Journal of Clinical Epidemiology 105 (2019) v–vii

Journal of Clinical Epidemiology

EDITORIAL

Multimorbidity and Comorbidity are now separate MESH headings

Nicholson et al. in a commentary point out that confusion continues within clinical epidemiology articles and elsewhere as to how to reliably distinguish between 'comorbidity' and 'multimorbidity'. Although differentiated in a paper by van der Akker (AK was an author) [1], only now in January 2018 has MeSH designated definitions and a different classification term for multimorbidity, distinct from comorbidity. Nicholson et al. re-emphasize that this is more than a semantic difference. While both terms focus on the occurrence of multiple chronic conditions within the same individual, the term ''comorbidity'' refers to the combined effects of additional conditions in reference to an index chronic condition (such as comorbidity in diabetes mellitus, stroke, depression or cancer). In comparison, the term ''multimorbidity'' indicates that no single condition holds priority over any of the co-occurring conditions from the perspective of the patient and the health care professional; risk factors are excluded from the above definition. Clinical guidelines too often ignore other morbidities beyond the index condition -and indeed multimorbidity may need a different approach that does not give primacy to the index condition but rather considers the patient preferences across all the multiple morbidities. These guidelines will need to be based on sound systematic reviews—the methods and guidance for these have yet to be developed by organisations such as Cochrane.

In a separate paper, Quinzler et al. describe a new multimorbidity index, the 'medication-based chronic disease score [medCDS) in independent patients outside of institutions that they propose for use in estimating disease burden and to facilitate tailoring of treatment. This score is based on the most common chronic conditions seen in primary care in Germany; for each chronic condition evidence based guidelines were identified and the suggested drug treatments were listed. A score was then derived from the number of drugs taken for each of the designated conditions and the total compared with mortality. In the three cohorts tested this produced better receiver-operating-curves for predicting mortality than existing medications-based and diagnosis-based scores. In addition to the need to explore how this is affected by under and over prescribing, this needs extending to assessing morbidity and quality of life.

A second Commentary focusses on issues around the external validation in a new population that is required of new predictive models of prognosis and diagnosis to demonstrate adequate receiver operating curve results. Pajouheshnia et al. examine reasons for the external validation showing differences from the original cohort -both the usual decrease as well as circumstances where the area under the curve actually increases. Using the example of a diagnostic prediction model using D-Dimer data of patients suspected of having deep venous thrombosis, they emphasize that differences in the distribution of predictor values across different studies need to consider both: a) true variation in the characteristics of patients (severity, clinical settings, geographical locations, or time); and b) also the ways that the predictors themselves (e.g., the D-Dimer) were measured.

Three articles address issues arising from the increased opportunities afforded by the internet. Comparisons of subsequent RCTs with observational web-based routinely collected health data have frequently found higher effect estimates in the latter so there is now increasing attention to exploring the use of Registry-based RCTs (RRCTs) to address the risk of selection bias. These are RCTs using the data or infrastructure of a pre-existing registry; these have the potential to combine the advantages of randomization (balance of unmeasured confounding) with the assumed advantages of registries (higher external validity) and enable long- term follow-ups of large sample sizes for relatively little effort. However, there are other concerns about their validity including the quality of the data and lack of blinding of the analysts. It is therefore reassuring that Mathes et al. found no differences in effect estimates in 15 mortality and 14 morbidity outcomes in cardiovascular and cancer comparable Registry—based RCTs and traditional RCTs published in 2016. However, more attention to the uniformity of outcome definitions, quality and validity checking is needed in registries.

Online internet trials offer an important format to empower the public and patients to contribute to rigorous evaluation of interventions such as self-management. However, they are susceptible to not only the usual biases seen in all trials but have additional risks that need to be systematically addressed. Price et al. reviewed a sample of 41 online trials published—the quality of these were similar to traditional trials but the authors identified a number of additional aspects and/or tailoring of existing reporting guidelines that should be followed—e.g. trial registration, reporting of methods, data management plans, and public and patient involvement. They propose the development and

https://doi.org/10.1016/j.jclinepi.2018.11.019

implementation of an online reusable protocol where reporting requirements would be embedded in the protocol to assist authors in writing up the online trials research.

Web based surveys are attractive as a means of contacting large numbers of respondents; they are efficient, inexpensive, help speed up post-data procedures such as data cleansing with built-in plausibility checks and lead to quicker provision of data for analysis but response rates are rarely high enough to address selection bias concerns. Weigi et al. provide a good example of this. For a colorectal cancer prevention study they invited 160,000 randomly selected persons aged 40-54 years in three large German cities from 2015 to 16 to complete a 15 minutes on-line questionnaire. The response rate although high in absolute numbers (>28,000) was low proportionately even after 2 reminders (19.6%). As the authors point out the possibility of selection bias even though the proportion of rick factors did not change after each reminder, does not meet standard epidemiologic criteria of over70% response rates [2] needed to guard against selection bias. So more safeguards are needed before such studies are used for policy making.

Two articles address issues of non-randomized Studies. Non randomised studies are beginning to be accepted for demonstrating comparative clinical efficacy by approval agencies, perhaps in part in response to the real world evidence (RWE) movement (and the large recent investments in this by pharma) defined by FDA as the ‘‘the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of real world data from observational or longitudinal patient studies’’. This has been hastened by the growing prevalence of electronic health records, combined with ‘big data;’ analytical techniques. So, the paper by Anderson et al. looking at NICE appraisals of pharmaceuticals between 2000 and 2016 is timely. They found 22 instances where comparative clinical effectiveness estimates were calculated using non-RCT data with 86% being approved (which is similar to rate of approval for submissions based on RCTs). They found that evidence base was of concern in that there was no uniform approach to the scientific assessment of the robustness of the evidence. They call for a) agreement on clear guidance to establish the comparative clinical effectiveness of pharmaceuticals with non-RCT data; b) monitoring and follow-up to assess the results of the real-world comparative clinical and cost-effectiveness of pharmaceutical technologies recommended on the basis of non-RCT data.

Turning to cohort studies, Binder et al. provide evidence of the widespread occurrence in top journals of likely bias in cohort studies due to the bias that Binder described in this journal in 2014 [3] - For example this occurs when the patient dies before the end of the cohort time period due to the risk factor of interest but the patient is then censored in the analysis. 58 of 125 eligible studies were at risk of this bias, but only 6 addressed this with appropriate sensitivity analyses. The authors make recommendations for better reporting and analysis.

Two papers describe provide exemplars of different questionnaire design methods. Patient-Reported-Outcome questionnaires are ‘coming-of- age’ in being accepted over indicators developed by experts alone. Van Melle et al. provide an exemplar of this being applied to the assessment of transitional patient safety (defined as any unintended or unexpected incident which could have or did lead to harm for one or more patients at transitions between general practice and hospital care settings). It has been shown that patients report different safety issues than health-care professionals, so excluding patients will lead to a biased view on transitional patient safety. The authors reviewed the literature and then worked with a patient panel to develop questionnaire items based on literature review, tested in the target group using a think-aloud procedure, and validated by a cross-sectional study among patients receiving health care at the interface between general practice and hospital care in two regions in the Netherlands. The resulting questionnaire achieved satisfactory reliability and validity in each of the 4 areas of: (1) personal relation with general practitioner, (2) personal relation with hospital physician, (3) information exchange, and (4) treatment consistency.

Flexibility of work functioning is a recent concept that has emerged in the 21st century yet Abna et al claim that the only measures of work productivity that includes this is a long 27 item questionnaire that is infeasible to use for screening. Abma et al. provide a good example of how a combination of indicators from classical test theory and item response theory methods were used along with evaluation of translatability and conceptual considerations to identify 5 and 10 item short version candidate items to reduce the responder burden whilst retaining sufficient discrimination.

Four articles report on systematic reviews and meta-analyses: all systematic reviews of interventions contain a narrative component even when a meta-analysis has been conducted. Further over 20% of Cochrane reviews of quantitative data in the effectiveness of interventions do not conduct a statistical meta-analysis (most commonly in reviews in public health, and effective practice and organisation of care) and report their findings and base their conclusions entirely on a narrative synthesis; this is likely even more common across all systematic reviews published in the peer review literature. Yet despite a number of guidance publications in the literature Campbell et al. in a review of nearly 500 systematic reviews, report that that there is an unacceptable degree of variation and indeed bias and ‘spin’ in the wording used in systematic reviews. They call for substantial improvements in clarity of reporting of narrative synthesis are required, and guidance is needed to improve clear and concise reporting guideline for narrative synthesis. This is critical if end-users including the public, practitioners and policy decision-makers are to have confidence in the results of systematic reviews that use narrative

synthesis. A second article on systematic reviews looks at the PRISMA reporting guidelines. Li et al. studied 60 Cochrane Systematic reviews and 60 non-Cochrane SRs in 2015. The uptake of the PRISMA reporting guideline for assessing benefit on the interventions (that was first published as the QUORUM Statement in 1996 and the PRISMA update in 2009) has been remarkably successful with over 80% adherence to the guidelines for 24 of the 29 PRISMA items to assess benefit-yet only one item (conclusions) of the 22 PRISMA Harms checklist items achieved over 80% adherence. Although to some extent unsurprising given that the PRISMA Harms checklist had not been published, but makes the point of how necessary such a checklist is. We look forward to seeing how the uptake improves now that this checklist in published and is prominent on the Equator reporting guidelines website.

Network meta-analyses also raise a number of new quantitative challenges−one such is how to best handle missing participant outcome data. Spineli makes a case for modeling, rather than excluding or imputing missing outcome data, to provide bias-adjusted results; this is demonstrated by showing the difference in results obtained when applied to a published network meta-analysis of seven pharmacologic treatments for chronic obstructive pulmonary disease. It will be interesting to see if this holds true in a representative sample of network meta-analyses. Another quantitative challenge: network meta-analysis assumes data 'coherence' (consistent internally, over time and across products and programmes) and 'transitivity' (if the relation holds between a first element and a second and between the second element and a third, it holds between the first and third elements equality is a transitive relationship). The GRADE guidelines group have identified one situation where these principles do not hold - where there is sparse data in the network. Brignardello-Petersen et al. demonstrate this in the example of a network meta-analysis of 8 trials including nearly 3500 patients; the direct risk ratio and the indirect were similar with overlapping confidence intervals−although these estimates were coherent, when they were combined the confidence intervals widened instead of narrowing as would be logical. This is implausible and would lead to wrong policy and practice decision-making. The authors demonstrate how a number of different statistical methods can be used to address this.

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

References

[1] Van den Akker M, Buntinx B, Knottnerus JA. Comorbidity or multimorbidity, what's in a name? A review of literature. Eur J Gen Pract 1996;2:65−70.
[2] Munday D. Solutions Corner. Science Editor. January − February 2002. Vol 25 No 1. 25-26
[3] Binder N, Schumacher M. Missing information caused by death leads to bias in relative risk estimates. J Clin Epidemiol 2014;67:1111−20.