



Original Research

Agreement of the American Academy of Orthopaedic Surgeons Appropriate Use Criteria With Treatment Recommendations From Arthroplasty Surgeons

T. Jacob Selph Jr., BS^{a, b}, Linda I. Suleiman, MD^a, Manasa S. Pagadala, BS^b, Rachel Bergman, MD^a, Patricia D. Franklin, MD, MBA, MPH^c, Adam I. Edelman, MD^{a, *}

^a Department of Orthopaedic Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

^b Northwestern University Feinberg School of Medicine, Chicago, IL, USA

^c Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

ARTICLE INFO

Article history:

Received 3 January 2024

Received in revised form

2 March 2024

Accepted 24 March 2024

Keywords:

Appropriate use criteria

Total knee arthroplasty

Total hip arthroplasty

Osteoarthritis

ABSTRACT

Background: The American Academy of Orthopaedic Surgeons (AAOS) Appropriate Use Criteria (AUC) for Surgical Management of Osteoarthritis of the Knee (2016) and Management of Osteoarthritis of the Hip (2017) are intended to provide treatment recommendations for osteoarthritis (OA). This study examined the agreement of AUC appropriateness classifications with arthroplasty surgeon recommendations for total knee arthroplasty (TKA) and total hip arthroplasty (THA).

Methods: The cohort included 558 OA patients (397 knee, 161 hip) referred to a specialty arthroplasty clinic. Surgeons completed the online AAOS AUC patient profiles to generate appropriateness ratings. Surgeons' recommendations for treatment were recorded. We performed univariate and bivariate analyses to evaluate relationships between AUC appropriateness and surgeon recommendations.

Results: The knee OA AUC classified TKA as "appropriate" for 309 (77.8%) of the 397 knee OA patients. Surgeons recommended TKA for 123 (31.0%), resulting in 46.8% (n = 186) higher rate of "appropriate" classification by AUC than TKA recommendation by surgeons. Weighted Cohen's κ demonstrated slight agreement ($\kappa = 0.06$, 95% confidence interval: 0.04, 0.09) between AUC appropriateness and surgeon TKA recommendation. The hip OA AUC classified THA as "appropriate" for 98 (60.9%) of the 161 hip OA patients. Surgeons recommended THA for 76 (47.2%), resulting in 13.7% (n = 22) higher rate of "appropriate" classification by AUC than THA recommendation by surgeons. Weighted Cohen's κ demonstrated moderate agreement ($\kappa = 0.47$, 95% confidence interval: 0.37, 0.57) between the AUC appropriateness classification and the surgeon's THA recommendation.

Conclusions: AAOS AUC guidelines indicated surgical appropriateness significantly more than arthroplasty surgeons. AUC agreed slightly with surgeons for TKA and moderately for THA.

© 2024 The Authors. Published by Elsevier Inc. on behalf of The American Association of Hip and Knee Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Total joint arthroplasty (TJA) is a common treatment for osteoarthritis (OA) of the knee and hip. By 2025, the annual primary total knee arthroplasty (TKA) utilization in the USA is projected to reach 1,272,000 (95% confidence interval [CI]: 1,200,000-1,710,000) and the primary total hip arthroplasty (THA) utilization is projected to reach 652,000 (95% CI: 610,000-696,000) [1]. Though common and

effective, TJA is a major surgery that carries the risk of unsatisfactory outcomes. Data indicate that 16% of TKA patients and 10.5% of THA patients report unsatisfactory outcomes during 1-year post-operative follow-up [2]. Efforts to minimize or eliminate suboptimal outcomes have included the creation of standardized decision-making tools to aid in the shared decision-making process between patients and surgeons.

In 2003, Escobar et al. used the RAND/University of California Los Angeles (UCLA) method to create the first appropriateness criteria for TKA [3]. In 2016 and 2017, the American Academy of Orthopaedic Surgeons (AAOS) used the RAND/UCLA method to create the current appropriate use criteria (AUC) for the surgical management of osteoarthritis of the knee and management of

* Corresponding author. 676 N. St. Clair St, Suite 1350, Chicago, IL 60611, USA. Tel.: +1 312 695 6800.

E-mail address: aedelst1@nm.org

osteoarthritis of the hip, respectively [4,5]. These criteria may be used to classify the decision to use specific treatment modalities as “appropriate,” “may be appropriate,” or “rarely appropriate.” After the AUC release, Riddle et al. analyzed the AUC for TKA and THA to identify which patient characteristics were most weighted within the AUC, although they did not compare the AUC results with surgeon recommendations [6,7]. Their analysis revealed that the strongest variables affecting TKA appropriateness classification in the surgical management of osteoarthritis of the knee were patient age, knee motion, radiographic OA severity, and pattern of knee compartment involvement, whereas function-limiting pain contributed little [6]. Their investigation of the management of osteoarthritis of the hip revealed that patient age and radiographic OA severity were, by far, the most powerful predictive variables leading to appropriateness for THA and that pain resulting in functional limitation did not contribute significantly to appropriateness for surgery [7]. Evidence suggests that arthritic pain (level of intensity, lifestyle impact, and ability to cope) is the most important factor for patients and surgeons in deciding whether to elect for TJA [8-10].

Given the discrepancy between the most powerful predictive variables in the AAOS AUC on the one hand and surgeon and patient priorities on the other, an evaluation of the utility of the AUC in modern practice is warranted. To date, validation studies on this topic have only been performed outside of the USA. [11-13]. We therefore evaluated the agreement of the AAOS AUC Surgical Management of Osteoarthritis of the Knee and Management of Osteoarthritis of the Hip with fellowship-trained arthroplasty surgeon treatment recommendations in a United States cohort. Previous research demonstrating discrepancies between the variables most predictive of AUC appropriateness classifications and the priorities of surgeons and patients prompted us to hypothesize that there could be a considerable level of disagreement between AUC surgical appropriateness ratings and fellowship-trained arthroplasty surgeons' recommendations.

Material and methods

Patients and data

This study was performed on a prospective cohort of 558 consecutive patients with OA (397 knee and 161 hip), who presented to a specialty arthroplasty clinic for evaluation of their respective condition by 1 of 4 participating fellowship-trained arthroplasty surgeons. We excluded patients with non-OA diagnoses from the analyses. All qualifying patients evaluated between January and April of 2023 were invited to provide informed consent and join the study cohort. At the time of visit, the knee OA sample group patient mean age was 66.5 years (range, 24-96) and mean body mass index (BMI) was 32.5 kg/m² (range, 18.5 -67.4). The sample contained 119 (30.0%) men and 278 (70.0%) women. Knee OA patient demographics are contained in Table 1. At the time of visit, the hip OA sample group patients mean age was 64 years (range, 27-91) and mean BMI was 30.2 kg/m² (range, 17.0-47.9). The group contained 60 (37.3%) men and 101 (62.7%) women. Hip OA patient demographics are included in Table 2.

Surgeons completed consultation with each patient according to their usual practice. At the end of each patient encounter, the surgeon's treatment recommendation was recorded. Surgeon recommendations were categorized as 1) recommend TKA/THA; 2) recommend nonsurgical care in the absence of medical contraindication to TKA/THA; and 3) recommend nonsurgical care due to medical contraindication to TKA/THA. At the conclusion of each patient visit, surgeons also completed the patient profile for the online knee or hip OA AAOS AUC (Fig. 1) and the resulting AUC

Table 1
Knee OA patient demographics.

| Variable | Mean | Range |
|------------------------------------|------------------------|-----------------------------|
| Age | 66.5 y | 24-96 y |
| BMI | 32.5 kg/m ² | 18.5-67.4 kg/m ² |
| Patient sex | N (total = 397) | % of knee OA patients |
| Men | 119 | 30.0 |
| Women | 278 | 70.0 |
| Insurance type | | |
| Medicare | 197 | 49.6 |
| PPO | 144 | 36.3 |
| Medicaid | 49 | 12.3 |
| HMO | 2 | 0.5 |
| Military | 1 | 0.3 |
| Workers' compensation | 1 | 0.3 |
| Uninsured | 3 | 0.8 |
| Ethnicity/race | | |
| White | 181 | 45.6 |
| Black | 145 | 36.5 |
| Asian | 2 | 0.5 |
| Native American | 1 | 0.3 |
| Hawaiian/Pacific Islander | 0 | |
| Hispanic ethnicity, race - "other" | 37 | 9.3 |
| Hispanic ethnicity, race - White | 6 | 1.5 |
| Hispanic ethnicity, race - Black | 0 | |
| "Other" | 3 | 0.8 |
| Unknown | 5 | 1.3 |
| Declined to report | 5 | 1.3 |

PPO, preferred provider organization; HMO, health maintenance organization.

appropriateness rating for each available intervention was recorded by nonsurgeon research staff. The AUC appropriateness classification result was not shared with surgeons to avoid biasing their recommendations. The knee AUC provides appropriateness for 3 surgical procedures (TKA, unicompartmental knee arthroplasty [UKA], and realignment osteotomy [RO]); the hip AUC provides appropriateness for surgical and nonsurgical treatments. Patient demographic information and clinical characteristics were recorded. This study was performed with supervision and approval by our institutional review board.

Table 2
Hip OA patient demographics.

| Variable | Mean | Range |
|-----------------------------------|------------------------|---------------------------|
| Age | 64 y | 27-91 y |
| BMI | 30.2 kg/m ² | 17-47.9 kg/m ² |
| Patient sex | N (total = 161) | % of hip OA patients |
| Men | 60 | 37.3 |
| Women | 101 | 62.7 |
| Insurance type | | |
| Medicare | 74 | 46.0 |
| PPO | 59 | 36.7 |
| Medicaid | 24 | 14.9 |
| HMO | 0 | |
| Military | 3 | 1.9 |
| Uninsured | 1 | 0.6 |
| Ethnicity/race | | |
| White | 83 | 51.6 |
| Black | 66 | 41.0 |
| Asian | 1 | 0.6 |
| Native American | 1 | 0.6 |
| Hawaiian/Pacific Islander | 0 | |
| Hispanic ethnicity, race- "other" | 5 | 3.1 |
| Hispanic ethnicity, race - White | 2 | 1.2 |
| Hispanic ethnicity, race - Black | 0 | |
| "Other" | 0 | |
| Unknown | 3 | 1.9 |
| Declined to report | 0 | |

PPO, preferred provider organization; HMO, health maintenance organization.

a

APPROPRIATE USE CRITERIA: OSTEOARTHRITIS OF THE KNEE: SURGICAL MANAGEMENT (2016)

INDICATION PROFILE

Function-Limiting Pain

- Function-limiting pain at moderate to long distances (walking moderate to long distances greater than one fourth mile)
- Function-limiting pain at short distances (limiting activity to two city blocks, the equivalent to walking the length of a shopping mall)
- Pain at rest or night

Range of Motion Extension/Flexion

- Full range of extension/flexion
- Lack of full extension (> 5 degree flexion contracture) and/or flexion < 110 degrees
- Lack of full extension (> 10 degree flexion contracture) and/or flexion < 90 degrees

Functional Instability

- No functional instability
- Functional instability

Pattern of arthritic involvement (medial tibiofemoral, lateral tibiofemoral or patellofemoral)

- Predominantly one compartment
- More than one compartment

Imaging (joint space in most involved compartment)

- Mild to moderate – Joint space narrowing as visible on imaging
- Severe

Limb Alignment

- Normal Alignment
- Varus/ valgus

Mechanical Symptoms (compatible with Meniscal Tear or Loose Body)

- Mechanical Symptoms
- No Mechanical Symptoms

Age

- Young
- Middle-aged
- Elderly

PROCEDURE RECOMMENDATIONS

SUBMIT

b

APPROPRIATE USE CRITERIA: OSTEOARTHRITIS OF THE HIP: MANAGEMENT (2017)

INDICATION PROFILE

Age

- Young (Approximately < 40)
- Middle-Aged (Approximately 40-65)
- Elderly (Approximately > 65)

Function Limiting Pain

- Function-Limiting Pain at Moderate to Long Distances
- Function-Limiting Pain at Short Distances
- Pain at Rest or Night

Radiographic Evaluation

- Minimal OA
- Minimal OA with acetabular dysplasia
- Minimal OA with FAI
- Moderate OA
- Severe OA

Range of Motion Limitation

- Minimal Range of Motion Limitation
- Moderate Range of Motion Limitation
- Severe Range of Motion Limitation

Risk of Patient for Negative Outcome

- Modifiable risk factors present
- No modifiable risk factors present

PROCEDURE RECOMMENDATIONS

SUBMIT

Figure 1. American Academy of Orthopaedic Surgeons appropriate use criteria indication profile input screens for (a) surgical management of knee osteoarthritis and (b) management of hip osteoarthritis.

Data analyses

Knee and hip OA patient data were compiled and analyzed by SAS software (Cary, North Carolina). We report knee and hip OA sample demographic data, AUC appropriateness ratings, and surgeon recommendations. G-testing was performed to confirm the significance of observed relationships between the AUC appropriateness classification and the surgeon’s TKA or THA recommendation. Significance cutoff was set at $\alpha = 0.05$. Where appropriate, odds ratios were calculated to assess the likelihood of surgeon recommendations in relation to AUC appropriateness ratings for TKA and THA. Interrater agreement between AUC appropriateness classification and surgeon TKA/THA recommendation was evaluated with weighted Cohen’s κ using Fleiss-Cohen weighting. Previous research established that κ statistic <0 indicates poor agreement, 0-0.2 slight agreement, 0.21-0.4 fair agreement, 0.41-0.6 moderate agreement, 0.61-0.8 substantial agreement, and 0.81-1 almost perfect agreement [14].

Results

Of the 397 knee OA patients, AUC classified TKA as “appropriate” for 309 (77.8%) and “may be appropriate” for 88 (22.2%) patients. AUC did not classify TKA as “rarely appropriate” for any patients in the sample. AUC classified UKA as “appropriate” for 13 (3.3%) patients, “may be appropriate” for 125 (31.5%) patients, and “rarely appropriate” for 259 (65.2%) patients. AUC did not classify RO as “appropriate” for any patient. AUC classified RO as “may be appropriate” for 23 (5.8%) patients and “rarely appropriate” for 374 (94.2%) patients. Knee OA AUC treatment classification data are contained within Table 3.

Surgeons recommended TKA for 123 (31.0%) of the 397 knee OA patients, resulting in an overall 46.8% higher probability of “appropriate” classification by AUC algorithm than arthroplasty surgeon recommendation for TKA. Patients for whom TKA was

rated as “appropriate” were less likely to be recommended surgery than nonoperative care with 36.9% probability of being recommended TKA, 41.1% probability of being recommended nonoperative care without specific contraindication against TKA, and 22.0% probability of recommendation against TKA due to contraindication (Fig. 2). Those for whom TKA was rated as “may be appropriate” were also less likely to be recommended surgery than nonoperative care with 10.2% probability of being recommended TKA, 76.1% probability of being recommended nonoperative care without specific contraindication to TKA, and 3.0% probability of being recommended against TKA due to contraindication (Fig. 2). Significance of observed relationships between AUC appropriateness and surgeon recommendations were supported by G-test ($\chi^2 = 26.5, P < .0001$). Patients for whom TKA was classified as “appropriate” were more likely to be recommended TKA than those for whom it was classified as “may be appropriate” (odds ratio: 5.1, 95% CI: 2.5, 10.6). Specific medical contraindications leading surgeons to recommend nonoperative treatment to patients for whom AUC rated TKA as “appropriate” included BMI, insufficiently controlled diabetes mellitus, and tobacco use. Surgeons’ rationale for recommending nonoperative treatment to patients for whom TKA was rated as “appropriate” by AUC are detailed within Table 4. No patients were recommended UKA or RO. Calculated weighted Cohen’s κ for interrater agreement demonstrated only slight agreement ($\kappa = 0.06, 95\% \text{ CI: } 0.04, 0.09$) between the AUC appropriateness classification and the surgeon’s TKA recommendation.

Table 3
Patients by AUC classification for TKA, UKA, and RO.

| AUC appropriateness category | Knee OA surgical management options | | |
|------------------------------|-------------------------------------|-------------|-------------|
| | TKA | UKA | RO |
| Appropriate | 309 (77.8%) | 13 (3.3%) | 0 |
| May be appropriate | 88 (22.2%) | 125 (31.5%) | 23 (5.8%) |
| Rarely appropriate | 0 | 259 (65.2%) | 374 (94.2%) |

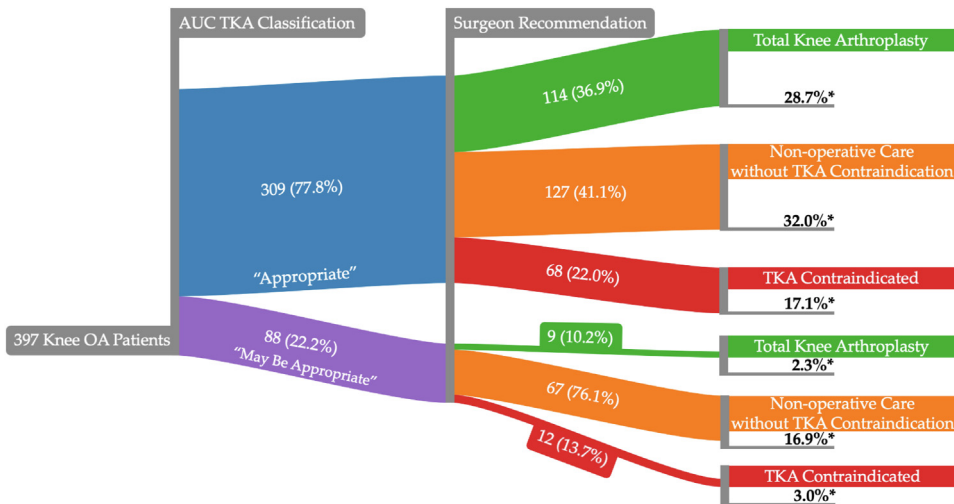


Figure 2. Sankey diagram demonstrating relationships between appropriate use criteria classification and surgeon total knee arthroplasty recommendations. Percentages listed in flow represent the probability of each surgeon’s recommendation for the respective appropriate use criteria classification group. %* represents the percentage of patients relative to the total knee osteoarthritis cohort.

Treatment categories included in the AAOS AUC for management of hip osteoarthritis are risk factor assessment and optimization, activity modifications, assistive devices, oral nonsteroidal anti-inflammatory drug (NSAID)/acetaminophen/tramadol, corticosteroid injection (CSI), physical therapy, THA, hip preservation surgery, and arthrodesis. Risk factor assessment and optimization, activity modifications, assistive devices, oral NSAID/acetaminophen/tramadol, and CSI were rated as “appropriate” for every hip OA patient in the study sample. Hip preservation surgery and arthrodesis were rated as “rarely appropriate” for every patient in the study sample. Physical therapy was rated as “appropriate” for 158 (98.1%) patients and “may be appropriate” for 3 (1.9%) patients. THA was rated as “appropriate” for 98 (60.9%) patients, “may be appropriate” for 29 (18.0%) patients, and “rarely appropriate” for 34 (21.1%) patients. Hip OA AUC classification data are represented in Table 5.

Surgeons recommended THA for 76 (47.2%) of the 161 hip OA patients, resulting in 13.67% (n = 22) higher overall probability of “appropriate” classification by the AUC algorithm than a recommendation for THA by fellowship-trained arthroplasty surgeon. Unlike the pattern observed for TKA, patients for whom THA was rated as “appropriate” were more likely to be recommended THA than nonoperative care with 72.4% probability of being recommended THA, 13.3% probability of being recommended nonoperative care without specific contraindication to THA, and 14.3% probability of being recommended against THA due to contraindication (Fig. 3). Like TKA, patients for whom THA was classified as “may be appropriate” were less likely to be recommended surgery than nonoperative care with 10.3% probability of being

recommended THA, 65.6% probability of being recommended nonoperative care without specific contraindication to THA, and 24.1% probability of being recommended against THA due to contraindication (Fig. 3). Patients for whom THA was classified as “rarely appropriate” were less likely to be recommended surgery than nonoperative care with 5.9% probability of THA recommendation and 94.1% probability of recommendation for nonoperative care without specific contraindication to THA (Fig. 3). Significance of observed relationships between AUC appropriateness and surgeon recommendations were supported by G-test ($\chi^2 = 72.8, P < .0001$). Patients for whom THA was classified as “appropriate” were more likely to be recommended for the procedure by surgeons than patients for whom THA was classified as “may be appropriate” or “rarely appropriate” (odds ratio: 30.5, 95% CI: 11.1, 84.2). Specific medical contraindications leading surgeons to recommend nonoperative treatment to patients for whom AUC rated THA as “appropriate” included BMI, insufficiently controlled diabetes mellitus, and tobacco use. Surgeons’ rationale for recommending nonoperative treatment to patients for whom THA was rated as “appropriate” by AUC are detailed within Table 4. Calculated weighted Cohen’s κ for interrater agreement demonstrated moderate agreement ($\kappa = 0.47, 95\% \text{ CI: } 0.37, 0.57$) between the AUC appropriateness classification and the surgeon’s THA recommendation.

Discussion

This study was the first to evaluate the validity of the AAOS Surgical Management of Osteoarthritis of the Knee and Management of Osteoarthritis of the Hip for TKA and THA in a US cohort. We examined the agreement between AAOS appropriateness ratings and fellowship-trained surgeon recommendations for TKA and THA based on clinical and radiographic evaluation. The results showed that the AUC classified TKA and THA as “appropriate” for 47% and 14% more patients than were recommended by surgeons, respectively. The AUC significantly over-recommend surgery compared to fellowship-trained arthroplasty surgeons’ clinical judgment. The AUC are intended to designate those patients for whom interventions offer expected health benefits that exceed the possible negative consequences by a sufficiently wide margin [4,5]. Our findings suggest that care providers should exercise caution

Table 4
Patients not recommended surgery with an AUC “appropriate” rating.

| | Knee OA | Hip OA |
|--|---------|--------|
| No medical contraindication | | |
| Patient has not failed nonoperative management | 123 | 9 |
| Mild joint symptoms only | 4 | 4 |
| Medical contraindication | | |
| BMI above recommended range | 44 | 7 |
| Insufficiently controlled diabetes | 3 | 0 |
| Tobacco use | 1 | 1 |
| Medical comorbidity not otherwise specified | 20 | 6 |

Table 5
Patients by level of appropriateness for nine treatment categories in hip osteoarthritis: appropriate use criteria.

| Hip OA treatment option | Appropriateness category | | |
|---|--------------------------|--------------------|--------------------|
| | Appropriate | May be appropriate | Rarely appropriate |
| Risk factor assessment and optimization | 161 (100%) | 0 | 0 |
| Activity modifications | 161 (100%) | 0 | 0 |
| Assistive devices | 161 (100%) | 0 | 0 |
| Oral NSAID, acetaminophen, or tramadol | 161 (100%) | 0 | 0 |
| CSI | 161 (100%) | 0 | 0 |
| PT | 158 (98.1%) | 3 (1.9%) | 0 |
| THA | 98 (60.9%) | 29 (18.0%) | 34 (21.1%) |
| Hip preservation surgery | 0 | 0 | 161 (100%) |
| Arthrodesis | 0 | 0 | 161 (100%) |

PT, physical therapy.

when utilizing the AAOS AUC appropriateness criteria and maintain a wariness of their tendency to over-recommend surgery.

An effective standardized decision-making tool would be expected to be able to classify management decisions in such a way as to be significantly aligned with those made by fellowship-trained arthroplasty surgeons. This expected pattern was not observed for TKA in our study, as patients for whom TKA was classified as “appropriate” received significantly fewer recommendations for TKA than nonoperative care by arthroplasty surgeons (Fig. 2). Our results suggest that the RAND/UCLA methodology used to create the AAOS AUC for Surgical Management of Osteoarthritis of the Knee may have yielded an inaccurate evaluation of TKA appropriateness in some cases. This is in line with previous work which suggests that the factors most important to both patients and surgeons (function-limiting pain) for TKA candidacy decisions are not the main drivers of the AUC appropriateness classification [6,8–10]. Another factor that may account for the AUC’s over-indication of TKA is the fact that specific medical contraindications are not accounted for in the AUC algorithm and are instead a factor that is left for surgeons to recognize and apply. This accounted for 34.8% (n = 68) of patients with a surgery rating of “appropriate” who were not recommended for surgery (Fig. 2). Our study contrasts with a previous work performed in a Qatari cohort suggesting that the TKA AUC was solidly in line with surgeons’ clinical

decision-making [11]. This divergence may stem from differing patient populations or differing surgeon practice patterns.

Though the AUC Management of Osteoarthritis of the Hip also demonstrated a tendency to over-recommend THA, as previously mentioned, our data showed that it did perform better in approximating the clinical decisions made by our surgeons; patients for whom THA was classified as “appropriate” were recommended for the procedure more often than nonoperative care. The observed tendency to recommend THA more often than surgeons may be partially accounted for by the findings of previous work that the case characteristics that patients and surgeons most often use for THA decision-making (functional pain) are not the main drivers of THA AUC classification [7–10]. This is likely a major reason why 13.3% (n = 13) of patients for whom THA was classified as “appropriate” were treated with nonoperative care instead of THA. A portion of the AUC tendency to over-recommend THA likely also stems from the fact that medical contraindications are not accounted for within the AUC algorithm and are instead left to the clinician’s judgment. This was the case for 51.9% (n = 14) of patients for whom THA was classified as “appropriate” but were not recommended for THA. Overall, our results regarding THA are comparable to those in previous studies evaluating the accuracy of hip OA AUC performed in a Qatari cohort [12].

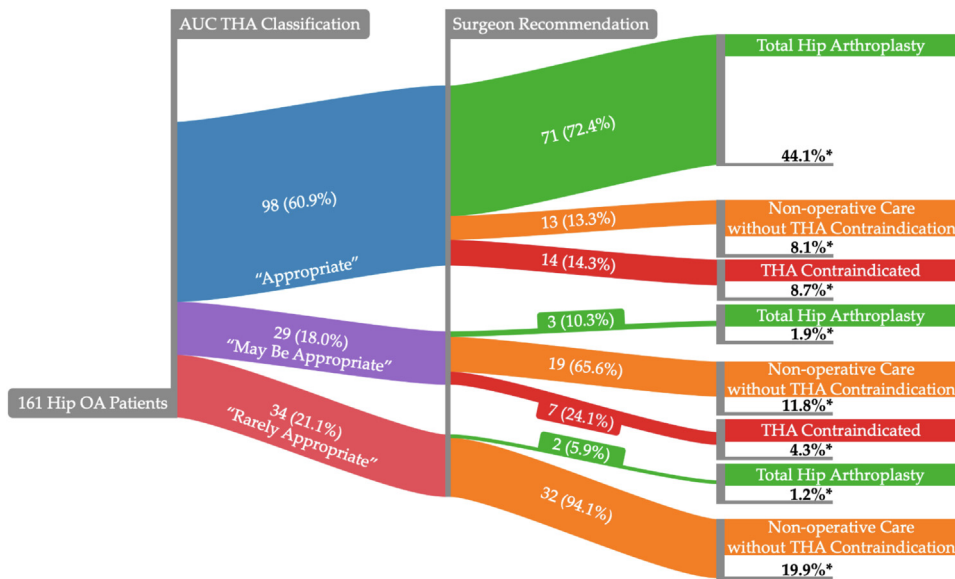


Figure 3. Sankey diagram demonstrating relationships between appropriate use criteria classification and surgeon total hip arthroplasty recommendations. Percentages listed in each flow represent the probability of each surgeon’s recommendation for the respective appropriate use criteria classification group. *% represents the percentage of patients relative to the total hip osteoarthritis cohort.

The utility that the hip OA AUC have in determining which patients are candidates for non-THA treatments is called into question by the fact that risk factor assessment and optimization, activity modifications, assistive devices, oral NSAID/acetaminophen/tramadol, and CSI were rated as “appropriate” for every single hip OA patient and that hip preservation surgery and arthrodesis were rated as “rarely appropriate” for every single hip OA patient. Further research into the agreement of the AUC Management of Osteoarthritis of the Hip with arthroplasty surgeon recommendations for non-THA treatment is merited.

Our study was subject to limitations. This study was performed at a single center within an academic tertiary care hospital in the USA. Findings may not have generalizability to population groups and practice locations that are different from our own. The results of our single-center validation study indicate that expansion to a multicenter study is warranted and necessary to investigate the agreement of the AAOS AUC for Surgical Management of Osteoarthritis of the Knee and Management of Osteoarthritis of the Hip with recommendation of TKA and THA from fellowship-trained arthroplasty surgeons in a US population. We also dichotomized the surgeon’s recommendation regarding surgical intervention, which may not fully represent the nuanced discussions that occur in the decision-making process. Nonetheless, our data accurately reflect whether the surgeon, through a process of shared decision-making with the patient, ultimately made a recommendation for surgery as the next best step. Furthermore, the double-blinded AUC algorithm data entry was subject to human error.

Conclusions

Our hypothesis that the AUC would demonstrate a considerable level of disagreement with the recommendations of our surgeons was supported for TKA and was moderately supported for THA. The AAOS AUC for Surgical Management of Osteoarthritis of the Knee showed a strong tendency to over-recommend TKA compared with our surgeons and demonstrated only slight agreement with them ($\kappa = 0.06$, 95% CI: 0.04, 0.09). The AAOS AUC Management of Osteoarthritis of the Hip appropriateness classifications for THA were mostly in line with our surgeons’ recommendations but still found to have only a moderate level of agreement ($\kappa = 0.47$, 95% CI: 0.37, 0.57). A multicenter study to verify these results is needed.

Conflicts of interest

L. Suleiman is a speaker for DePuy Synthes Company and Ethicon Inc., is a paid consultant for Stryker Corporation, and is a board/committee member of the AAHKS Program Committee, the AAHKS Women in Arthroplasty, and the Ruth Jackson Orthopaedic Society. P. D. Franklin is a board/committee member of the National Scientific Advisory Committee for the Arthritis Foundation. A. I. Edelstein is a paid consultant for Corin Group and DePuy Synthes Company, is an editorial board member of *Arthroplasty Today*, and is a board/committee member of the AAHKS Publications Committee. All other authors declare no potential conflicts of interest.

For full disclosure statements refer to <https://doi.org/10.1016/j.artd.2024.101386>.

CRedit authorship contribution statement

T. Jacob Selph: Writing – original draft, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization. **Linda I. Suleiman:** Supervision, Resources, Conceptualization, Investigation, Methodology, Project administration, Writing – review & editing. **Manasa S. Pagadala:** Writing – review & editing, Data curation. **Rachel Bergman:** Data curation, Writing – review & editing. **Patricia D. Franklin:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Adam I. Edelstein:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization, Formal analysis, Investigation.

References

- [1] Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: future projections to 2020–2040 using the national inpatient sample. *J Rheumatol* 2019;46:1134–40. <https://doi.org/10.3899/jrheum.170990>.
- [2] DeFrance MJ, Scuderi GR. Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature. *J Arthroplasty* 2023;38:594–9. <https://doi.org/10.1016/j.arth.2022.10.011>.
- [3] Escobar A, Quintana JM, Aróstegui I, Azkárate J, Güenaga JI, Arenaza JC, et al. Development of explicit criteria for total knee replacement. *Int J Technol Assess Health Care* 2003;19:57–70. <https://doi.org/10.1017/S0266462303000060>.
- [4] American Academy of Orthopaedic Surgeons. *Surgical Management of Osteoarthritis of the Knee Appropriate Use Criteria*. Rosemont, IL: American Association of Orthopaedic Surgeons; 2016.
- [5] American Academy of Orthopaedic Surgeons. *Appropriate Use Criteria for the Management of Osteoarthritis of the Hip*. Rosemont, IL: American Association of Orthopaedic Surgeons; 2017.
- [6] Riddle DL, Perera RA. Appropriateness and total knee arthroplasty: an examination of the American Academy of Orthopaedic Surgeons appropriateness rating system. *Osteoarthritis Cartilage* 2017;25:1994–8. <https://doi.org/10.1016/j.joca.2017.08.018>.
- [7] 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. <https://doi.org/10.3899/jrheum.180911>.
- [8] 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. <https://doi.org/10.1016/j.joca.2012.05.008>.
- [9] 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. <https://doi.org/10.1002/art.39124>.
- [10] 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. <https://doi.org/10.1111/jep.12449>.
- [11] Ahmed GO, Elswify K, Ahmed AF. Usability of the AAOS Appropriate Use Criteria (AUC) for the surgical management of knee osteoarthritis in clinical practice. *Knee Surg Sports Traumatol Arthrosc* 2020;28:2077–81. <https://doi.org/10.1007/s00167-020-05908-7>.
- [12] Alkaramany E, Murshid A, Alhaneedi GA. Appropriateness of hip osteoarthritis management in clinical practice compared with the American Academy of Orthopaedic Surgeons (AAOS) criteria. *Int Orthop* 2021;45:2805–10. <https://doi.org/10.1007/s00264-021-05150-x>.
- [13] Karunaratne S, Harris IA, Trevena L, Horsley M, Solomon M. Observing the use of knee arthroplasty appropriateness tools in clinical practice: do appropriateness criteria tools predict surgeon decision-making? *Osteoarthritis Cartilage* 2021;29:1275–81. <https://doi.org/10.1016/j.joca.2021.06.009>.
- [14] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. <https://doi.org/10.2307/2529310>.