ELSEVIER

Journal of Clinical Epidemiology 107 (2019) vi−viii

Journal of Clinical Epidemiology

EDITORIAL

Stepped wedge designs are coming of age in clinical epidemiology

An important step toward the coming of age of the stepped wedge design has been the publication of a Stepped Wedge Cluster Randomised Trial extension to CONSORT [1]. This design has definite ethical and logistic advantages so needs to be in the research methods armamentarium. However for this design to be accepted by those making practice and policy decisions it is important that they be reassured that this design has been conducted in a rigorous manner. The development of consensus reporting guidelines is the first step but such guidelines then need to be 'road-tested' to ensure their feasibility and accuracy. Hemming et al assessed the quality of reporting in stepped wedge studies according to the 26 items in the new guideline using the participants of a workshop attended by 50 individuals familiar with this design. Most items performed well but better descriptions are frequently needed when reporting on the exact format of the design with justification of how clusters and individuals were identified for inclusion in the study, and whether this was done before or after randomisation of the clusters.

Although stepped wedge designs are often chosen due to ethical concerns, Eichner et al warn that stepped-wedge cluster randomized trials often did not reach their planned sample size. In a literature search they identified forty-six individual stepped-wedge studies from 2010 to 2017 and found that 30% of these studies failed to reach their planned sample size. The most common reasons were dropout of clusters and delayed implementation of the intervention.

Applicability of RCTs to clinical practice is addressed in two papers. Pawson challenges the current thinking on the extent to which pragmatic trials, even those that fully satisfy criteria such as PRECIS [2], are truly applicable to clinical practice. He claims the pragmatic trial literature has not been well informed by these discussions and fails to distinguish between the three categories that he favors: simple generalization, extensional generalization, and applicability. He proposes that generalization cannot be achieved in a single trial be it explanatory or pragmatic. It requires a multicase, multimethod approach building an understanding of the biological and behavioral mechanisms of action that result in heterogeneous treatment outcomes.

The second article on applicability comes from Malmivaara who concluded that many of 161 randomized controlled trials in four leading general medical journals (BMJ, JAMA, Lancet, and NEJM) published in 2017 provided insufficient information to be applicable to clinical practice. This is based on a benchmarking scale developed by this author that assesses the five categories of patient selection, baseline characteristics of patients, interventions and co-interventions, outcomes, and statistical issues. This requires more detail than is included in other clinical trial reporting guidelines for (i) patient selection; (ii) study setting; (iii) patient characteristics related to functioning, comorbidities; (iv) environment; (v) equity. One or more of these were deficient in almost all of the randomized controlled trials. For this benchmarking scale to be adopted consensus should be sought by not only journals but also clinicians and decision makers using a process such as that recommended for reporting guidelines by the Equator Group [3].

Four articles address statistical issues: Twiske and De Vente look at how to distinguish between and within subject relationships, when studying associations where one of the variables is time−variant e.g., to look at the longitudinal relationship between cholesterol and skinfold thickness, as an indicator for body fatness; the latter is a time dependent variable. The regression coefficient of a standard mixed model analysis is a weighted average of the between and within-subject part of the relationship. However one wants to know the relative contribution of each − Twiske and De Vente show how the two components can be usefully separated by a hybrid model where the between-subject part of the relationship is obtained using the individual mean value over time, whereas the within-subject part is obtained using the deviation score, that is, the differences between the observations and the individual mean value.

Speich et al address a major criticism of the RCT bedrock principle that only 'Intention to Treat' analyses are valid. However in making practice or policy treatment health decisions the size of the predicted benefit and harms, the intention to treat analysis [although it carries the least risk of bias), is not enough. Also essential for policy and practice decisions are the results of two other analyses a) The benefits and harms of the 'Per Protocol' [patients followed as in the protocol irrespective if they switched groups e.g., for worsening disease] and b) 'As Treated' (only patients that did receive the full 'dose' of therapy). A major difference compared to that of most current practice is the need for modeling in addition to the subgroup analyses that is required to provide the 'estimands' (modified estimate) for those classified as Per Protocol and As Treated. The authors demonstrated this when evaluating 12 RCTs reporting 138 analyses for 24 clinical trials; they found that

https://doi.org/10.1016/j.jclinepi.2019.02.013
0895-4356/© 2019 Published by Elsevier Inc.

structural modeling outperformed traditional statistical approaches since it better handled selection bias and confounders that change over time, that is, time-varying confounders, such as patient characteristics (e.g., body weight) or even the treatment that the study aims to explore.

Saad et al describe a new way for meta-analyses to ‘embrace patient heterogeneity’ to identify interventions that show efficacy in only a subset of treated patients [4], by analysing the distribution of an intervention-effect in random-effects models. Using elegant superiority/inferiority plots two examples are shown: one illustrating that evidence from a meta-analysis did not support authors’ highly publicized conclusion that hypericum is as effective as other antidepressants; the other of a subgroup analysis of the effect of ribavirin in hepatitis C, demonstrating clear important benefit in one subgroup but not in others.

Dose-response meta-analyses may not be well known to readers: they are defined very simply as meta-analyses that systematically synthesize dose-specific findings from multiple studies to yield more precise estimates of putative dose-response effects [5]. In addition to meeting the best methods criteria as laid out in AMSTAR, Xu et al propose five specific statistical considerations when deciding on the appropriate analysis for dose-response meta-analyses. These additional criteria were met in fewer than 30% of 529 dose-response meta-analyses published since 2011. Before these proposals are formalised they need to be endorsed by a process such as that recommended for reporting guidelines by the Equator Group [3].

Stone et al take issue with the usual strategy of assessing the impact of quality in meta-analysis by excluding lower or including higher quality studies, which they argue induces collider stratification bias- a form of selection bias. They demonstrated that this explains the documented inconsistency of effects using quantitative measures of plot asymmetry to reanalyze 17 general surgery clinical trials.

Two articles address reporting guidelines: reporting guidelines are endorsed by many journals but to varying degrees. Sharp et al looked at the frequency of any endorsement of the STROBE (Strengthening Reporting of Observational Studies in Epidemiology) reporting guidelines for observational studies [6], a widely accepted reporting guideline. Of 257 unique journals who publish observational studies, 54% did not mention STROBE. 5% required STROBE on submission, 9% suggested use, 5% recommended a ‘‘relevant guideline,’’ 28% mentioned it indirectly (via editorial policies or International Committee of Medical Journal Editors recommendations) and the relevant extension was required by 2 (!1%) journals.

CONSORT reporting guidelines of surgical treatment as a subset of non-pharmacological treatments date back to 2008 with an update in 2017 [7]. Conroy et al reviewed a wide variety of trials of surgery, both by surgical discipline and by geographic location, published within a cohort of leading surgical and medical journals. There was poor adherence to established reporting guidelines, especially the reporting of items specifically relating to surgical expertise and center clustering are recommended. Solutions are proposed about when and how to address differing surgical expertise and centre-clustering in the design, conduct and analysis of randomised surgical trials.

Clark et al report on an interesting approach to the priority setting for the Cochrane Library. The Cochrane organisation (formerly the Cochrane Collaboration), despite containing over 5,000 reviews of treatments, is unable to cover anywhere near all available pharmacologic, surgical and natural treatments. Cochrane Review Groups must therefore prioritize reviews to best allocate their limited resources. A variety of different approaches are used by the over 50 review groups. In this issue of JCE Clark et al report on one approach that uses the most frequently asked questions used for searches of TRIP database (originally ‘‘Trip’’ stood for Turning Research Into Practice but now only the acronym TRIP is used) that was established in 1977. From 2010 to 2017 30,541 acute respiratory infection searches in the TRIP database addressed 20 clinical treatment questions. These are all now addressed by Cochrane reviews or protocols.

Endocrine-Disrupting Chemicals (EDCs: In the first of a number of articles in press in JCE on the methodological challenges in evaluating the impact on humans of endocrine disruptive chemicals, Lee and Jacob review the challenges of the unpredictable net effects of diverse EDC mixtures, low reliability of exposure assessment, nonmonotonic dose-response relationships, nonexistence of an unexposed group, and complicated interactions with diet and obesity.

Finally, Oreel et al describe a new perspective on assessing the relationship between the different components of quality of life. They studied how a network analysis of sequential data on self reported symptoms, mood and physical state has the potential to delineate the sequence of different aspects of health related quality of life. They analyzed 30 patients with stable coronary artery disease; the patients completed a set of items assessing health−related quality of life on a computer nine times a day for seven consecutive days. Patients differed in which items preceded or followed each other. The authors suggest that to the extent that network models are meaningful representations of health−related quality of life dynamics, they may help deepening our insight and provide targets for personalized treatment.

Peter Tugwell
J. André Knottnerus
E-mail address: ltugwell@uottawa.ca (P. Tugwell)

References

[1] Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. The CONSORT extension for stepped-wedge cluster randomised trials. BMJ 2018;363:k1614.

[2] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe K, Zwarenstein M, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

[3] Available at. https://www.equator-network.org/toolkits/developing-a-reporting-guideline/.

[4] Embracing patient heterogeneity. Nat Med 2014;20:689.

[5] Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology 1993;4:218−28.

[6] Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiology 2007;18:805−35.

[7] Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials for nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med 2017;167:40−7.