**Abstract:** The American Academy of Orthopaedic Surgeons (AAOS) 2013 guidelines for knee osteoarthritis recommended against the use of viscosupplementation for failing to meet the criterion of minimum clinically important improvement (MCII). However, the AAOS’s methodology contained numerous flaws in obtaining, displaying, and interpreting MCII-based results. The current state of research on MCII allows it to be used only as a supplementary instrument, not a basis for clinical decision making. The AAOS guidelines should reflect this consideration in their recommendations to avoid condemning potentially viable treatments in the context of limited available alternatives.

Value, quality, comparative effectiveness: these are terms we see and hear with increasing frequency in relation to the delivery of health care. They are used in connection with the term evidence-based medicine (EBM), which is a concept meant to promote the scientific proof of treatment efficacy throughout medicine and surgery. In the past, physicians were the arbiters of proof of efficacy through their analysis and interpretation of the medical literature. Orthopaedic surgeons reviewed their literature, mostly case series (Level IV) and case-control studies (Level III), and treatment recommendations were made to patients based on the analysis leavened with expert opinion and experience. The increasing reliance on third-party reimbursement of medical services and the resulting cost escalation associated with this payment method have forced the government and insurers to examine exactly what it is they are paying for when services are performed. One of the main purposes of EBM when performed by payers, therefore, is economic.

Evidence-based analyses of medical services are now commonplace and becoming widespread. These analyses are used in determining coverage decisions for medical services. The federal government has invested over $1 billion in the Patient-Centered Outcomes Research Institute, and insurers each perform their own evidence synthesis of the literature. Many states use these methodologies, especially when evaluating high-cost and innovative procedures. In addition, international EBM analyses such as those performed by the National Institute for Health and Clinical Excellence in England are cited in the United States when EBM analyses are performed. There is no standardization throughout these various programs in choosing the methodology used to determine the value of studies in the literature. In addition, expert input from orthopaedic surgeons is also variable. This lack of standardization and expert input has led to confusion about the value of much of the orthopaedic literature in EBM studies.

The American Academy of Orthopaedic Surgeons (AAOS) has developed its own quality initiative and EBM process including clinical practice guidelines (CPG) and appropriate use criteria. This quality initiative is meant to make sure that the real experts in musculoskeletal care are involved in the EBM process. The CPG process has been controversial because of its evidence-synthesis rules that stress randomized clinical trials (Level I) and do not allow inclusion of much of the Level III and IV evidence available in the literature. In addition, some statistical metrics such as minimum clinically important improvement (MCII) and minimum clinically important difference (MCID)
result in significant changes in the way even Level I studies are assessed for value. In the recently published CPG on osteoarthritis (OA) of the knee, the MCII was used as a metric in the strong recommendation against the use of viscosupplementation in treating knee OA, despite studies that showed statistically significant improvement of symptoms after treatment with certain types of hyaluronic acid.

Because one of the tenets of EBM is the incorporation of expert opinion, it is important for orthopaedic surgeons to understand the various methodologic rules and analytic metrics used in these analyses. Armed with this understanding, physicians can critically evaluate EBM analyses to provide important expert-opinion filters to such programs. Physicians can then advocate for or against EBM studies, providing the expert opinion that is essential to their proper use. In this way, access to safe and effective care will be maintained for patients.

**Commentary**

The most recent AAOS CPG for knee OA raised controversy by “strongly recommending” against the use of viscosupplementation. This negative shift from the 2008 “inconclusive” recommendations followed changes in CPG methodology rather than an evidence trend. Furthermore, the significant limitations of the AAOS methodology called into question the guidelines’ verdict and highlighted the profound impact that unfavorable recommendations may have on multiple stakeholders. We aim to discuss the methodologic flaws of the AAOS guidelines that led to recommendations against certain treatments, such as viscosupplementation, potentially depriving some patients of effective remedies for a disease with a very limited range of therapeutic options (Fig 1).

The guidelines included only 14 hyaluronic acid studies, in stark contrast with the 89 articles covered by the latest systematic review on viscosupplementation. To compare treatments, the AAOS 2013 guidelines used a relatively new metric termed the “minimum clinically important improvement”—that is, the “smallest clinical change that is important to patients.”

Although the AAOS’s meta-analyses showed statistically significant treatment effects of viscosupplementation, it was nevertheless deemed ineffective because it did not meet MCII thresholds. In theory, MCII is a helpful tool to show how statistically significant improvements may not always be clinically relevant. However, as intuitive as it appears, this instrument requires proper understanding and grasp of its limitations from those using these guidelines. Several caveats should be considered for the appropriate interpretation of MCII-based conclusions.

First, it is essential to critically appraise how the MCII was derived. The AAOS sourced the MCII/MCID values from Tubach et al. (2005) and Angst et al. (2001). Angst et al. used the metric of MCID, which assesses both improvement and deterioration, whereas Tubach et al. used MCII (which assesses only improvement). The studies evaluated cohorts of knee/hip OA patients undergoing different interventions: nonsteroidal anti-inflammatory drugs or inpatient rehabilitation. However, the AAOS did not consider that MCID or MCII values are context specific. For example, Angst et al. observed that MCID absolute values differed for improvement versus deterioration, whereas Tubach et al., as well as other authors before them, found that the main confounding factor in the determination of MCII/MCID was the baseline symptom severity. Besides regression-to-the-mean and “floor” and “ceiling” phenomena, this dependence of MCID/MCII on baseline values could be explained by the influence of patients’ expectations on their determination of treatment effect. Thus the MCII could depend on other factors shaping patients’ expectations, such as the types of treatments in a randomized trial. Current literature also associates MCII/MCID levels with time between outcome assessments and age. Consequently, MCIs calculated for specific therapies may not be applicable to other treatments or patients with different baseline characteristics. Furthermore, Tubach et al. anchored MCII to a “good” level of improvement in 75% of patients or more, surpassing a “fair” positive change (defined as “reasonable effect but could be better”). This may have set the criterion for the “smallest” clinically important improvement too high to be used as a standard for determining a treatment’s efficacy.

Second, one should be aware of how MCII values were presented to readers. For forest-plot presentation, the AAOS guidelines “standardized” MCIs by dividing them by the standard deviation of a “mean baseline score” and then displayed them as bold red lines. Because of the dependence of MCID/MCII on a study population’s heterogeneity, previous research has suggested dividing MCID/MCII by an agreed-on population-based standard deviation rather than standard deviations pooled from individual trials and then using the same measure throughout. This should lead to consistent MCII cutoff values for the same condition and outcome across all treatments, which was not the case in the AAOS guidelines. Moreover, no adjustment was made for the fact that effect sizes calculated from each trial (less precision) would be biased toward the null-effect, whereas MCII standardized by the pooled standard deviation (high precision) would be biased away from the null-effect. Hypothetically, featuring clinically and statistically significant cutoffs in the same plot may be helpful. However, presenting MCII as a red line more prominent than the null-effect line could have been misleading to the AAOS voting panel, especially given the lack of transparency in how the standardized MCII cutoffs were different for different treatments.

Finally, it is critical to correctly interpret the results of analyses performed using the MCII. It is debatable
whether MCIIs should be used at all for between-group comparisons when the cutoff values were derived from within-patient comparisons. Notably, this would disregard the placebo effect and still require the between-group difference to be as large as the within-patient improvement to achieve "clinical significance." Even if we accept the MCI for between-treatment comparison, the interpretation of results could be overly simplistic. Contrary to statistically significant cutoffs for effect sizes, MCIIs are derived from within-patient comparisons and may not be applicable for between-group comparisons. MCIIs are a distributional characteristic that should be expressed as a range, with proportions of patients in each group that achieved outcomes reaching the lower bound.

**MCII methodological points**
- MCIIs are context-specific and may not be applicable across treatments or patient populations.
- Standardized MCII values for the same condition and outcome should be uniform across all treatments.
- MCIIs are derived from within-patient comparisons and may not be applicable for between-group comparisons.
- MCIIs are a distributional characteristic that should be expressed as a range, with proportions of patients in each group that achieved outcomes reaching the lower bound.

**AAOS 2013 guideline**
- The AAOS evaluated various treatments with MCIIs obtained from 2 nonrandomized trials of everyday-practice NSAIDs and rehabilitation.
- The AAOS standardization procedure lacked transparency, and MCII cut-offs inexplicably differed across treatments.
- The AAOS used MCIIs for between-group comparisons while failing to account for the placebo effect.
- MCIIs were expressed as a single cut-off point, thereby dismissing a share of patients who have achieved clinically significant improvement.

**Fig 1. Summary of key points.**

whether MCIIs should be used at all for between-group comparisons when the cutoff values were derived from within-patient comparisons. Notably, this would disregard the placebo effect and still require the between-group difference to be as large as the within-patient improvement to achieve "clinical significance." Even if we accept the MCI for between-treatment comparison, the interpretation of results could be overly simplistic. Contrary to statistically significant cutoffs for effect sizes, MCIIs are a distributional characteristic that carries a degree of uncertainty, and it should be expressed as a range, not a single point. Therefore, for between-group comparison, the results should be expressed as proportions of patients in each group that achieved outcomes reaching or exceeding the lower bound of the MCI range. Accordingly, even when treatment effects did not reach an MCI "cutoff," there could still be a significant proportion of patients who crossed that threshold and for whom clinically significant improvement was achieved. In conclusion, the AAOS should be commended for its efforts to highlight the fact that significant effect sizes do not always translate into clinically pertinent findings. However, the question arises as to whether its committee’s recommendations were biased by improper use of a single metric in an analysis that used a very small subset of studies based on arbitrary criteria. Applied appropriately, MCIIs can be a useful tool for comprehensive presentation of study results; however, it should not be a cornerstone of clinical decision making. In addition, multiple variables associated with treatment of a patient with OA, such as arthritis severity, the number of compartments involved, alignment, age,
body weight, activity level, and medical comorbidities, were not addressed in these guidelines. Although guidelines can assist treatment decisions, they are not definitive and must therefore be used with caution and as an ancillary to clinical reasoning on a case-by-case basis. Guideline developers should word recommendations carefully to avoid losing therapeutic options that may benefit a subset of patients in a field in which few safe and effective treatments are available.

References