

COMMENTARY

Open Access



# Cochrane: the unfinished symphony of research synthesis

Ian Roberts\* and Katharine Ker

## Abstract

The NHS needs valid information on the safety and effectiveness of healthcare interventions. Cochrane systematic reviews are an important source of this information. Traditionally, Cochrane has attempted to identify and include all relevant trials in systematic reviews on the basis that if all trials are identified and included, there should be no selection bias. However, a predictable consequence of the drive to include all trials is that some studies are included that are not trials (false positives). Including such studies in reviews might increase bias. More effort is needed to authenticate trials to be included in reviews, but this task is bedevilled by the enormous increase in the number of 'trials' conducted each year. We argue that excluding small trials from reviews would release resources for more detailed appraisal of larger trials. Conducting fewer but broader reviews that contain fewer but properly validated trials might better serve patients' interests.

## Background

The UK National Institute for Health Research (NIHR) provides public funding for systematic reviews because the information they provide informs healthcare decision making in the NHS and the commissioning of new research. For both purposes, it is essential that the information reviews provide is valid and up-to-date. In October 2015, NIHR announced an evaluation of its investment in Cochrane infrastructure for the production and dissemination of reviews. It will consider the health and economic impact of Cochrane reviews by 'assessing the quantity, quality and impact of reviews on policy, practice and research'. Of the parameters assessed (quantity, quality, impact), quality is the most important. Systematic reviews will not bring health and economic benefits if their conclusions are misleading. We consider the main threats to validity in systematic reviews and how they can be minimised cost-effectively.

The conclusions of a systematic review can be misleading if it includes a biased sample of trials (selection bias in study identification and inclusion) and/or the effect estimates in the included trials are biased. Validity requires an unbiased sample of unbiased effect estimates. Neither criterion is easy to satisfy. About half of

all trials are not published, and the results of published and unpublished trials differ systematically [1, 2]. Even if all trials could be identified, selective reporting of outcomes remains an important source of bias [3]. Because of poor methodology, the results of many trials are biased, and incomplete or inaccurate reporting of trial methods frustrates quality assessment.

## Find all the trials

Since its inception, Cochrane has attempted to avoid selection bias in study identification and inclusion by conducting exhaustive searches for all trials, published or unpublished, irrespective of language of publication. If all trials are included, there should be no such selection bias. A 'highly sensitive search strategy' was developed to identify all trials in the main electronic bibliographic databases [4]. Because many trials in indexed journals were not coded as such by the National Library of Medicine, trial identification was supplemented by hand-searching. This resulted in a major initiative to re-tag records that were not properly coded. The Cochrane Central Register of Controlled Trials (CENTRAL) launched in 1996 includes trials identified in electronic databases and by hand-searching [5, 6]. Trial identification has become a highly specialised activity, and almost all Cochrane Groups have a dedicated trial search co-ordinator to undertake this role. However, despite extensive efforts to identify and include all relevant

\* Correspondence: [Ian.Roberts@shtm.ac.uk](mailto:Ian.Roberts@shtm.ac.uk)  
Clinical Trials Unit, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

trials, selection bias (in study identification and inclusion) still casts doubt on the conclusions of many reviews.

### The problem of false positives

A predictable consequence of the drive to find and include all trials is that some studies are included that are not trials (false positives). Maximising sensitivity often decreases specificity. Many reports claiming to be 'randomised trials' are not in fact randomised, often owing to a lack of an understanding of trial design among the authors and sometimes due to downright deception. An interview study of Chinese authors of so-called randomised controlled trials found that only 7 % were authentic [7]. Even for trials conducted at university-affiliated hospitals, only 56 % were authentic [7]. Although it would be inappropriate to assume that the prevalence of 'false positives' is as high in other settings, it would also be inappropriate to ignore these findings and take trials at face value without conducting further checking.

Meta-analyses of baseline variables in systematic reviews of randomised trials often reveal surprising imbalances suggesting that randomisation was either subverted or absent [8, 9]. And some reviews contain fraudulent data. A Cochrane review showing that high-dose mannitol reduced the risk of death after head injury was rewritten after an investigation was unable to confirm that three of the included trials took place [10]. The conclusions of a review of starch solution in critically ill patients changed importantly after excluding seven 'trials' from an investigator whose research was retracted due to misconduct [11, 12]. Many journal editors and systematic reviewers take trial reports at face value with little or no effort to confirm whether a trial actually took place or how reliably it was conducted. A recent survey of authors of systematic reviews found that 38 % had no contact with the authors of the original studies [13]. Indeed, the amount of contact with authors is rarely reported in reviews. Checking with ethics committees that ethical approval was obtained for a trial is one of the few ways to obtain independent confirmation that a trial took place and even then is not completely reliable. However, very few (3 %) reviewers check whether the included studies had such approval [13]. Unless sensitive searching is accompanied by similarly rigorous efforts to ensure the integrity of the included trials, the precision gained by including a larger number of 'trials' could be outweighed by an increase in bias.

NIHR funding for Cochrane Groups is proportional to the number of reviews published and the number of trials included in reviews. Both criteria have seriously detrimental unintended consequences. First, it incentivises the fragmentation of evidence. For example, there are over 100 Cochrane reviews on the treatment of hypertension, including separate reviews of treatment trials in people with and without diabetes, rather than a single

review assessing whether diabetes is an effect modifier. Second, there is a financial incentive to include every study purporting to be a 'trial' with no incentive to root out false positives.

### The challenge of validating trials

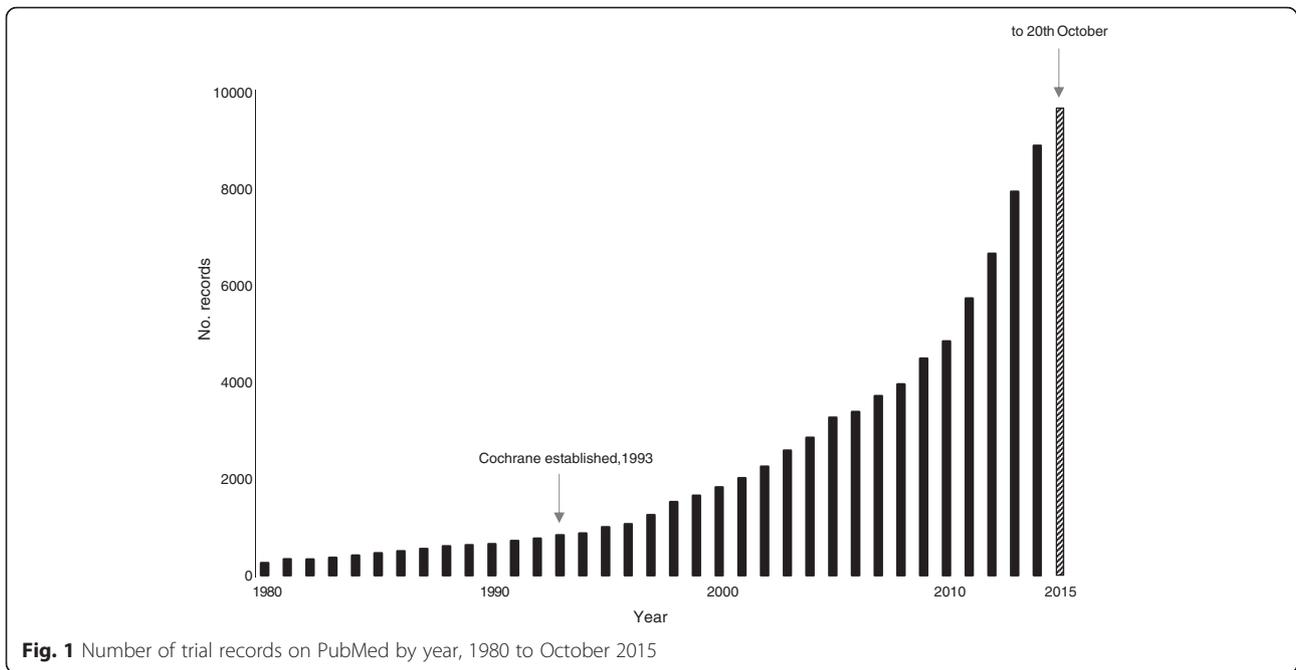
Unless rigorous efforts to identify trials are accompanied by at least as much effort to assess their integrity and completeness, extensive searching may have the unintended consequence of *reducing* the reliability of reviews. Results from individual patient data meta-analyses are likely to be more reliable in this respect since they involve checking the integrity and completeness of all included data. However, they can be challenging and costly and are not immune from selection bias [14]. Nevertheless, simpler checks are possible. We believe that authors should be required to provide evidence that the trial actually took place and that the participants were properly randomised, that there were no post-randomisation exclusions and no selective reporting. This could be combined with the use of software to detect plagiarism and statistical data checking (e.g. detection of extreme between study homogeneity). Reviews should routinely report their validation strategies, including which trials were confirmed and which authors were unresponsive.

However, the task of identifying and authenticating all potentially relevant trials is bedevilled by the enormous increase in the number of 'trials' conducted each year (Fig. 1). In 1993, the year the Cochrane Collaboration was founded, 823 trials were indexed in PubMed. Twenty-two years later, the corresponding figure is 9673 trials. If efforts to prevent bias through the complete ascertainment of trials have failed to date, it is unlikely they will succeed in future without a considerable increase in resources. Many Cochrane reviews are already out-of-date and the situation can only be expected to get worse.

### Small trials—are they worth the effort?

Most trials are small. Over half of registered trials include fewer than 100 participants (Fig. 2). Many important health outcomes (death, myocardial infarction, stroke, sepsis, pneumonia) are dichotomous and occur in only a small proportion of trial participants. Most small trials register only a few such outcome events in each arm. With such sparse outcome data, there is a substantial risk of selective under-reporting of trials and patient outcomes. This is a major threat to validity. One fifth of Cochrane meta-analyses with a statistically significant result became non-significant after adjusting for outcome reporting bias [3].

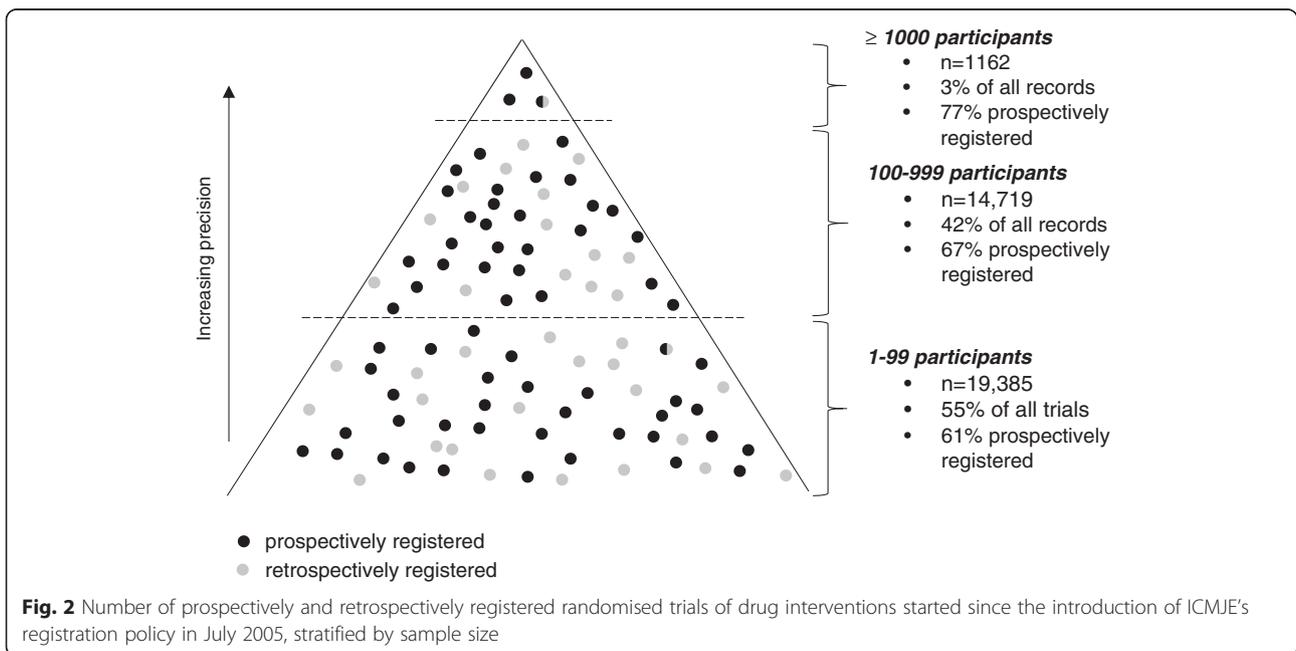
Meta-epidemiological studies show that small, single-centre trials generally provide larger estimates of treatment effects than large, multi-centre trials [15–18]. This



could be due to strict patient selection and better intervention compliance in small trials. However, it is often due to the poor quality of small trials and their greater risk of selection bias. Furthermore, small trials that are stopped early for apparent benefit often provide implausibly large effects (the overestimates are smaller in large trials) [19, 20]. Random effects models exacerbate small study bias. In the presence of heterogeneity, random effects models give greater weight to small studies, which are more susceptible to bias. Small studies with

large effects appear to ‘anchor’ the meta-analysis such that the result is largely unchanged regardless how large the subsequent trials. When meta-analyses are restricted to larger studies, treatment effects that appear large and statistically significant when all trials are combined usually become smaller [16]. This suggests that including all apparent trials in systematic reviews can increase rather than decrease bias.

The well-documented unreliability of small trials offers an opportunity to increase the validity of reviews whilst



reducing the burden and cost of conducting and maintaining them. Although large trials provide more information about the treatment effect than small trials, the time and effort required to evaluate a trial is often inversely related to sample size. Limiting inclusion to larger trials would release resources that could be redirected to a more thorough critical appraisal of the trials that provide the most information. The sample size cut-off would clearly depend on the objectives of the review and the outcomes of interest. The information content of a trial depends on the number of events rather than the number of participants. Quantitative research is needed on the sample size cut-off that would provide an optimal balance of sensitivity and specificity, and on other markers of unreliable data.

The exclusion of grossly underpowered trials from systematic reviews and meta-analyses should reduce bias, but it is not a guarantee. A more reliable approach would be to identify trials for inclusion in systematic reviews from trial registers [21]. Systematic reviews that entail complete ascertainment of results from all (or an unbiased sample of) prospectively registered trials (i.e. trials registered before the first patient is enrolled) should not be affected by selection bias in trial identification and inclusion. Indeed, preventing such selection bias is one of the main purposes of registers, and trials with large sample sizes are more likely to be prospectively registered (Fig. 2). However, including only prospectively registered trials is not sufficient. Trials that were prospectively registered but not reported must be sought out and included to avoid bias. Similarly, data on health outcomes that were collected but not fully reported must be pursued and included. This will take time and effort and will be greatly facilitated by excluding the myriad small trials that provide negligible amounts of information.

### Counter arguments

Some argue that including all trials, regardless of size or quality, allows reviewers to draw attention to the scandal of low-quality, underpowered trials [22]. We agree that the main contribution of systematic reviews has been to highlight the miserable unreliability of most biomedical research. However, in the light of the methodological advances from meta-epidemiological studies, it is questionable whether highlighting poor quality remains a legitimate use of public funds. Indeed, including such trials in reviews gives them unwarranted endorsement.

Another argument for including small trials in reviews is that the combined results from small trials often motivate larger high-quality studies. For example, Chalmers and Glasziou argue that 'fundlers and regulators cannot be expected to support and endorse large studies without some reassurance from the results of smaller existing studies that the substantial investment needed is

justified' [23, 24]. Whilst we agree that investment in new research should be preceded by systematic assessment of existing evidence, it is essential to avoid making funding decisions that are heavily influenced by biased research. Because effect estimates from systematic reviews often inform sample size calculations, there is a danger that inflated effect estimates from reviews of small trials motivate new trials that are underpowered to detect realistic treatment effects. For example, a Cochrane review of randomised trials of the effect of preoperative statins on the risk of post-operative atrial fibrillation included 17 small trials with a total of 2138 participants and reported a halving of the odds of post-operative atrial fibrillation with statin treatment (OR = 0.54; 95 % CI 0.43 to 0.67;  $p < 0.01$ ) [25]. This claim was later refuted by a randomised trial (the Statin Therapy in Cardiac Surgery (STICS) trial) that included more outcome events (cases of atrial fibrillation) than all the previous trials combined that found no reduction in atrial fibrillation (OR = 1.04; 95 % CI 0.84 to 1.30;  $p = 0.72$ ) [26]. Fortunately, the STICS trial investigators were appropriately sceptical of the large treatment benefits suggested by the Cochrane review of small trials and conducted a trial with sufficient power to exclude a more plausible effect. However, although the new trial appears to refute the conclusion of earlier smaller trials because of the anchoring effect of small studies with large treatment effects, the updated meta-analysis will still suggest a significant reduction in atrial fibrillation with statin treatment, a conclusion that seems unlikely.

### Conclusion

The NHS needs valid information on the safety and effectiveness of healthcare interventions. This information must be provided cost-effectively. Although investment in systematic reviews can be more cost-effective than conducting new trials, this should not rule out consideration of how to improve the cost-effectiveness of conducting reviews. We argue that attempting to identify and include all apparently relevant trials might increase rather than decrease bias and may not be the most cost-effective approach. Many supposed 'trials' are not in fact randomised trials. The gulf between individual participant data meta-analyses, where all data is thoroughly checked for accuracy and completeness, and systematic reviews based on published data from apparent trials is too wide. More effort is needed to validate trials and obtain data on unreported outcomes. Excluding grossly underpowered small trials from reviews might increase validity and release resources for more detailed appraisal of included trials. NIHR incentives to conduct narrowly focused reviews including everything purporting to be a trial should be re-considered. Incentives to conduct fewer but broader reviews that contain fewer but properly validated trials might better serve patients' interests.

**Abbreviations**

NHS, National Health Service; NIHR, National Institute for Health Research

**Acknowledgements**

The authors thank Anthony Rodgers and Iain Chalmers for providing detailed comments on the manuscript.

**Funding**

Ian Roberts is an NIHR Senior Investigator. This article was prepared in response to the NIHR evaluation of Cochrane infrastructure spending.

**Availability of data and materials**

Not applicable

**Authors' contributions**

Both authors discussed the issues. IR prepared the first draft which was revised following the comments from KK. KK prepared the figures. Both authors approved the manuscript.

**Authors' information**

Ian Roberts is a Professor of Epidemiology and Public Health, and Katharine Ker is a Lecturer at the London School of Hygiene & Tropical Medicine.

**Competing interests**

Ian Roberts is the co-ordinating editor of the Cochrane Injuries Group.

**Consent for publication**

Not applicable

**Ethics approval and consent to participate**

Not applicable

Received: 4 April 2016 Accepted: 17 June 2016

Published online: 14 July 2016

**References**

- Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, Hing C, Kwok CS, Pang C, Harvey I. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*. 2010;14:8. [http://www.journalslibrary.nihr.ac.uk/\\_\\_\\_data/assets/pdf\\_file/0005/64751/FullReport-hta14080.pdf](http://www.journalslibrary.nihr.ac.uk/___data/assets/pdf_file/0005/64751/FullReport-hta14080.pdf). Accessed July 2016.
- <http://www.alltrials.net/>. (Accessed 29 Jan 2015)
- Kirkham J, Dwan K, Altman D, Gamble C, Dodd S, Smyth R, Williamson P. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365. doi:10.1136/bmj.c365.
- Lefebvre C, Glanville J, Wieland S, Coles B, Weightman A. Methodological developments in searching for studies for systematic reviews: past, present and future? *Systematic Reviews*. 2013;2:78.
- Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, Enkin E, Dickersin K. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. *Control Clin Trials*. 1986;7:306–24.
- Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Control Clin Trials*. 1985;6:306–17.
- Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials*. 2009;10:46. doi:10.1186/1745-6215-10-46.
- Clark L, Fairhurst C, Hewitt CE, Birks Y, Brabyn S, Cockayne S, Rodgers S, Hicks K, Hodgson R, Littlewood E, Torgerson DJ. A methodological review of recent meta-analyses has found significant heterogeneity in age between randomized groups. *J Clin Epidemiol*. 2014;67:1016–24.
- Clark L, Fairhurst C, Cook E, Torgerson DJ. Important outcome predictors showed greater baseline heterogeneity than age in two systematic reviews. *J Clin Epidemiol*. 2015;68:175–81.
- Roberts I, Smith ES. Doubts over head injury studies. *BMJ*. 2007;334:392–4.
- Bunn F, Alderson P, Hawkins V. Colloid solutions for fluid resuscitation. *Cochrane Database Syst Rev*. 2001;2, CD001319.
- Zarychanski R, Abou-Setta A, Turgeon A, Houston B, McIntyre L, Marshall J, Fergusson D. Association of hydroxyethyl starch administration with mortality and acute kidney injury in critically ill patients requiring volume resuscitation: a systematic review and meta-analysis. *JAMA*. 2013;309(7):678–88.
- Elia N, von Elm E, Chatagner A, Pöpping D, Tramèr M. How do authors of systematic reviews deal with research malpractice and misconduct in original studies? A cross-sectional analysis of systematic reviews and survey of their authors. *BMJ Open*. 2016;6:e010442. doi:10.1136/bmjopen-2015-010442.
- Ahmed I, Sutton A, Riley R. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*. 2012;344:d7762.
- Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ*. 2013;346: f2304. doi:10.1136/bmj.f2304.
- Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ*. 2010;341:c3515.
- Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiological study. *Ann Intern Med*. 2011;155:39–51.
- Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ*. 2012;344, e813.
- Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203–9.
- Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303:1180–7.
- Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E. The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ*. 2015;350:h2463.
- Altman D. The scandal of poor medical research. *BMJ*. 1994;308:283.
- Roberts I, Ker K. How systematic reviews cause research waste. *Lancet*. 2015; 386:1536.
- Chalmers I, Glasziou P. Systematic reviews and research waste. *Lancet*. 2016; 387:122–3.
- Kuhn EW, Slottoch I, Wahlers T, Liakopoulos OJ. Preoperative statin therapy for patients undergoing cardiac surgery. *Cochrane Database Syst Rev*. 2015; 8:CD008493.
- Zheng Z, Jayaram R, Jiang L, Emberson J, Zhao Y, Li Q, Du J, Guarguagli S, Hill M, Chen Z, Collins R, Casadei B. Perioperative rosuvastatin in cardiac surgery. *N Engl J Med*. 2016;374:1744–53.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

