Journal of Clinical Epidemiology 115 (2019) A6−A8

EDITORIAL

Can university medical center trial investigators do more to ensure timely publication of clinical trial results?

The all−trials initative [http://www.alltrials.net/find-out-more/why-this-matters/the-alltrials-campaign/] calls on governments, regulators and research bodies to register all trials past and present and report the full methods and the results within 12 months on the completion of the trial. Without this information, the results of these trials could be lost forever to doctors and researchers, leading to missed opportunities for good medicine, and trials being repeated. The initiative also calls for all universities, ethics committees and medical bodies to enact a change of culture, recognise that underreporting of trials is misconduct and police their own members to ensure compliance. In this issue the paper by Wieschowski et al reports on the performance of university medical centers in Germany. They found that the performance was poor, with over 40% of trial results not being published within 24 months and 25% still unpublished at 6 years after trial completion. The authors point out the ethical concerns of the burden and risks to participants not generating any knowledge gain. They suggest strategies to improve these statistics including regular public reporting, providing rewards to those reporting within the 2 year standard such as performance-allocated funds for new projects, and even sanctions for those that do not manage to report at least summary results in the registry within 24 months after trial completion.

Opt-in/opt−out design trials are being increasingly seen in order to address ethical concerns in RCTs of variations of health services in a community. This can result in substantive differences between experimental and control groups. Ng et al report on an Australian RCT looking at the utilisation and costs of adding health coaching to a universal care package; the characteristics of the control group and experimental group were well matched at baseline, but sizable proportions of the patients 'opting−in' to the coaching option in the control group, and patients 'opting-out' of the health coaching option in the experimental group, resulted in a biased sample that did not allow a fair comparison. This was resolved by a 'nearest neighbor' matching approach that these authors suggest can be used to look at efficacy issues in future studies with likely substantive selection bias.

Gkioni et al. address the challenge that trials failing to recruit the designated sample size are all too common-this has been reported to occur in up to 45% of trials in the UK. Decisions to fund clinical trials need to take into account the balance of factors including the importance of the clinical question and the time and costs required. One approach to attempt to avoid such failures has been the development of statistical models that can incorporate information for one or more of time-dependent factors such as staggered center initiations and seasonal variations, and ability to specify rates per center. Thirteen eligible articles were identified of which 12 focused on stochastic approaches, and two on deterministic models. Deterministic models require specification of few parameters but are less realistic although easy to implement. They argue that recruitment should be viewed as a stochastic rather than a deterministic process because the costs are highly driven by the length of time to recruit to the target. However, stochastic models require greater parameter specification, which, along with greater complexity may be a barrier to their implementation. It will be interesting to see if the use of these models reduces the number of trials failing to meet their recruitment targets.

Clinical prediction scores and decision rules that accurately identify those with a poor prognosis are much liked, especially for triaging in primary and emergency care. However their accuracy requires validation by testing in different populations and their benefit and cost-effectiveness need to be demonstrated in randomised trials - this is a major and expensive undertaking. Decision analytic modeling has been proposed to improve the RCT design by looking at what the key variables will be and the likelihood of making sound conclusions given potential variations in these variables. Jenniskens et al report on a case study of this where a decision analytic model was developed to assess the impact of using the HEART score for predicting major adverse cardiac events. Impact on patient health outcomes and health care costs was assessed in scenarios by varying compliance with and additional clinical knowledge. Although the subsequent trial did not exactly match the model, it did predict the challenges that occurred in the subsequent trial, especially the impact of poor compliance with acting on the HEART score results.

Another paper on prognostic prediction rules by Tan et al, reviews 91 clinical prognostic prediction models for obstetric care. As they point out, many countries have established comprehensive antenatal care systems, which

https://doi.org/10.1016/j.jclinepi.2019.09.020
0895-4356

document a full spectrum of health care data during regular antenatal visits and examinations. This provides great opportunities for developing prognostic prediction models. However, this potential is not being realised due to poor methods being employed. They found major problems such as lack of prespecified protocols, absence of validation, inappropriate measurement of outcomes and predictors, unclear intended timing of model use, and insufficient sample sizes. They call for the implementation of best practice methods for prognostic prediction models for obstetric care.

Are systematic reviews made more robust by contacting the authors of primary studies? The robustness of systematic reviews is highly dependent on the quality of the primary studies addressing the topic of interest. A major challenge for systematic reviewers is thus obtaining the required information to assess the quality of the primary studies. Despite the increasing use of reporting guidelines, many peer reviewed papers reporting the primary study fail to document all the needed key components needed to demonstrate low risk of bias of the studies. Given that many policy makers and practitioners depend upon the robustness of the best synthesis of the evidence of such systematic reviews, organisations such as Cochrane encourage systematic review authors to contact the investigators to obtain the missing information. Reynders et al surveyed the systematic review authors of the 319 Cochrane new systematic reviews published in 2016 and found that the systematic review authors did contact the authors of primary studies in these review for 74% of these reviews.They note a number of different considerations both for and against contacting the authors of the primary studies. Positive considerations include the finding that 10-50% of responses resulted in one of more of the following modifications of crucial outcomes: risk of bias scores, the GRADE scores, the summary primary or secondary outcomes of the review, and the summary effect size of the primary outcome of the review. However, on the negative side this slows the review process. Contact addresses are often out of date, especially for older studies with the resulting risk of bias of getting more information for more recent trials, resulting in better ratings for bias; less than half of those with correct addresses reply; and many systematic reviewers consider this post-hoc information to be unreliable given that authors sometimes contradict details of the published paper. These authors also noted that documentation of the process of solicitation and how judgements are made was poor, so they recommend a format for better documentation to improve transparency. In this context a controlled trial by Danko et al, showed that the additional effort of personalising the request such as by making a phone call to an author of the primary research does improve response rates, providing useful information for the systematic review.

Three papers address new developments in outcomes research. Standardised ultra-brief patient reported outcomes instruments are attractive for clinical practice and for research, especially registries. However, it is important to demonstrate that such ultra brief instruments have satisfactory clinimetric properties such as feasibility and responsiveness. One such instrument is the SPADE questionnaire which consists of one numeric rating scale for each of the domains of sleep, pain, anxiety, depression, and low energy/fatigue. Kroenke et al report on its satisfactory correlations with the relevant PROMIS 4 items/domain and other legacy measures[i.e., measures that have been widely used in the past] at baseline and at 3 months.

Meta-research to look at outcomes is becoming much easier with the advent of public databases such as Clinicaltrials.gov and the World Health Organization-International Clinical Trials Registry Platform (WHO-ICTRP). Kirkham et al provide a good example of this in their study to assess what types of trials use core outcome sets (COS). Interestingly, industry funded studies in 341 studies from the field of rheumatology were more likely [80% vs. 60%] to use core outcome sets. The authors recommend that more attention be paid to persuade public funding agencies of the importance of core outcome sets to avoid selective reporting and facilitate systematic reviews.

A further article on core outcome sets, by Young et al., looks at the issue of how unique outcome domains are defined in COS before ranking of them to decide which ones should be required [i.e., core] in all trials. A systematic literature review of the outcomes reported in past trials is one of the first steps [plus best practice is to complement by asking the views of patients with the condition]. This almost always results in many outcomes, a substantial number of which measure the same attribute or domain, so these need integrating and de-duplicating before being voted on for importance. Young et al. assessed a sample from the COMET database of all studies relevant to the development of core outcome sets for use in clinical trials [http://www.comet-initiative.org/resources] and showed the procedures followed are rarely reported, and when documented the process varies greatly. They call for a group to develop consensus recommendations on best practices for this stage.

Li et al report on a nice study using a controlled trial design to fill a critical methodological gap in current understanding of data abstraction best practices for systematic reviews. The study was an online randomized cross-over trial with 26 pairs of data abstractors. Each pair abstracted data from six articles, two per approach. The investigators compared three approaches, one of which used a Data Abstraction Assistant software tool. Disappointingly this did not show better results than regular approaches without the software. The error rate of approximatley 15% across all groups was similar to previous studies, but the results do support the need for independent abstraction by two assessors, especially for outcomes and results. The authors propose a useful practical guide to best practice given the current state of the evidence.

Missingness is a term that has appeared this century, defined as "*When there are missing data in a dataset, the reasons for it being missing may be explained by different covariates.*" https://en.wiktionary.org/w/index.php?title=missingness&action=history. To complement the GRADE (Grading of Recommendations Assessment, Development and Evaluation) paper on guidance for assessing risk of bias associated with missing data in trials being included in systematic review evidence [1], Kahale et al expand their guidance on definitions of key terms, categorising participants described in RCT reports and who might have missing data, and providing a flowchart on how to judge the outcome data missingness for each category.

It is reassuring that Tsujimoto et al report that, in contrast to primary research papers, evidence from their study on the time−lag of non-Cochrane reviews in the PROSPERO database agrees with previous evidence on Cochrane systematic reviews that the lack of statistical significance in the results is not a factor in delays if getting these systematic reviews published. Factors that investigators perceive as causing delays in publishing systematic reviews are mainly lack of time, rejection by journals, and duplication of similar topics.

Psychology journals publish a sizable number of policy and practice-relevant systematic reviews with meta-analyses, so it is disappointing that most do not meet the reporting best practice listed in the PRISMA Reporting Guidelines. Le Clercq et al found that only 20% of reviews indicate that the PRISMA reporting guidelines have been used, and that the compliance was substantially poorer in those that did not cite PRISMA for key elements such as: adequate reporting summary, protocol, information sources, search strategy, study characteristics, results of individual studies, funding, study selection, risk of bias in individual studies, and bias across studies. This suggests that psychology journal editors can improve the reporting of these factors by insisting that appropriate reporting guidelines such as PRISMA are used in their papers.

Comparative effectiveness systematic reviews based on network meta-analysis are needed for the majority of clinical and policy situations where more than one therapeutic option is being considered. However, this makes it much more difficult to summarise the results in a format that non-experts can easily understand. Ypes−Nunes et al present the format of the Summary of Findings table that the Cochrane Collaboration studies use. This includes details of the clinical question (PICO-Population, Intervention, Comparison Group, Outcome), a plot depicting network geometry, relative and absolute effect estimates, certainty of evidence, ranking of treatments, and interpretation of findings.

Scoping rapid reviews have an important place in surveying the literature landscape before time is spent on a detailed review. JCE has published on the methods of rapid quantitative scoping reviews [*Journal of Clinical Epidemiology* 67 (2014) 1291−1294] − now this is complemented by the article in this issue by Campbell et al. reviewing the methods used for rapid qualitative systematic reviews. Only 15 examples were found; the shortcuts used were similar to those used on the rapid quantitative scoping reviews such as abbreviated search strategies, including date and language restrictions, the use of a single reviewer for screening, and data extraction.

Dose-response is one the Bradford Hill criteria for assessing causation. However assessing this relationship is not straightforward nor, as Jiang et al show in their review of 93 meta-analyses of dose-response studies published in 2017, is there consensus on which metrics to report. For example, criteria are needed to report on the doses for the open-ended intervals of exposure or intervention doses, as well as how to report nonlinear dose-response relationships.

Reference

[1] Guyatt GH, Ebrahim S, Alonso-Coello P, Johnston BC, Mathioudakis AG, Briel M, et al. GRADE guidelines 17: assessing the risk of bias associated with missing participant outcome data in a body of evidence. J Clin Epidemiol 2017;87:14−22.