Osteoarthritis and Cartilage (2008) **16,** 636–637

doi:10.1016/j.joca.2008.02.009

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Editorial
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

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 disablement in the population[1], yet no especially effective intervention or therapy has emerged. As a result, there is as 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[2]. Even among experts, there are examples of a striking propensity for differing interpretations of available data[3–5]. 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[6], 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 OA task force, for example, solicited propositions from the expert panel as a first step, and next sought evidence on those propositions[7]. This may have constrained the scope of therapies brought under consideration.

Thus, with a remit from OARSI, and evidently mindful of the methodological milieu, Zhang and the guideline development panel planned a more logical but burdensome approach that reversed these steps – evidence-driven *first* and clinically-supported *second* (6 and guideline Ref.[10]). However, the enormity of the task inherent in the first step (appraisal of the totality of evidence on OA treatments) seems to have obligated the OARSI panel to reduce their burden by performing instead an appraisal of existing 2003 EULAR guidelines, updated by a systematic review of subsequent evidence – itself not a trivial endeavor! This seems like a reasonable reductionist step to enable the feasibility of the global data synthesis but it might still leave vulnerability to omission of data on treatments not considered in prior guidelines. It might also make readers wonder about the level of detail that informed the experts' decisions. How, for example, would they weigh a flawed meta-analysis (viewed as the top level of evidence in the hierarchy) against high quality clinical trial results (which would have been omitted in the systematic review of recent evidence)? What would they do with pooled effect size estimates from high quality meta-analyses that doubted their own results because of problems with the included trials (a common situation apparently[8,9])?

So what emerged on the formidable journey through evidence gathering towards formulation of consensus? Quite a few illuminating pieces of information, as it turns out. Firstly, among the 23 guidelines that met inclusion criteria for critical appraisal, there was considerable variation in setting (e.g., rheumatology vs orthopedics) and also in what treatments were considered. There were significant differences in mean quality scores between guidelines subsets with opinion-based formulations scoring lowest. Among 51 treatment modalities collectively addressed, 20 were recommended by all, but the denominator varied (e.g., treatment with 'herbs' was addressed in only two guideline sets but recommended by both – '100% consensus'!). Somewhat paradoxically, the strength of consensus for a treatment did not necessarily reflect the level of evidence in any of the guideline sets (including this one). For example, ultrasound therapy received weak support, notwithstanding positive results in clinical trials, yet there was strong support for surgical procedures, which were supported only by uncontrolled or observational data. Clearly, a great number of unmeasured factors influence the decision to recommend.

The guideline development process also rests on the capacity of the expert panel to accurately synthesize the presented information and reach consensus. It is gratifying that the panel here was reasonably representative of

*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

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 mellifluence seems primarily intended to avoid contention rather than 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 approaches 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 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

1. Fenter TC, Naslund MJ, Shah MB, Eaddy MT, Black L. The cost of treating the 10 most prevalent diseases in men 50 years of age or older. Am J Manag Care 2006;12(Suppl 4):S90–8.
2. Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, *et al*. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. N Engl J Med 2006;354(8):795–808.
3. Greenwald RA. Marginal efficacy of glucosamine: comment on the article by VLAD, *et al* and the editorial by Reginster. Arthritis Rheum 2007;58(1):332.
4. Reginster JY. The efficacy of glucosamine sulfate in osteoarthritis: financial and nonfinancial conflict of interest. Arthritis Rheum 2007;56(7): 2105–10.
5. Vlad SC, LaValley MP, McAlindon TE, Felson DT. Glucosamine for pain in osteoarthritis: why do trial results differ? Arthritis Rheum 2007; 56(7):2267–77.
6. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15(9):981–1000.
7. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, *et al*. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003;62(12):1145–55.
8. Arrich J, Piribauer F, Mad P, Schmid D, Klaushofer K, Mullner M. Intra-articular hyaluronic acid for the treatment of osteoarthritis of the knee: systematic review and meta-analysis. CMAJ 2005;172(8):1039–43.
9. Lo GH, LaValley M, McAlindon T, Felson DT. Intra-articular hyaluronic acid in treatment of knee osteoarthritis: a meta-analysis. JAMA 2003;290(23):3115–21.
10. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16(2):137–62.