The evidence base of taming continuously proliferating evidence

The annual number of research publications listed in Medline is exponentially increasing, and there is no reason to assume that this trend will not continue in the years to come. In addition, a growing number of online texts are becoming available on a wealth of topics relevant to clinicians. This is a blessing on the one hand, as it is both a result of and a major contribution to progress. But on the other hand, it is a matter of concern because it is increasingly difficult to overview the literature, not only with regard to content but also, and in particular, with regard to quality. ‘Taming’ published evidence, by using approaches that summarize the continuously growing stream of publications and help to assess and select with regard to relevance and quality, is increasingly important. At the same time we must be keen on demonstrating whether these approaches really work, by making approaches themselves subject to evidence-based evaluations. This is a common theme in a number of contributions in this issue.

Pottie and colleagues report on their experiences with integrating the GRADE approach into the guideline development by the new Canadian Task Force on Preventive Health Care [1]. They provide us a good insight into how the GRADE guidance for rating quality of evidence and grading strength of recommendations in health care is working out, and in what key methodological challenges may arise. Their experience is clearly positive, but they emphasize the need for better methods for efficiently seeking and selecting the best evidence when faced with lower quality evidence and modeling studies. In a commentary on this paper, Guyatt et al. address this topic, which is a challenge not unique to GRADE, in relation to population screening studies. They emphasize that a narrow or rigid interpretation of GRADE guidance should be avoided.

In summarizing and assessing accumulating evidence, reviews of systematic reviews are becoming increasingly frequent. Therefore, Pieper and coworkers have taken up the challenge to examine the descriptive and methodological characteristics of published overviews of systematic reviews. In their review of overviews of systematic reviews, based on 126 overviews of reviews, they come to the important conclusion that overviews of reviews often lack methodological rigor, and they make a plea for methodological and reporting standards for overviews.

An important question in evaluating the basics of evidence-based approaches is whether journals, in one way or another, select the articles that they publish in relation to study outcome. Based on a study of 67 meta-analyses of 964 randomized trials, Evangelos and his team found a correlation between information gain from a study and the impact factor of the journal in which the study was published. They conclude that publication in influential journals with high impact factors is apparently driven by the perceived information gain, and discuss the implications of their findings.

As online medical texts are increasingly important as evidence base for medical decision-making, Prorok et al. evaluated the quality of evidence reporting of 10 selected online medical texts. Using an 11-item scale, also addressing breadth of coverage and timeliness of content updating, they found that no single resource was ideal. The authors advise those seeking answers to clinical questions not to rely on one online resource, and also make recommendations for authors, editors, and publishers.

An important step forward as compared with meta-analysis based on published studies that has been advocated now for many years is to collect and meta-analyze individual patient data (IPD) [2]. In this context, the findings of Kovalchik are striking. She approached a large number of authors of meta-analyses in an e-mail survey on their efforts to obtain IPD, and found that most meta-analysts never attempt to collect IPD. The most frequently reported reason for not doing so is the belief that participant data cannot be obtained. Clearly, this is an urgent issue with much room for improvement for investigators, meta-analysts, and their institutions, given their responsibility for optimal use of data of patients and study participants.

Of course, in discussing the added value of evidence-based approaches, the debate on the appropriateness and validity of study designs is and will always be important and welcome. In the Variance and Dissent section, Kotz et al. and Mdege et al. discuss the stepped wedge design, connecting to a systematic review by Mdege and her group that published earlier in this journal [3]. The challenges and pitfalls associated with using the stepped wedge design, circumstances under which it may be used, and relevant precautions, are highlighted in this debate.

Other contributions are also addressing relevant questions on the strengths and limitations of presented evidence. Ertel et al. showed the importance of addressing conflicting results from previous studies from a methodological perspective, illustrated with the important question whether maternal perinatal depression and child overweight are
independently associated. Their findings, based on data from a large number of mother-child pairs in a population-based prospective study, add evidence that postpartum depression is not associated with child weight. The investigators emphasize the importance of methodological approaches, including adequate handling of missing data, to understand conflicting evidence.

In a review, Mills and colleagues study the clinically important but difficult question of how to combine the effects of treatments that have not been evaluated in combination. They provide necessary assumptions to be able to consider combining data from single or multiple trials to determine a combined effect, and conclude that additive effects may be useful tools to estimate the effectiveness of treatment combinations.

In a systematic review, van Oort and coworkers summarize the methodological quality and development stage of prediction models for primary care physiotherapy of musculoskeletal complaints. They found that, while prediction models are available for a wide range of patient populations, the development stage of most models is preliminary and the study quality is moderate. Additional work is needed before these models can be recommended for practice.

A number of articles address the evaluation of the performance of measures and instruments. In a cross-sectional study in 17 primary care clinics, Légare et al. assessed the psychometric properties of dyadic measures for shared decision-making. The validity and reliability of dyadic measures were confirmed for various elements. The authors discuss the importance of valid and reliable dyadic measures for shared decision-making research. For the large Chinese-speaking population and the clinical researchers working in Chinese-speaking communities, the study of Wei and his team provides useful data on the reliability and validity of the Chinese version of the Quebec Back Pain Disability Questionnaire (SC-QDS). Based on a study in patients with low back pain and controls, the authors conclude that the SC-QDS is appropriate for clinical and research use among Chinese-speaking patients. In a study population of nurses, Gärtnert et al. evaluated the interpretability of individual changes, minimal important change for improvement, and the smallest detectable change of the Nurses Work Functioning Questionnaire. Their results may guide researchers and practitioners in their conclusions on whether changes in individual workers are ‘real’ and ‘relevant.’ Further research needs are described. Morze et al. examined the test-retest reliability and validity of an instrument to capture self-reported melanoma risk factors in the context of a cohort study, and found a fair-to-good test-retest reliability for most self-reported risk factors. This result may be helpful for developing risk prediction tools in the future.

The study by Mitchell and coworkers adds to our insight into how to reduce attrition in randomized controlled trials. The authors assessed the effectiveness of pre-notification using a newsletter to increase response rates in the context of a randomized trial on a screening program to identify osteoporosis risk. The results support previous research suggesting that pre-notification increases response rates and the authors recommend sending out newsletters to trial participants before the follow-up questionnaires.

J. André Knottnerus
Peter Tugwell

E-mail address: kim.luyten@maastrichtuniversity.nl
(J.A. Knottnerus)

References