

How Appropriate Are Appropriate-use Criteria?



In their paper titled “Appropriateness and Total Hip Arthroplasty: Determining the Structure of the American Academy of Orthopaedic Surgeons System of Classification,” Riddle and Perera have analyzed the American Association of Orthopaedic Surgeons (AAOS) appropriate-use criteria (AUC) for total hip arthroplasty (THA)¹. They aimed to determine the contribution of each of the variables included by the AAOS (age, function-limiting pain, hip radiographic evaluation, range-of-motion limitation, presence or absence of modifiable risk factors) to the classification of appropriateness. An appropriate procedure is commonly defined as one for which “the expected health benefits significantly exceed the expected health risks by a wide margin,” based on the best available evidence². The aim of AUC is to improve patient care and outcomes, and to identify the complexities of clinical decision making, helping practitioners and patients make a decision about a specific procedure in a specific clinical condition. The US Center for Medicare and Medicaid Services (CMS) established a program to promote AUC in response to both overuse and underuse of medical procedures, and to link them to physician payments (now pushed back to 2020). In response to the CMS and cognizant of wide regional variations in the use of arthroplasty, the significant proportion of recipients who are dissatisfied, and the expenditure of billions of dollars annually, the AAOS developed AUC to guide management of osteoarthritis of the hip, including performance of THA³. Appropriateness differs from guideline recommendations, which provide overarching approaches to healthcare but cannot determine whether the procedure should be performed in an individual patient’s situation. This is where AUC can be used for guidance in decision making, because AUC can identify gradations in severity of disease or risk in specific clinical situations.

In a process using the RAND/University of California at Los Angeles Appropriateness Method, the AAOS participants performed a review of published evidence to identify the key predictor variables for THA outcomes, and then created 270 brief vignettes that included those variables ranked by severity³.

The selected variables included age, function-limiting pain, hip radiographic findings, range-of-motion limitation, and the presence or absence of modifiable surgical risk factors. Choices included specific age ranges: young (< 40 yrs), middle-aged (around 40–65 yrs), or elderly (around 65 yrs or more), while the choices for function-limiting pain included pain while walking moderate to long distances through pain at rest or at night, and so forth for each variable. The vignettes were then graded by an expert panel as “appropriate” (scored 7, 8, 9), “may be appropriate” (scored 4, 5, 6), or “rarely appropriate” (scored 1, 2, 3), and a consensus method was used to determine mean appropriateness rating scores³. By responding to hypothetical scenarios, a clinician can use the publicly available AAOS AUC with their patients as a decision-making tool and generate a score ranking the appropriateness of a procedure, and this score is informed by the expert panel’s ranking choices.

However, there are potential limitations in AUC, and careful analysis of the components of an AUC is important to ensure accuracy. The definition of appropriateness can vary depending on the composition of the expert panel and whether patients are included, whether cost or outcome data are considered, as well as the quality of the evidence and whether it is current. These factors may vary between groups and may lead to bias. For total knee arthroplasty, for example, there is significant discordance in the cases determined to be appropriate when 2 different validated appropriateness algorithms are applied^{4,5}. It is important to study AUC carefully, as their application may affect patient access to care and payment for care.

Riddle and Perera used the AAOS vignettes and performed a multinomial regression to predict the relative contribution of each variable to the appropriateness classification¹. They additionally used a classification tree — a machine learning approach to predictive modeling — that permits inclusion of more than 2 observed values, and tests each value to determine which variable most strongly associates with the appropriateness classification. They

See Appropriateness and hip arthroplasty, page 1127

found that age and radiographic severity increased the odds of being classified as appropriate significantly more than the other variables selected by the AAOS and included in the model. The authors noted that the THA AUC were highly dependent on traditional measures of age and radiographic severity, and the effect of function-limiting pain on the classification of appropriateness was small, even though most patients report that function-limiting pain is the primary motivation for surgery^{6,7,8,9}. Although there were 5 factors included after literature review, 2 of the factors dominated the panel's determinations.

How can we explain their work, and what does this mean for the AAOS THA AUC? The discrepancy described by Riddle and Perera may be due to the lack of inclusion of patients and their perspectives in the AAOS panels for either writing the vignettes or voting on appropriateness¹. However, concordant with the results of the literature review, surgeons also rate function-limiting pain and pain effect in appropriateness for THA, and voice concerns in interviews about the relationship of pain to radiographic structural damage, while cognizant of data that indicate that radiographic severity predicts postoperative improvements in pain and function^{7,10,11}. Moreover, the literature also reveals smaller improvements for older patients than younger patients and no correlation between age and multiple outcome measures. Moreover, implant longevity is less of a deterrent for young patients because the longterm performance of highly cross-linked polyethylene demonstrates minimal wear and osteolysis well into the second decade¹⁰. Riddle and Perera conclude that the AAOS THA AUC should be used with caution because the variables of age and radiographic severity played a disproportionate role in the panel's rankings, despite selection of 5 predictor variables through their literature review and synthesis¹.

Understanding the variables that are included and ranked in AUC is critical, so the concerns voiced by Riddle and Perera are helpful and lead to an assessment of the qualities that should determine appropriateness for any procedure. Everyone involved — including patients — should be included in the determination of AUC. The AAOS used 2 panels. First, the vignettes were written by one panel based on variables selected after synthesis of an extensive literature review, and next a separate panel, all specialists in hip surgery, voted and ranked the variables in 2 rounds. No patients were included, so factors of importance to patients were not given the same weight as those ranked by surgeons, even among the variables selected by literature review, and it is known that priorities can differ. The balance of benefit to harm that defines appropriateness should include the patients' perspective^{12,13}.

Second, AUC should be validated, using a different cohort of patients to determine whether the predictor variables function well, with an anchor such as patient satisfaction or change in quality of life as the outcome. A procedure

considered appropriate should have a high likelihood of an anticipated result. Do radiographic severity and patient age most significantly determine patient-reported outcome measures or are there specific levels of function-limiting pain that better predict a satisfactory outcome? Further study could clarify these questions and could inform the variable selection and rankings. When appropriateness criteria were retrospectively applied to arthroplasty cohorts by the authors of the study under discussion, those cases classified as appropriate or indeterminate had significant improvements in pain and function as measured using the Western Ontario and McMaster Universities Arthritis Index, while those classified as inappropriate did not¹⁴. However, when the same data were analyzed using the final pain score (the "destination") to rate outcome rather than the change in score (the "journey"), the appropriateness rating did not predict the outcome⁵.

Finally, surgical risks and techniques have changed substantially. The AAOS considers modifiable risk factors as a key variable, and given the changes in anesthesia, surgical technique, and component design, this variable requires vigilance to ensure that risk assessments are up to date.

Riddle and Perera have provided a useful and thoughtful analysis of the AAOS AUC for THA, and determined that the contribution of patient age and the radiographic evaluation were disproportionate in the AAOS classification of appropriateness over the other important predictor variables identified in their literature review and synthesis¹. While it is likely that AUC will be important in determining access to care and CMS payments, the authors' concern about the excessive weight given to 2 of the 5 variables (which may reflect the lack of broad input) in turn certainly raises more concern. Determining AUC for THA is a valuable effort that will improve patient care with improved decision-making algorithms; however, the performance of the criteria should be validated and the patients' perspective should be included.

SUSAN M. GOODMAN , MD,
PETER K. SCULCO , MD,
Hospital for Special Surgery,
Weill-Cornell Medical School,
Department of Medicine,
New York, New York, USA.

Address correspondence to Dr. S.M. Goodman, Department of Rheumatology, Hospital for Special Surgery, 535 East 70th St., New York, New York 10021, USA. E-mail: goodmans@hss.edu. Dr. Goodman receives research support from Novartis, and is a member of the Guidelines Committee for the American College of Rheumatology. Dr. Sculco receives consultant fees paid by Lima Corporate.

REFERENCES

1. Riddle DL, Perera RA. Appropriateness and total hip arthroplasty: determining the structure of the American Academy of Orthopaedic Surgeons system of classification. *J Rheumatol* 2019;46:1127-33.
2. Brook RH, Chassin MR, Fink A, Solomon DH, Koseoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986; 2:53-63.

3. American Academy of Orthopaedic Surgeons. Management of osteoarthritis of the hip. Evidence-based clinical practice guideline. [Internet. Accessed February 1, 2019.] Available from: www.aaos.org/uploadedFiles/PreProduction/Quality/Guidelines_and_Reviews/OA%20Hip%20CPG_3.13.17.pdf
4. Ghomrawi HM, Alexiades M, Pavlov H, Nam D, Endo Y, Mandl LA, et al. Evaluation of two appropriateness criteria for total knee replacement. *Arthritis Care Res* 2014;66:1749-53.
5. Katz JN, Winter AR, Hawker G. Measures of the appropriateness of elective orthopaedic joint and spine procedures. *J Bone Joint Surg Am* 2017;99:e15.
6. Hawker G, Bohm ER, Conner-Spady B, De Coster C, Dunbar M, Hennigar A, et al. Perspectives of Canadian stakeholders on criteria for appropriateness for total joint arthroplasty in patients with hip and knee osteoarthritis. *Arthritis Rheumatol* 2015;67:1806-15.
7. Hoang A, Goodman SM, Navarro-Millán IY, Mandl LA, Figgie MP, Bostrom MP, et al. Patients and surgeons provide endorsement of core domains for total joint replacement clinical trials. *Arthritis Res Ther* 2017;19:267.
8. Gossec L, Paternotte S, Bingham CO 3rd, Clegg DO, Coste P, Conaghan PG, et al. OARSI/OMERACT initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. An OMERACT 10 Special Interest Group. *J Rheumatol* 2011;38:1765-9.
9. Frankel L, Sanmartin C, Conner-Spady B, Marshall DA, Freeman-Collins L, Wall A, et al. Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. *Osteoarthritis Cartilage* 2012;20:967-73.
10. Hofstede SN, Gademán MG, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. *BMC Musculoskelet Disord* 2016;17:212.
11. Frankel L, Sanmartin C, Hawker G, De Coster C, Dunbar M, Bohm E, et al. Perspectives of orthopaedic surgeons on patients' appropriateness for total joint arthroplasty: a qualitative study. *J Eval Clin Pract* 2016;22:164-70.
12. Gibofsky A, Galloway J, Kekow J, Zerbin C, de la Vega M, Lee G, et al. Comparison of patient and physician perspectives in the management of rheumatoid arthritis: results from global physician- and patient-based surveys. *Health Qual Life Outcomes* 2018;16:211.
13. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ* 2001;323:1218-22.
14. Riddle DL, Perera RA, Jiranek WA, Dumenci L. Using surgical appropriateness criteria to examine outcomes of total knee arthroplasty in a United States sample. *Arthritis Care Res* 2015;67:349-57.

J Rheumatol 2019;46:1064–6; doi:10.3899/jrheum.190012