EDITORIAL

When is a health-care quality indicator ready to use?

Quality indicators and how rigorous they need to be for internal use and external reporting are debated in two interesting letters in this issue. Stelfox et al. focused on quality improvement, accountability, and research as the potential primary goals of quality; Doggen et al. usefully expand this to what they term external uses that include accountability, selective referral (choosing a doctor or hospital), or increasing health knowledge. Both groups agree that judgment is needed but disagree on when an indicator is ready for use using a framework such as that of the Institute of Medicine (IOM). These frameworks are conceptually attractive: IOM defines five broad criteria for good indicators: importance, scientific soundness, feasibility, alignment, and comprehensiveness (safety, effectiveness, patient centeredness, timeliness, efficiency, and equity). The clinical epidemiology community should engage more actively in establishing criteria for the clinical sensibility of the measures used; for example, by developing consensus on the threshold of scientific evidence for adopting quality indicators to optimize value.

Another field which the Journal of Clinical Epidemiology has published articles on but needs more attention in clinical epidemiology research, is frailty [1,2]. Frailty (defined as decline in physiologic reserves and resilience) is a concept that is achieving global acceptance as an identifiable health state contributing to the global burden of morbidity and is of increasing importance with the increased life expectancy. One of the most widely used sets of measures is the phenotypic model introduced by Fried et al. [3] that uses five indicators of various phases of the frailty cycle: unintentional weight loss, exhaustion, muscle weakness, slow gait, and low levels of physical activity [4]. Bellinda et al. assess the trustworthiness of the results comparing frailty between 12 countries in Europe (Denmark, Germany, Sweden, The Netherlands, Spain, Italy, France, Austria, Greece, Switzerland, Belgium, and Israel) in a Survey of Health, Ageing and Retirement in Europe (n = 27,938), in a population-based study of community-dwelling adults aged >50 years between 2004 and 2006. The authors note that a simple frailty sum score of the above five indicators using a single question for each is useful in the clinic, but when used across countries, it is important to check and correct for factor invariance. For seven countries, there was no bias detected, but for the other five countries, bias was detected. Robustness and stability of these measures are needed as a baseline and to monitor individual and community approaches to preventing and managing frailty. This constituency of researchers in the area of “frailty” should link up with those working on the clinical epidemiology challenges of “comorbidity” which again means breaking out of the single disease specialty concept.

How best to report absolute results for survival curves? Best practice in clinical epidemiology now calls for both relative and absolute measures of treatment effect when presenting the results of interventions. When the follow-up is long term, median or mean comparisons should be enriched by showing survival curves to show how early any differences appear and whether and/or when these differences are sustained. The default method is the Kaplan—Meier curve. To address concerns of right censoring and evolving selection bias over time, Coory et al. make a case with examples for the need to complement them with the “risk difference curve”.

Should hospital inpatient mortality rates be superseded by a “hospital-patient 1-year mortality risk score”? van Walraven tackles the challenge that the health services benchmark of in-hospital mortality rates are increasingly meaningless as we discharge patients earlier and earlier. He suggests such an index is important not only for comparing health-care performance between communities or hospitals but also for allowing patients and their physicians to make more informed decisions about their health care during the hospitalization and afterward. He reports here that the risk of death within 1 year of admission to hospital can be accurately estimated by a risk index (the hospital-patient one-year mortality risk score) that quantifies the influence of a dozen patient and hospital factors on long-term survival.

A field of growing importance is text messaging in clinical epidemiologic research [4]. Social media is changing the way many of us communicate with others, and as Brabyn et al. note, it has been used already in clinical care for appointment reminders and chronic disease management, so it makes sense to also use this for clinical research. This study uses short message service text messaging very successfully (>90% response rate) in obtaining daily texts for appointment reminders and chronic disease management, so it makes sense to also use this for clinical research. This study uses short message service text messaging very successfully (>90% response rate) in obtaining daily texts as a tool for data collection in a trial of treatment for irritable bowel syndrome by texting a single number in answer to a text asking for a number out of nine on the severity of their symptoms. This immediacy has the great appeal of obliterating recall bias in data collection but clearly is limited by how much can be asked at one time.
Can one trust proxy measures of quality of life and well-being for assessing the burden of illness in long-term care institutions such as nursing homes? Devine et al. studied this for 565 pairs of care home residents greater than 85 years of age and proxies by administering the EQ-5D to both independently and comparing the results. The agreement is good for those with depression and, not surprisingly, those with lower cognitive impairment. Poor levels of agreement were found for single domains but achieved a minimal level of agreement for the whole index score. This level may be acceptable for assessing burden of illness at a macro level but is almost certainly insufficient for assessing change at the group and/or programmatic level and certainly for assessing individual interventions.

How important is authorship order in journal publications? For clinical epidemiology researchers, this is indeed important because it is increasingly rare to see single-authored articles, and decisions on promotion, tenure, and annual merit increases for individuals [and for institutional funding] are based on journal publications as a major criterion. Bhandari et al. surveyed chairpersons of departments of medicine and surgery in North America. They confirmed its importance, but there was considerable variation in interpretation of the importance of the order itself, the corresponding author role, and that of the last author. This adds impetus to the universal adoption by all peer-review journals of the practice of formally listing each authors contributions that is advocated in a few leading journals.

Three articles address systematic review methods. Oral medicine and dentistry systematic reviews are coming of age so how good is their quality? Papageorgiou et al. analyzed the quality of 281 meta-analyses containing over 3,000 trials in eight databases up to 2013. Ten percent [28 of 281] of meta-analyses were not listed in Medline, so they recommend that other databases need searching for this oral health topic area. We look forward to there being criteria and data for some of the special design features in oral health not assessed here but noted by the authors, such as the unit of analysis (patient, mouth quadrant, jaw, tooth, or even tooth surface), surrogate end points, and blinding.

Is “Table 1, with the baseline comparisons” as important in systematic reviews as it is in randomized controlled trial publications? For primary data such as from randomized trials, we all appreciate that randomization does not guarantee equal allocation of known confounders so one needs to check for baseline differences. This concept needs taking a step further in meta-analyses: when there are small but statistically nonsignificant differences in individual studies, then if these are in the same direction overall, a substantive difference in potential confounders can occur in systematic reviews and meta-analyses. Clark et al. provide a nice example of this, finding that age was imbalanced overall in five of a sample of twelve recently published systematic reviews with 503 component randomized trials, published in leading journals in 2011–2012. These authors offer four techniques to assess this phenomenon and, if present, suggest using these results to drive sensitivity analyses.

How comprehensive should searches for new studies be for network meta-analyses (NMA) systematic reviews? NMA systematic reviews are increasingly popular, because they address comparative effectiveness questions of more than one intervention even when there are no direct head-to-head comparisons. As Lun Li et al. point out both traditional and NMA meta-analyses need to include all available evidence and apply a systematic and unbiased approach to estimate potential differences; thus it is important to ensure the same level of the best practice as is recommended for traditional systematic reviews and meta-analyses. Yet, in reviewing 248 NMA identified in the major search databases, the search standards recommended in the Cochrane Handbook were not met in >50% of these NMA publications.

Finally, an interesting article on animal research conduct collaboration on methods between those working with animals and with those studying humans should be encouraged for each side to learn from each other [5]. Bello et al. examine the impact of not blinding outcome assessors on estimates of intervention effects in animal experiments modeling human clinical conditions. Although blinded outcome assessment is recommended, it is not often employed in practice. In 10 experimental studies involving more than 2,000 animals, each contained blinded and unblinded assessments. There were dramatic differences between the blinded and unblinded assessments averaging a 59% larger effect in the unblinded studies.

Peter Tugwell
J. André Knottnerus
Editors
E-mail address: anneke.germeraad@maastrichtuniversity.nl
(J.A. Knottnerus)

References