Article

AMERICAN ORTHOPAEDIC FOOT & ANKLE SOCIETY

Foot & Ankle International®
2021, Vol. 42(5) 624–632

DOI: 10.1177/1071100720977842
journals.sagepub.com/home/fai

Effect of Patient Demographics on Minimally Important Difference of Ankle Osteoarthritis Scale Among End-Stage Ankle Arthritis Patients

Jason M. Sutherland, PhD[1], Carmela Melina Albanese, BSc[1], Kevin Wing, MD[2], Yixiang Jenny Zhang, BSc[1], Alastair Younger, MD[2], Andrea Veljkovic, MD[2], and Murray Penner, MD[2]

Abstract

Background: Ankle replacement and ankle arthrodesis are standard treatments for treating end-stage ankle arthritis when conservative treatment fails. Comparing patient-reported outcome scores to the instrument's minimal important difference (MID) helps physicians and researchers infer whether a meaningful change in health from the patient's perspective has occurred following treatment. The objective of this study was to estimate the MID of the Ankle Osteoarthritis Scale among a cohort of operatively treated end-stage ankle arthritis patients undergoing ankle replacement or arthrodesis.
Methods: A survey package including the Ankle Osteoarthritis Scale was completed by participants preoperatively and 2 years postoperatively. Distribution and anchor-based approaches to calculating the MID were used to estimate the MID of the Ankle Osteoarthritis Scale and its 2 domains. The distribution-based approaches used were the small and medium effect size methods, while the mean absolute change method and linear regression method were the anchor-based approaches. Bootstrap sampling was used to obtain the variance of MID estimates. The MID was estimated for sex, age, operative, and baseline health subgroups. The cohort comprised 283 participants, totaling 298 ankles.
Results: The MID did not vary with sex or operative procedure. Age-based differences in MID values may exist for the Ankle Osteoarthritis Scale total score, and MID values were generally smallest among the oldest patients. Patients with the best and worst ankle-related health preoperatively had higher MID values than patients reporting mid-range Ankle Osteoarthritis Scale values preoperatively.
Conclusion: The best estimate of the MID of the Ankle Osteoarthritis Scale total score is 5.81. Our findings indicate that the MID of the Ankle Osteoarthritis Scale may not vary by sex or operative subgroups but likely varies by age and preoperative Ankle Osteoarthritis Scale score.
Level of Evidence: Level II, prospective comparative study.

Keywords: ankle replacement, ankle arthrodesis, Ankle Osteoarthritis Scale, minimally important difference

Introduction

End-stage ankle arthritis (ESAA) is a debilitating chronic condition characterized by pain and swelling at the ankle joint as a result of inflammatory, idiopathic, or post-traumatic causes.[1,18,21] ESAA is associated with poor physical functioning, pain, and reduced health-related quality of life (HRQoL).[2,21,39] Patients who fail conservative treatment of their ESAA are most commonly treated with either ankle arthrodesis (AA) or total ankle replacement (TAR).[14,32,40,42] AA is often recommended for younger

[1]Centre for Health Services and Policy Research, School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[2]Department of Orthopaedics, University of British Columbia, Vancouver, British Columbia, Canada

Corresponding Author:
Jason M. Sutherland, PhD, Centre for Health Services and Policy Research, School of Population and Public Health, University of British Columbia, 201-2206 East Mall, Vancouver, BC V6T 1Z3, Canada.
Email: jason.sutherland@ubc.ca

and more active, high-demand patients or patients with comorbidities compromising soft-tissue healing[32,44] and TAR is often recommended for patients aged 65 years or older with lower demands.[44] TAR has become increasingly popular[35] and has been found to be more cost-efficient than AA among both older and younger patients.[12,34]

To measure health-related quality of life, symptoms, and functioning from the patient's perspective, patient-reported outcomes (PROs) are increasingly being used.[16,37] Most PROs generate an ordinal value that represents the severity of disease, symptoms, or condition that the PRO purports to measure. With repeated application of PROs, changes in patients' health status, symptoms, or function, including before and after surgery[5,27] can be measured. PROs, including the Ankle Osteoarthritis Scale (AOS),[19,29] have become widely used for reporting patients' perspectives of their foot and ankle–related health,[19,21,29,36,39,41,44] including among ESAA patients throughout the perioperative period.[14,36,41]

Among the PRO literature, the minimal important difference (MID)—sometimes referred to as the minimal clinically important difference (MCID)—is the smallest amount of change in PRO value that a patient considers to be meaningful.[37] For example, if the MID of a symptom-based instrument is determined to be 5 points, patients reporting a 3-point change in their PRO value relative to their baseline score would not consider their symptoms to have changed. The MID is often used to detect clinical meaningfulness of operative interventions[5,27] by comparing changes in PRO scores between the pre- and postoperative time points to the MID or as an endpoint in clinical effectiveness research.[5,23] The MID can provide insight into how patients' PRO values are expected to change over time and can be compared between subgroups.[6,16]

Two studies to date have estimated the MID of the AOS total score among ESAA patients. One study estimated the MID of the AOS total score among ESAA patients undergoing surgery to be 28.0 ± 17.9 at 2 years' follow-up.[8] Another reported the MID as 12.35 at an average follow-up of 5 years postoperation.[44] Although estimated at different points in time, the disparity between the MIDs is large, a 50% difference, though no evidence of a time-related "decay" in the MID has been reported, and the contrast leaves clinicians and interventionists without a definitive approach to reconciling the MID for AA or TAR.

There is little evidence indicating whether the variability in MID estimates can be attributed to subgroup differences or study-specific contextual characteristics.[37] This insight is striking given that others have recommended that instrument values of AA and TAR patients be analyzed separately, the surgeries have different rates of revision and may also be subject to gender, age, or etiology of arthritis differences.[10,17,20,25,45,46] The use of a single estimation of MID may result in inappropriate conclusions regarding interventions or misleading patient advice.[10] For example, if older patients had a much lower MID than younger patients, they may be denied surgery that would benefit them if a universal MID was used.

To improve understanding of what constitutes a meaningful change in AOS score among operatively treated ESAA patients, the objective of this study was to use 4 different approaches to calculate MIDs for the AOS and its subscales among surgery (TAR or AA), sex, and age subgroups of patients being treated operatively for end-stage ankle arthritis. The study also estimates the MID of the AOS among preoperative health subgroups. We aimed to compare MID estimates across the 4 methods and test for MID differences between patient subgroups. We hypothesized that MID values would vary between patient subgroups and that the small effect size would provide the smallest MID estimates while the medium effect size would produce MID values similar to those obtained using anchor-based approaches.

Methods

Patient Recruitment and PRO Collection

This study is based on analysis of a prospectively recruited longitudinal cohort of patients enrolled in the Vancouver site of the Canadian Orthopaedic Foot and Ankle Society (COFAS) study who underwent elective ankle arthrodesis or ankle arthroplasty for the treatment of end-stage ankle arthritis in Vancouver, Canada. Patients were recruited into the study between March 6, 2002, and December 6, 2013. This study is based on follow-up data collected until August 9, 2017.

Prospective participants had been treated unsuccessfully with nonoperative management. To be eligible, patients had to be capable of responding to the survey questions in English (with or without assistance). Exclusion criteria were Charcot arthropathy, osteonecrosis of talus, previous ankle arthrodesis or replacement surgeries, and current or prior infection.[17,46]

Participants completed a survey package derived from the Musculoskeletal Outcomes Data Evaluation and Management System (MODEMS) questionnaire from the American Academy of Orthopaedic Surgeons. Participants completed this survey package preoperatively (baseline), 6 months postoperatively, and annually thereafter.[15] Surveys included the 36-Item Short Form Health Survey (SF-36) and the AOS, as well as questions about demographics, comorbidities, expectations of the operation, and outcome satisfaction.[8,15] As part of the baseline surveys, patients recorded their demographics, including sex, age at surgery, body mass index, history of smoking, diabetes, inflammatory arthritis, and surgery type. Each patient completed 1 survey package per ankle, reflecting that each ankle was individually treated. Ankles missing baseline PROs were excluded and removed from the data. This study required a minimum follow-up period of 2 years, including a complete response to the anchor question at 2 years. Additional study details can be found in previous publications.[17,46] Two-sample *t* tests, adjusted for multiple comparisons, were used to test for demographic differences between surgery subgroups and between sex subgroups.

Participating patients provided written informed consent for study enrollment prior to data collection. The study is approved by the Research Ethics Board at St. Paul's Hospital, Vancouver, Canada.

Patient-Reported Outcome Measures

MODEMS. The MODEMS package was designed to improve quality of care for patients with musculoskeletal disorders and contains validated PROs, including the SF-36, the AOS, and an expectation and satisfaction scale.[49] The expectation and satisfaction scale is a 6-item instrument; its earlier version that used value-based responses has demonstrated adequate internal consistency (Cronbach alpha = 0.71) and test-retest reliability (Cohen kappa = 0.91) among orthopedic surgery populations.[49]

AOS. The AOS[19] is an 18-item PRO instrument consisting of two 9-item subscales that measure ankle-related pain and disability, separately. To complete each item, patients placed a mark along a 100-mm horizontal line bounded on the left by "No pain" and to the right by "Worst pain imaginable" for the pain scale and "No difficulty" to "So difficult unable" for the disability scale. Each subscale produced a numeric score between 0 and 100, and an overall score was calculated as the average of the 2 subscale scores.[19] Higher scores indicate greater pain, disability, or overall ankle-related health, respectively.[19] The AOS has exhibited strong psychometric properties, including construct validity,[19,29] test-retest reliability,[19] and negligible ceiling and floor effects.[43] It is widely used in research specific to operative treatment of ESAA.[14,36,41]

Analysis

There are 2 commonly used approaches to estimating MID values: distribution- and anchor-based methods. Distribution-based approaches to calculating the MID express observed change in the form of a standardized metric (ie, the MID) based solely on the distribution of observed scores in a sample.[37] The effect size method is one distribution-based approach that relies entirely on the SD of the baseline data.[37] Using this method, which is based on Cohen *d* statistic,[9,38] the MID is taken as either 0.2 or 0.5 times the SD of baseline scores.[11] In this article, we refer to the former as the *small* effect size method and the latter as the *medium* effect size method.

Anchor-based approaches to estimate the MID utilize an external indicator that identifies patients whose health has changed by a small but meaningful degree and that has a nontrivial association with the PRO of interest.[37] In this study, the anchor question was found within the expectation and satisfaction scale of the MODEMS package (question 36),[13,49] which queries respondents on their expectations of the operation and satisfaction with the outcome.[48] Preoperatively, the item asks, "What results do you expect from your treatment?" The responses range from "not at all likely" to "extremely likely." Postoperatively, the question asks, "Are the results of your treatment what you expected in terms of relief from symptoms (pain, stiffness, swelling, numbness, weakness, instability)?" The responses are "definitely yes," "probably yes," "not sure," "probably not," "definitely not," and "not applicable." The responses can be treated as categorical and are also recorded as numeric values measuring from 1 ("Not at all likely"/"Definitely yes") to 6 ("Not applicable"). Distributions of item responses were assessed for floor or ceiling effects.

Pearson correlation between the anchor question was found to exceed 0.36 for the AOS overall score and its 2 subscales, indicating that the expectation question met the recommended correlation of a suitable anchor.[22,37] In addition, previous research has used this anchor question to estimate the MID of the AOS total score, classifying the "probably yes" and "probably no" subgroups, at the postoperative timepoint, as the minimally changed group, and taking the mean of the absolute value of the AOS score change as the MID estimate.[8] The current study took the same approach to estimating the MID[8] to ensure the results of this study were comparable to existing research.

A multiple linear-regression method was employed as an alternative anchor-based approach.[7,28,31] This regression-based method takes the slope of the linear relation between the anchor (independent variable) and PRO score of interest (dependent variable) as the MID and enables adjustment for confounding variables. In this study, we fit a multiple linear regression model with the change in MODEMS expectation/satisfaction question score as the independent variable and the change in AOS score as the dependent variable, adjusting for age at surgery, smoking status, and baseline AOS scores, which are associated with a change in AOS scores.[8] The model slope corresponding to the anchor was identified as the MID value for the respective AOS subscale or total score.[3,24,33]

To compare MID values between subgroups, a bootstrap resampling method, with 1000 samples with replacement, was applied to obtain estimates of the standard error of the MID for each surgery, sex, age category, and preoperative (baseline) subgroup. Two methods were used to test the hypothesis that MID values differed between patient subgroups. First, 2-sample *t* tests were used to compare the MIDs between AA and TAR, as well as between males and

Table 1. Summary of Participant Demographics, Stratified by Surgery Type and Sex.

Characteristic	All (n = 298; 100%)	AA (n = 111; 37.2%)	TAR (n = 187; 62.8%)	*P* value	Male (n = 166; 55.7%)	Female (n = 132; 44.3%)	*P* value
Age at surgery, mean (SD)	62.0 (11.2)	57.2 (12.6)	64.8 (9.1)	< .001	62.6 (10.9)	61.2 (11.5)	.63
Age, n (%)							
≤50	43 (14.4)	31 (27.9)	12 (6.4)	–	20 (12.0)	23 (17.4)	–
51-60	87 (29.2)	37 (33.3)	50 (26.7)	–	47 (28.3)	40 (30.3)	–
61-70	87 (29.2)	23 (20.7)	64 (34.2)	–	49 (29.5)	38 (28.8)	–
≥70	81 (27.2)	20 (18.0)	61 (32.6)	–	50 (30.1)	31 (23.5)	–
Body mass index, mean (SD)	28.1 (5.03)	28.8 (5.38)	27.7 (4.77)	.90	27.6 (4.21)	28.7 (5.87)	.95
Diabetic, n (%)	25 (8.4)	11 (9.9)	14 (7.5)	–	27 (10.2)	8 (6.1)	–
Inflammatory arthritis, n (%)	58 (19.5)	15 (13.5)	43 (23.0)	–	17 (10.2)	41 (36.1)	–
Smoking history, n (%)							
Currently smoking/quit <12 mo ago	37 (12.5)	17 (15.4)	20 (10.7)	–	27 (15.1)	12 (9.1)	–
Never smoked/quit (for >12 mo)	261 (87.6)	94 (84.6)	167 (89.3)	–	141 (84.9)	120 (90.9)	–

Abbreviations: AA, ankle arthrodesis; TAR, total ankle replacement.

females. A 1-way analysis of variance (ANOVA) was used to test for differences in MID values between age categories and baseline AOS categories, each with more than 2 subgroups. When the results of ANOVA indicated a significant difference between age categories, pairwise *t* tests were used to conduct post hoc analyses to determine which MIDs significantly differed from each other. We visually assessed the frequency distributions of the MID statistics for normality. To adjust for multiple comparisons, we used an alpha of 0.08 as a threshold for reporting pairwise comparisons in post hoc testing, applying Bonferroni correction to our original alpha of 0.05.

The analyses were conducted using R studio, version 1.1.463, and SAS 9.4. For statistical tests, alpha was set at 0.05.

Results

The final sample contained 283 participants totalling 298 unique ankles. Overall and subgroup samples did not reveal significant departures from normality,[30] allowing for the use of parametric statistical analyses, although nonparametric bootstrap resampling methods were used to test for differences between MID values. Substantial floor and ceiling effects were not observed.

Patients' demographic characteristics and differences between subgroups are presented (Table 1). In the sample, more patients underwent TAR (63%) than AA (37%). The mean age of all participants was 61.97 years. TAR patients were older, on average, than those who underwent AA ($P < .001$). Categorizing patients by age revealed different age distributions between subgroups, with the majority of TAR patients being 61 years or older (66.8%). In contrast, the majority of AA patients were 60 years or younger (61.2%). More than one-fifth of TAR patients (23%) had a history of inflammatory arthritis, whereas this condition was only present in 13.5% of AA patients. There were no observed differences in age, body mass index, diabetes diagnosis, smoking status, or smoking history between male and female patients.

A summary of participants' preoperative and 2-year postoperative follow-up scores of the AOS are presented in Appendix 1; at 2 years post-surgery, the difference between preoperative and postoperative AOS total score means (±SD) was 29.19 ± 25.78.

The MID estimates, using the distribution- and anchor-based approaches, for the overall cohort as well as each surgery and sex subgroup are shown (Table 2). The analysis found no statistically significant differences in the 2-year MIDs between AA and TAR or between males and females using the 4 approaches. The MID for the AOS total score, determined using the anchor-based linear regression approach, was 5.81.

There were some statistically significant differences between age categories' MID estimates. After adjusting for

Table 2. MID (Standard Error) of the AOS Scales and Instrument Total, Stratified by Operative Procedure and Sex.

		Surgery			Sex		
MID calculation	Overall	AA	TAR	*P* value	Male	Female	*P* value
Distribution-based approach							
Effect size, small, 0.2 × SD							
AOS total	3.04 (0.13)	3.03 (0.20)	3.05 (0.17)	.987	3.08 (0.18)	2.89 (0.18)	.751
Pain	3.29 (0.13)	3.21 (0.21)	3.34 (0.18)	.844	3.18 (0.19)	3.35 (0.21)	.787
Disability	3.36 (0.13)	3.51 (0.23)	3.28 (0.17)	.712	3.53 (0.18)	3.02 (0.20)	.403
Effect size, medium, 0.5 × SD							
AOS total	7.61 (0.32)	7.60 (0.50)	7.62 (0.43)	.979	7.70 (0.45)	7.23 (0.46)	.616
Pain	8.23 (0.34)	8.03 (0.52)	8.34 (0.46)	.755	7.95 (0.47)	8.38 (0.53)	.669
Disability	8.40 (0.34)	8.03 (0.56)	8.19 (0.42)	.560	8.83 (0.45)	7.55 (0.49)	.187
Anchor-based approach							
Linear regression							
AOS total	5.81 (0.72)	6.64 (0.74)	5.05 (1.03)	.232	5.47 (1.05)	6.41 (0.99)	.510
Pain	5.42 (0.70)	6.22 (0.82)	4.65 (0.95)	.240	5.05 (0.97)	6.00 (1.00)	.497
Disability	6.22 (0.79)	7.13 (0.86)	5.44 (1.22)	.243	5.96 (1.22)	6.82 (1.17)	.575
Mean absolute change							
AOS total	25.30 (1.58)	26.19 (2.67)	24.81 (2.11)	.529	25.26 (2.00)	25.23 (2.53)	.952
Pain	25.70 (1.66)	28.23 (2.93)	24.27 (1.91)	.074	25.79 (2.15)	25.58 (2.68)	.926
Disability	26.70 (1.78)	24.89 (2.35)	27.72 (2.32)	.193	27.51 (2.20)	25.71 (2.81)	.424

Abbreviations: AOS, Ankle Osteoarthritis Scale; AA, ankle arthrodesis; MID, minimal important difference; TAR, total ankle replacement.

Table 3. MID Estimates (Standard Error) for All AOS Scales and the Total Score, Stratified by Age Category.

MID calculation method	Overall	Age ≤50 y	Age 51-60 y	Age 61-69 y	Age ≥70 y	*P* value
Distribution-based approach						
Effect-size, small, 0.2 × SD						
AOS total	3.04 (0.13)	2.64 (0.29)	2.99 (0.22)	3.13 (0.23)	3.25 (0.29)	.517
Pain	3.29 (0.13)	2.76 (0.29)	3.27 (0.26)	3.28 (0.23)	3.61 (0.30)	.304
Disability	3.36 (0.13)	3.27 (0.30)	3.27 (0.22)	3.48 (0.25)	3.41 (0.31)	.925
Effect-size, medium, 0.5 × SD						
AOS total	7.61 (0.32)	6.59 (0.73)	7.47 (0.55)	7.82 (0.58)	8.12 (0.73)	.517
Pain	8.23 (0.34)	6.91 (0.74)	8.18 (0.65)	8.20 (0.57)	9.03 (0.75)	.304
Disability	8.40 (0.34)	8.16 (0.75)	8.17 (0.54)	8.71 (0.63)	8.53 (0.78)	.925
Anchor-based approach						
Linear-regression						
AOS total	5.81 (0.72)	5.22 (0.90)	6.09 (0.89)	7.75 (0.74)	3.44 (1.21)	.012
Pain	5.42 (0.70)	4.87 (0.84)	5.16 (0.89)	8.22 (0.69)	2.27 (1.08)	<.001
Disability	6.22 (0.79)	5.56 (1.01)	7.01 (0.98)	7.28 (1.06)	4.62 (1.45)	.310
Mean absolute change						
AOS total	25.30 (1.58)	25.36 (4.14)	24.32 (2.50)	25.80 (2.99)	25.92 (3.44)	.980
Pain	25.70 (1.66)	30.31 (3.79)	24.58 (2.68)	23.53 (2.87)	26.44 (3.78)	.580
Disability	26.70 (1.78)	22.22 (4.07)	25.77 (2.63)	29.05 (3.61)	28.13 (3.93)	.623

Abbreviations: AOS, Ankle Osteoarthritis Scale; MID, minimal important difference.

pairwise comparisons, there were differences in MID between the oldest patients and younger patients; the oldest patients had the smallest MID in most comparisons (Table 3). The pairwise comparisons of the AOS subscales are shown in Appendices 2, 3, and 4.

There were statistically significant differences in MID estimates between preoperative AOS value categories (Table 4). Depending on preoperative AOS value, the MID differed for each of the approaches. For the linear regression method, the post hoc comparisons indicated that the MID was statistically significantly larger (7.09) among those who scored between 51 and 60 than among patients whose AOS value was less than 40. Using the mean absolute change method, the MID was statistically

Table 4. MID Estimates (Standard Error) for All AOS Total Score, Stratified by Preoperative (Baseline) AOS Value, and the Pairwise Comparison of MID Estimates Stratified by Baseline Score Categories.[a]

MID calculation method	Overall	Score ≤40 (n = 66)	Score 41-50 (n = 84)	Score 51-60 (n = 80)	Score >60 (n = 68)	*P* value
Distribution-based						
Effect size, small 0.2 × SD	3.04 (0.13)	1.76 (0.14)	0.60 (0.03)	0.61 (0.03)	1.55 (0.15)	<.001
Effect size, medium 0.5 × SD	7.61 (0.32)	4.40 (0.35)	1.50 (0.07)	1.52 (0.07)	3.87 (0.36)	<.001
Anchor-based						
Linear regression	5.81 (0.72)	3.87 (0.78)	5.49 (0.92)	7.09 (0.83)	6.14 (1.36)	.148
Mean absolute change	25.30 (1.58)	17.70 (1.80)	21.63 (2.21)	24.46 (2.87)	39.09 (4.18)	<.001
Pairwise comparisons						
Distribution-based, 0.2 effect size						
Score ≤40			Y	Y	N	
Score 40-50			–	N	Y	
Score 50-60			–	–	Y	
Distribution-based, 0.5 effect size						
Score ≤40			Y	Y	N	
Score 40-50			–	N	Y	
Score 50-60			–	–	Y	
Anchor-based (linear regression)						
Score ≤40			N	Y	N	
Score 40-50			–	N	N	
Score 50-60			–	–	N	
Anchor-based (mean absolute change)						
Score ≤40			Y	N	Y	
Score 40-50			–	N	Y	
Score 50-60			–	–	Y	

Abbreviations: AOS, Ankle Osteoarthritis Scale; MID, minimal important difference; N, no; Y, yes.
[a]Pairwise tests are modified with *t* test. Pairs with significant *P* value are marked with Y.

significantly higher among the group of patients who scored higher than 60 on the AOS preoperatively than all lower scoring subgroups. These results indicate that the MID of the AOS total score varies between patient groups with different preoperative AOS scores, with the MID typically highest among the best functioning (AOS score less than 40) and worse functioning patients (AOS score greater than 60), preoperatively.

Discussion

The purpose of this study was to calculate the MID of the AOS and its 2 subscales among patients who underwent operative treatment for ESAA. Validated MIDs are important to clinicians and other researchers because they provide an independent end point for evaluating interventions. It is recommended that minimally important differences reflect population and context differences[37]; however, published MIDs for the AOS have not reflected differences among subgroups. This study contributes a deeper understanding of how the MID differs by surgery or patient characteristics for the 2-year end point by reporting and comparing MID values across treatment, sex, age, and preoperative AOS score categories.

Overall, our sample showed improvement in ankle-related health, indicated by a decrease in mean AOS total score at 2 years postsurgery of 29.2 ± 25.8. The improvement observed in our sample is consistent with measures in the existing literature. Another study that also measured outcomes at 2 years postsurgery reported a mean improvement in AOS total score of 31.2 ± 22.7.[8] A similar level of improvement has been reported at 6 months postsurgery, represented by an AOS score change of 28.8,[36] as well as among both TAR and AA subgroups in a sample with a mean follow-up time of 5.5 years postoperation.[14]

This study found that the 2-year MID in AOS total score was 25.3, standard error of 1.6, using the mean absolute change method, which corresponds to a standardized effect size of 0.92. This value is consistent with the 2-year MID reported elsewhere (28.0 ± 17.9) using the absolute mean change method.[8] Using the linear regression anchor-based method and the 2 distribution-based approaches, we observed much smaller MIDs for the AOS total score (ranging from 3.1 to 7.6).

The study's findings present a conundrum; the MID for the overall instrument varies appreciably by the approach used to calculate its value: distribution- and anchor-based methods can yield very different estimates of the MID. The challenge of reconciling the different values is not unique to the AOS, and the literature supports triangulating results from multiple methods.[4,37,47]

Distribution-based approaches have been criticized for not capturing meaningfulness from a patient perspective[37] and do not always measure the "minimal" change perceived as meaningful to patients, but rather provide a mathematical estimate of how far away from the mean is likely clinically meaningful. On the other hand, it has been postulated that half an SD (ie, the medium-effect size) yields a universal estimate of the MID, supported by some empirical and psychophysiological evidence.[26,37] Distribution-based methods are also acknowledged for generating a metric that is easy too for clinicians to understand.[37]

In the present study, the estimates of the MID obtained using the anchor-based absolute mean change method were much larger than calculated using the other approaches. This method may articulate the significant difference that the patient perceives as they move in either direction from "not sure" to "probably yes" or "probably no." In contrast, the distribution-based methods, also restricted by AOS scoring, are a function of baseline variation transformed by a factor of 0.2 or 0.5; the lower MID estimates produced using these methods reflect the low variability in baseline AOS scores in our sample and are a function of the methods themselves. However, it is recognized that different methods will yield a range of MID estimates; some researchers suggest that anchor-based methods be assigned the most weight while drawing on experience from clinical trials and conceptual understanding of the relationship between anchor and PRO of interest to narrow down the range of MID values.[37]

Based on recommendations in the literature,[26,37] and given the estimated effect size in this sample, we suggest placing the most weight on the MID estimates for the AOS obtained using the anchor-based linear regression method and recommend the MID total score estimate to be 5.81. It is reassuring that the small and medium effect size distribution-based methods produced MID values on either side of the linear regression estimate since the reputed strongest method yielded estimates within the overall range of MID values.

Regarding subgroup differences in MID values, the analyses did not find evidence that the MID varies by AA or TAR and sex, though our results suggest that a different MID may be needed for older patients. This finding is predominantly reassuring because it provides some evidence that the MID is robust to surgery and sex, though some additional care is needed for evaluating the effectiveness of surgery among older patients.

Importantly, the results support that the MID varies by baseline ankle-related health; the MID of the AOS total score was highest among patients with the worst and best health. This is inconsistent with previous research investigating foot-related health in a sample that underwent bunion correction surgery; the study identified that smaller improvements in functioning were meaningful for those with higher preoperative functioning than those with worse preoperative functioning.[16] These discrepancies may, however, be clinical in nature, owing to the different challenges that bunions and end-stage arthritis present to patients. A possible explanation for our finding is that patients with the worst health have higher expectations for their outcomes and perceive only full symptom resolution as operative success because they are admitted to surgery with the most severe pain and disability.

In spite of our robust sample and thorough analyses, the present study is not without limitations. Similar to other research on MID, this study may be affected by selection bias, because patients with worse health are more likely to return their PROs.[44] Additionally, although the MODEMS expectation and satisfaction question is the only anchor relevant to ankle outcome instruments, a disease-specific anchor may be more suitable for measuring the MID of the AOS, but no alternative resources have yet been published.[48,49] Moreover, these results are limited by the reliability of the items under study. An additional limitation of the present study is that we did not calculate an a priori population size for this study. The findings of this study point to future work that is needed to explore the interaction of the effects of age and operative groups.

Conclusions

Based on analysis of a prospective cohort of patients who underwent TAR or AA for ESAA, the evidence supports that differences in MID values for the AOS likely exist for different age groups and patient groups with varying levels of preoperative ankle-related health. No effects on MID were found for sex or operative subgroup. Future research can explore the etiology of subgroup differences in AOS MID values.

Authorship

Each author has participated in the design of the study, has contributed to the collection of the data, has participated in the interpretation of the results, writing of the manuscript, and assumes full responsibility for the content of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Canadian Institutes for Health Research (CIHR), Strategy for Patient-Oriented Research, Methods Cluster and in-kind support of Vancouver Coastal Health (VCH) Authority. The last author is a scholar of the Michael Smith Foundation for Health Research (MSFHR). CIHR, VCH, and MSFHR had no role in developing the methods, data analyses, interpreting the results or manuscript preparation.

ORCID iDs

Jason M. Sutherland, PhD, https://orcid.org/0000-0002-1857-2432

Carmela Melina Albanese, BSc, https://orcid.org/0000-0003-3310-380X

Murray Penner, MD, https://orcid.org/0000-0003-1349-2510

Supplemental Material

Supplementary material is available online with this article.

References

1. Adukia V, Mangwani J, Issac R, Hussain S, Parker L. Current concepts in the management of ankle arthritis. *J Clin Orthop Trauma*. 2020;11(3):388-398.
2. Agel J, Coetzee JC, Sangeorzan BJ, Roberts MM, Hansen ST. Functional limitations of patients with end-stage ankle arthrosis. *Foot Ankle Int*. 2005;26(7):537-539.
3. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol*. 2017;82:128-136.
4. Bell ML, Dhillon HM, Bray VJ, Vardy JL. Important differences and meaningful changes for the Functional Assessment of Cancer Therapy-Cognitive Function (FACT-Cog). *J Patient Rep Outcomes*. 2018;2(1):48.
5. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
6. Çelik D, Çoban Ö, Kılıçoğlu Ö. Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *J Clin Epidemiol*. 2019;113:44-57.
7. Clement ND, MacDonald D, Simpson AHRW. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sport Traumatol Arthrosc*. 2014;22(8):1933-1939.
8. Coe MP, Sutherland JM, Penner MJ, Younger A, Wing KJ. Minimal clinically important difference and the effect of clinical variables on the ankle osteoarthritis scale in surgically treated end-stage ankle arthritis. *J Bone Joint Surg Am*. 2015;97(10):818-823.
9. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
10. Cook CE. Clinimetrics corner: the Minimal Clinically Important Change Score (MCID): a necessary pretense. *J Man Manip Ther*. 2008;16(4):82E-83E.
11. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7(5):541-546.
12. Courville XF, Hecht PJ, Tosteson ANA. Is total ankle arthroplasty a cost-effective alternative to ankle fusion ? *Clin Orthop Relat Res*. 2011;469:1721-1727.
13. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American Spine Society lumbar spine outcome assessment instrument: reliability and validity tests. *Spine (Phila Pa 1976)*. 1996;21(6):741-749.
14. Daniels TR, Younger ASE, Penner M, et al. Intermediate-term results of total ankle replacement and ankle arthrodesis: a COFAS multicenter study. *J Bone Joint Surg*. 2014;96:135-142.
15. Daum WJ, Brinker MR, Nash DB. Quality and outcome determination in health care and orthopaedics: evolution and current structure. *J Am Acad Orthop Surg*. 2000;8(2):133-139.
16. Desai S, Peterson AC, Wing K, et al. Minimally important difference in the foot and ankle outcome score among patients undergoing hallux valgus surgery. *Foot Ankle Int*. 2019;40(6):694-701.
17. Desai S, Sutherland JM, Younger A, et al. Role of postoperative patient-reported outcomes to predict ankle arthroplasty and arthrodesis revision. *Foot Ankle Int*. 2020;41(1):10-16.
18. DiDomenico LA, Gatalyak N. End-stage ankle arthritis. arthrodiastasis, supramalleolar osteotomy, or arthrodesis? *Clin Podiatr Med Surg*. 2012;29(3):391-412.
19. Domsic RT, Saltzman CL. Ankle osteoarthritis scale. *Foot Ankle Int*. 1998;19(7):466-471.
20. Gill GS, Casey Chan K, Mills DM. 5- to 18-year follow-up study of cemented total knee arthroplasty for patients 55 years old or younger. *J Arthroplasty*. 1997;12(1):49-54.
21. Glazebrook M, Daniels T, Younger A, et al. Comparison of health-related quality of life between patients with end-stage ankle and hip arthrosis. *J Bone Joint Surg Am*. 2008;90(3):499-505.
22. Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77(4):371-383.
23. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life—a systematic review. *J Clin Epidemiol*. 2017;89:188-198.
24. Jayadevappa R, Malkowicz SB, Wittink M, Wein AJ, Chhatre S. Comparison of distribution-and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Serv Res*. 2012;47(5):1902-1925.
25. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. *BMJ Open*. 2015;5(10):e007953.
26. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184.
27. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Educ*. 2017;17(4):137-144.
28. Lee WC, Kwan YH, Chong HC, Yeo SJ. The minimal clinically important difference for Knee Society Clinical Rating System after total knee arthroplasty for primary osteoarthritis. *Knee Surg Sport Traumatol Arthrosc*. 2017;25(11):3354-3359.

29. Liu G, Peterson AC, Wing K, et al. Validation of the ankle osteoarthritis scale instrument for preoperative evaluation of end-stage ankle arthritis patients using item response theory. *Foot Ankle Int*. 2019;40(4):422-429.
30. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*. 2002;23:151-169.
31. Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. *Knee Surg Relat Res*. 2020;32(1):19.
32. Morash J, Walton DM, Glazebrook M. Ankle arthrodesis versus total ankle arthroplasty. *Foot Ankle Clin*. 2017;22: 251-266.
33. Musoro ZJ, Hamel J-F, Ediebah DE, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *BMJ Open*. 2018;8:19117.
34. Nwachukwu BU, McLawhorn AS, Simon MS, et al. Management of end-stage ankle arthritis: cost- utility analysis using direct and indirect costs. *J Bone Joint Surg*. 2015;97:1159-1172.
35. Raikin SM, Rasouli MR, Espandar R, Maltenfort MG. Trends in treatment of advanced ankle arthropathy by total ankle replacement or ankle fusion. *Foot Ankle Int*. 2014;35(3):216-224.
36. Rajapakshe S, Sutherland JM, Wing K, et al. Health and quality of life outcomes among patients undergoing surgery for end-stage ankle arthritis. *Foot Ankle Int*. 2019;40(10): 1129-1139.
37. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-109.
38. Sawilowsky SS. New effect size rules of thumb. *J Mod Appl Stat Methods*. 2009;8(2):597-599.
39. Segal AD, Shofer J, Hahn ME, Orendurff MS, Ledoux WR, Sangeorzan BJ. Functional limitations associated with end-stage ankle arthritis. *J Bone Joint Surg*. 2012;94(9):777-783.
40. Southerland JT, Boberg JS, Downey M, Nakra A, Rabjon LV. *McGlamry's Comprehensive Textbook of Foot and Ankle Surgery*. 4th ed. Philadelphia, PA: Wolters Kluwer Health; 2013.
41. Sutherland JM, Rajapakshe S, Younger A, et al. Do preoperative patient-reported outcomes predict hospital length of stay for surgically-treated end-stage ankle osteoarthritis patients? *Foot Ankle Surg*. 2020;26(2):175-180.
42. Thomas RH, Daniels TR. Ankle arthritis. *J Bone Joint Surg Am*. 2003;85(5):923-936.
43. Veltman ES, Hofstad CJ, Witteveen AGH. Are current foot- and ankle outcome measures appropriate for the evaluation of treatment for osteoarthritis of the ankle? Evaluation of ceiling effects in foot- and ankle outcome measures. *Foot Ankle Surg*. 2017;23(3):168-172.
44. Waly FJ, Yeo EMN, Wing KJ, Penner MJ, Veljkovic A, Younger ASE. Relationship of preoperative patient-reported outcome measures (PROMs) to postoperative success in end-stage ankle arthritis. *Foot Ankle Int*. 2020;41(3):253-258.
45. Williams DP, Price AJ, Beard DJ, et al. The effects of age on patient-reported outcome measures in total knee replacements. *Bone Joint J*. 2013;95-B(1):38-44.
46. Wing KJ, Chapinal N, Coe MP, et al. Measuring the operative treatment effect in end-stage ankle arthritis: are we asking the right questions? A COFAS multicenter study. *Foot Ankle Int*. 2017;38(10):1064-1069.
47. Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. *Value Health*. 2005;8(2):117-127.
48. Younger ASE, Wing KJ, Glazebrook M, et al. Patient expectation and satisfaction as measures of operative outcome in end-stage ankle arthritis. *Foot Ankle Int*. 2015;36(2):123-134.
49. Zywiel MG, Mahomed A, Gandhi R, Perruccio AV, Mahomed NN. Measuring expectations in orthopaedic surgery: a systematic review. *Clin Orthop Relat Res*. 2013;471(11): 3446-3456.