There are few therapeutic fields in which the need for an impartial, balanced, synthesis of the evidence is as great as for osteoarthritis (OA). The disorder is one of the most frequent and costly causes of physical disability in the population, yet no especially effective intervention or therapy has emerged. As a result, there is an astonishing range of promulgated therapies, many of which are marketed directly to consumers—often, apparently, in the face of compelling scientific evidence of their inefficacy. Even among experts, there are examples of a striking propensity for differing interpretations of available data. A robust and respected evidence synthesis would clearly be of value to assist health professionals and individuals with OA in making wise treatment choices and optimizing resource allocation.

Historically, the development of broadly acceptable treatment guidelines for OA has been a challenge. As Zhang et al. highlight in this most recent endeavor, previous iterations from different constituencies produced no universally agreed recommendations, even for a core group of safe and effective therapies. It is evident that the acceptability of guidelines is almost entirely contingent on the perceived robustness of their methodological origin. The sophistication of methodologies for achieving evidence-based consensus has advanced considerably, but so has their complexity and level of effort. Furthermore, the field has exhibited shifting paradigms and may not yet have solidified. For example, consensus on the optimal methodology shifted from opinion-based (with its potential for bias) to a more purist evidence-based synthesis of high quality evidence (which could be patchy and theoretical) to a more balanced hybrid approach that aims to integrate research evidence with clinical expertise. Though there are advantages to the hybrid approach, there are also potential pitfalls including potential for variability in its deployment. The 2003 EULAR task force, for example, solicited propositions from the expert panel as a first step, and next sought evidence on those propositions. This may have constrained the scope of therapies brought under consideration.

Thus, with a remit from OARSI, and evidently mindful of the decision to recommend. The 2007 OARSI recommendations for the management of hip and knee osteoarthritis: towards consensus?

T. McAlindon D.M., M.P.H., Chief of Rheumatology*, Natalie V. Zucker, Milton O. Zucker Professorship in Rheumatology and Immunology
Tufts Medical Center 800 Washington Street, Boston, MA 02111, United States

*Address correspondence and reprint requests to: Dr Tim McAlindon, Division of Rheumatology, Tufts-New England Medical Center, 800 Washington Street, Tufts Medical Center, Box 406, Boston, MA 02111, United States. Tel: 1-617-636-5645; Fax: 1-617-636-1542; E-mail: http://www.nemc.org/medicine/rheumatology, tmcalindon@tufts-nemc.org

© 2008 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2008.02.009
experts in pertinent disciplines, of Europe and North America, and provided illuminating information on Industry affiliations. Importantly, panelists that perceived themselves to have potential conflict of interest were able to (and did) recuse themselves from voting on individual propositions.

The Delphi process consisted of developing and refining a series of data-based propositions among the panel until a level of consensus was reached. In fact, the exercise appears to have been so effective in crafting carefully-worded text that, at times, its mellifluousness seems primarily intended to avoid contention rather than to provide specific directive. Nevertheless, the propositions are generally down-to-earth statements about commonly-used modalities about which some consensus will have great utility.

The most appealing output of this analysis, however, lies in the aim to quantify level and strength of consensus achieved for each proposition. This information should enable us to weigh, or rank, the level of conviction held by the panel for each recommendation. In accepting that treatment issues are rarely dichotomous, this approach has the additional advantage of removing the overly directive tone of many guidelines. However, there is some complexity in interpretation of their metrics. For example, the visual analog scale used to measure strength of recommendation does not have clearly defined anchor points for the extremes. So, it was unclear to me whether '0' represents equipoise or a view that the evidence suggests that the modality actually does not work. The first situation would present the possibility of bias due to a floor effect, while the second would push equipoise into the middle of the VAS and hinder interpretation of the numeric value. It also proved difficult at times to reconcile the scores for level and strength of consensus.

What does it mean, for example, that the proposition for acetaminophen (a pharmaceutical recommended in 100% of prior guidelines) received only 77% consensus yet a 92% strength of recommendation? Or, in contrast, that the proposition for joint lavage/arthroscopic debridement has consensus of 100% yet a strength of recommendation of 60%? Scrutiny of the phrasing of the proposition, or the detailed proposition in the manuscript, does provide clarification in many cases, since votes for strength of recommendation were for the propositions along with included caveats. Until we grow more familiar with the external benchmarks of these two parameters, it may be prudent to view them as ranking metrics for (1) the proposition as stated, and (2) the level of conviction of the usefulness of that modality in a clinical setting.

So how do the OARSI recommendations perform in a test run? I was jarred only by proposition 1, an articulation of OA folklore that does not contain a scientifically-testable hypothesis, appears to be based entirely on opinion, yet received among the highest strengths of recommendation. Propositions for controversial treatments (e.g., glucosamine and hyaluronate) or interventions with limited or negative data (e.g., joint lavage) were present but ranked appropriately low for strengths of recommendation. What did emerge was the omission of some prominent studies that were published subsequent to the Delphi exercise. This was probably unavoidable and, according to the sensitivity analysis, non-influential. However, this remains a more general dilemma if, as the authors assert, a consequence of their burden and complexity is that guideline development in fast-moving fields will never be quite current. Could this whole exercise be reconfigured to facilitate frequent updates, or even real-time adjustments, perhaps leveraging the power of the Internet? We should challenge the OARSI guidelines panel, at the completion of this considerable endeavor, to apply their evident creativity in the accomplishment of such an ideal.

References