presented reached a certain level: direct that the presentations would no longer be published in a special issue of *JAMA*, but study authors should rather submit their research for publication to various journals. Although there were still many challenges, he felt the Congresses had seen a number of victories, notably the introduction of reporting structures such as SORT and CONSORT (note: a new CONSORT

Statement is due to be published next year), and the requirement for clinical trial registration by the ICMJE journals.

EQUATOR Pre-meeting Satellite Event

The EQUATOR network (see May 2008 issue of *EON*) held a successful and well-attended reporting workshop the day before the Congress and hosted its second Annual Lecture in the evening: *Redescribing medicine: reporting or reclaiming research for health?* by Richard Horton, editor-in-chief of *The Lancet*. This was a fascinating talk, delivered in Horton's usual passionate and thought-provoking way. He likened journals to systems of justice (drawing on concepts in Amartya Sen's 2009 book, *The Idea of Justice*), stressing that journals are more than repositories of research – they have personalities. I liked this idea, as I think it's very true. He questioned whether journals accurately describe the world of health as it truly is and then went on to talk about research reporting, asking 'How can we all help EQUATOR to become a global network?' He was concerned about publication bias against low-income countries and stressed that it is the responsibility of editors to ensure that divergent perspectives are given space in their journals.

The Congress

And so to the Congress proper. All the sessions were plenary sessions – 10-minute presentations followed by 10 minutes for

Congress on Peer Review

continued

discussion based on questions from the audience – so no difficult choices needed to be made as to which to attend. The sessions were well attended throughout, even on the last day and despite the meeting being held in such a beautiful city with so many potential distractions. The Congresses concentrate on biomedical publications and, although there were sessions on authorship/ contributorship, peer review, data sharing and conflicts of interest, editorial training, decisions, policies and ethics, publication pathways and publication bias, reporting in relation to clinical trials predominated – there were three sessions on this, plus one on trial registration. Coming from a science journal, I think I was in the great minority. I would have liked more on other aspects of peer review. But, as Drummond Rennie explained, the committee can only select presentations from what's submitted. He, like others, had been surprised that there was so little electronic-related material and urged everyone to 'start working'.

What Were the Highlights of the Meeting?

Invited Presentations

As well as the research presentations, each day started with a longer invited talk, and these were very good.

Steven Goodman opened the first day with an entertaining and stimulating talk entitled *Jumping without Parachutes: numbers, knowledge and biomedical journals*, focussing on how things work in the reporting of studies. This included both humorous and serious examples of premature or inaccurate causal pathways (from the 3.9-year survival advantage of Oscar winners to the protective effect of Omega-3 oils from different sources), which are frequently reversed in subsequent studies, and the claim that medical journals are still predominantly using an 80-year-old statistical technology. We are witnessing an

'avalanche of new methods' entering the field, but Goodman felt they are often being used by poorly trained people and being reviewed by those who are even less knowledgeable. A couple of interesting points came up: the tension between what is known before a study and what is found, and the phenomenon of 'data peeking', where if you look at the data during a study you're more likely to find the effect you expect – so don't.

Harold (Hal) Sox gave a very nice and topical talk at the start of the second day on *Comparative Effectiveness Research*. He has recently retired from 8 years as the editor of *Annals of Internal Medicine* and gave an overview of four important lessons he'd learned during his editorship: (1) most clinical research is either not novel or not valid (i.e., conclusions not fully supported by the evidence) and unlikely to change patient care; (2) journal editing has diverse 'receptor sites' and these groups expect different things (e.g., the public, a stamp of approval on clinical research; the author's peers, transparent reporting and details of what was done and found; the authors, valuable help from expert editors); (3) predicting which articles will be influential is an imperfect science; and (4) to minimise mistakes in what they publish, journals need statisticians who are active participants in the culture of the journal, and this involves a big financial investment.

The main focus of Sox's talk was Comparative Effectiveness Research (CER), which compares the benefits and harms of different treatments and identifies interventions that work in improving health care. A lot of money is being put into this in the United States in the form of stimulus funding (see www.hhs.gov/recovery/programs/cer/index.html) and he saw this as both an opportunity and a challenge for medical journals. An opportunity because medical journals will have a central role in CER, a challenge because there will be a large

Congress on Peer Review

continued

increase in the number of publications, some of which will involve complex modelling, and this will impact peer review. Medical journal editing is expensive, and he asked the question: 'Who will pay for this in the future?' He sees free access to journal content as a threat to medical journals' traditional income sources, and if journal income falls he is concerned about what will happen to first-class journal editing. (Note: Sox used his farewell editorial in the 7 July issue of *Annals of Internal Medicine*, vol 151, pp 68-69 to cover this.)

Drummond Rennie's historical overview of the Congresses (see above) opened the third and final day.

Research Presentations

With 46 research presentations it's not possible to highlight more than a few. The abstracts are all available on the Congress web site (www.ama-assn.org/public/peer/peerhome.htm), but quite a few speakers included updated results so these abstracts don't represent exactly what was presented. It's important to remember that although the presentations were selected for inclusion by a double-blind review process, they haven't yet been subject to scrutiny by journal peer review. A number of the presentations have already been reported in the press and on blogs; some have received a lot of coverage, including some criticism (see below).

Authorship. The first research presentation on the opening day (Thursday 10th, morning) reported an interesting study with important implications for the perception of authors' contributions. Little is known about how readers assign credit based on author position and corresponding-author (CA) designation. Jason Busse and colleagues (Bhandari *et al.*), therefore, looked at how chairpersons in departments of

surgery and medicine across North America perceived author contribution (in study conception and design, analysis and interpretation of data, and statistical analysis) based on position in the author listing and CA status, and asked their view of the most prestigious position. They presented a hypothetical study with five authors to nearly 300 chairpersons, making either the first or last author the CA. From the 165 responses (57% response rate), they found that when the last author was the CA, perception of the first author's contribution decreased significantly for a number of areas, whereas their own overall prestige increased significantly. The respondents varied widely in their perceptions of the contributions of the other authors, irrespective of who was the CA. The second author's contribution was affected by whether the first or last author was the CA. Academic department chairs were, therefore, influenced by CA designation. The authors of the study concluded that without explicit details of the contributions of authors, false conclusions can be drawn about author credit and accountability.

Joseph Wislar (Thursday 10th, morning session) presented the results of a study (Wislar *et al.*) to determine the prevalence of honorary and ghost authorship in general medical journals and how this compared to the results of an earlier study in 1996. He and his three co-authors, who are *JAMA* editors, surveyed the corresponding authors of 900 articles of various types published in 2008 in six general medical journals (*Annals of Internal Medicine*, *JAMA*, *The Lancet*, *Nature Medicine*, *New England Journal of Medicine (NEJM)*, and *PLoS Medicine*) about honorary and ghost authorship and obtained 630 responses (about a 70% response rate). Honorary authors were reported for 21% of articles, ghost authors for 8%, and both for 2% (cf, 19%, 11% and 2%, respectively, for results in the 1996

Congress on Peer Review

continued

survey). The prevalence of honorary authors was highest in *Nature Medicine* (39%) and lowest in *NEJM* (16%), and the prevalence of ghost authors was highest in *NEJM* (11%) and the lowest in *Nature Medicine* (2%). They concluded that the prevalence of honorary and ghost authors has not changed significantly since 1996 and found this worrying given the increased awareness of authorship issues and the introduction and publication of author contributions by some medical journals. However, they found no significant differences between journals requiring author contribution disclosures and those that do not.

The study has been the subject of much discussion, and the *NEJM* has been reported to be sceptical about the findings and concerned about some of the definitions and methodology. Ginny Barbor, chief editor of *PLoS Medicine*, was at the meeting and found the results ‘quite shocking’ because her journal has tough and explicit policies on ghostwriting and contributorship, and she felt they’d ‘been lied to by authors’.

Peer Review. There were only three presentations in this session (Thursday 10th, morning), which was a bit disappointing—one on grant review, the other two on peer reviewers and review quality. The study presented by Michael Callaham (*The natural history of peer reviewers: changes over time*) has also generated some interest in the media but maybe shouldn’t be taken at face value. Callaham wanted to see if reviewer performance changes as reviewers get older. So all the reviews for *Annals of Emergency Medicine* (where he is EiC) from 1994 to 2008 (14,808 by 1498 reviewers), which had been rated on a five-point scale by 84 editors, were assessed. The performance of 93% of the reviewers deteriorated over time at a mean rate of 0.04 points a year. He also looked to see if doing more reviews over time led to better reviews – it

didn’t. Kent Anderson has written a good post on this study on The Scholarly Kitchen blog highlighting the limitations and suggesting other interpretations; it’s worth a look (<http://scholarlykitchen.sspnet.org/?s=vancouver+peer+review>). Kent Anderson also comments that ‘attempts to make scientific peer-review more scientific seem misguided’ and ‘the purpose of tools like these eludes me’.

Editorial Policies. An important study was presented by Benjamin Djulbegovic (Wager *et al.*) on the Thursday afternoon in the editorial policies session. In July 2005, *JAMA* introduced a new policy requiring independent statistical analysis for industry-funded randomized controlled trials (RCTs). The authors looked at all the RCTs published in *JAMA* and two other journals as a control (*The Lancet* and *NEJM*) from July 2002 to June 2008. They found that the total number of RCTs and the proportion with commercial funding decreased significantly in *JAMA* after July 2005. During the same period, the number of RCTs published in *NEJM* increased, but funding did not change significantly; the same number were published in *The Lancet*, but the proportion of industry-funded RCTs rose (but not significantly). The authors concluded that *JAMA*’s requirement for independent statistical analysis for industry-funded studies was associated with a change in pattern of RCTs published, but they were unable to tell whether the policy had affected the number submitted, the acceptance rate (i.e., quality), or both. Fiona Godlee, editor-in-chief of the *BMJ*, made the point from the floor that if it could be shown that the quality had improved, other medical journals would be ‘flocking to do the same’. Catherine DeAngelis, the *JAMA* editor-in-chief, got up to explain the policy and reasons why she’d introduced it. The aim had been to protect the public against poor studies and she had no regrets about introducing it.

Congress on Peer Review

continued

Publication Bias. An interesting study was presented by Seth Leopold (Emerson *et al.*) in the Friday morning session on publication bias. The authors wanted to see whether more reviewers would recommend publication of a 'positive' outcome report compared to a 'no-difference' outcome report of an otherwise identical fabricated RCT. Positive-outcome publication bias can skew evidence-based medicine, so this is an important area of study. The two versions of the reports they prepared (on dosage and timing of perioperative antibiotics) were identical except for outcome and were randomly allocated to 209 reviewers at two orthopaedic journals - *Journal of Bone and Joint Surgery (JBJS)* and *Clinical Orthopaedics and Related Research (CORR)*. At both journals the 'positive' outcome manuscript was more likely to be recommended for publication (but only significantly at *JBJS*), suggesting that positive-outcome bias is present. A surprising, and worrying, finding was that the 'positive' outcome manuscript received better reviews for methodological rigour, even though the methods sections of the two versions were identical, and was checked less carefully for errors. The authors recommended that editors should consider providing reviewers with more explicit guidelines for the review of 'no-difference' manuscripts.

Rhetoric. The Friday afternoon session started with an unusual but interesting session on 'rhetoric', where research involving linguistic analysis of the use of words and phrases in published articles was reported. Studies were presented by Isabelle Boutron on the nature and frequency of 'spin' ('the manipulation of the content and rhetoric of reporting to convince the reader of the likely truth of a result') in published reports of RCTs with nonstatistically significant primary outcomes, and by Lisa Bero on the rhetoric used to frame research results reported in drug studies to determine if statistically significant numerical data support

claims about the effectiveness or safety of a drug. The final presentation, by Eileen Gambrill, advocated the introduction of a 'propaganda index'. The principle behind this was commendable, namely to devise an index to serve as a complement to methodological filters such as CONSORT in reviewing manuscripts and the published literature, which could identify and reduce spin and inappropriate claims and emphasis. I didn't, however, like the use of the term 'propaganda index' in this context, as it denotes a rather extreme and doctrinaire activity and has derogatory and political connotations, which makes it unsuitable for general and international use. In submissions to most journals authors do try to hype their findings to make their data seem more important than they are and help get their work published, but many times authors are just sloppy in their use of language. In both cases it's up to reviewers and editors to spot inflated claims and inappropriate use of language and get the authors to address these issues. The problem is of course much more serious in the biomedical literature because claims about drug effectiveness and safety impact health care, and positive results in industry-sponsored studies can result in great financial returns. So an index that can help reviewers and readers detect exaggerated and unsupported claims could be valuable.

Posters

There were many interesting posters – covering a much wider subject base than the actual presentations – but, unfortunately, they weren't on display for the whole meeting and there wasn't enough time to view them all and discuss the findings with the poster presenters. I only managed to look at a fraction and would have loved to have seen them all. 🌐