AMERICAN ORTHOPAEDIC FOOT & ANKLE SOCIETY

Foot & Ankle International®
2014, Vol. 35(10) 1031–1036

DOI: 10.1177/1071100714543647
fai.sagepub.com

Comparison of the Self-Reported Foot and Ankle Score (SEFAS) and the American Orthopedic Foot and Ankle Society Score (AOFAS)

Maria C. Cöster, MD[1,2], Björn E. Rosengren, MD[1], Ann Bremander, PhD[3], Lars Brudin, MD[4], and Magnus K. Karlsson, MD[1]

Abstract

Background: The Self-reported Foot and Ankle Score (SEFAS) is a patient-reported outcome measure, while the American Orthopedic Foot and Ankle Society Score (AOFAS) is a clinician-based score, both used for evaluation of foot and ankle disorders. The purpose of this study was to compare the psychometric properties of these 2 scoring systems.

Methods: A total of 95 patients with great toe disorders and 111 patients with ankle or hindfoot disorders completed the 2 scores before and after surgery. We evaluated time to complete the scores in seconds, correlations between scores with Spearman's correlation coefficient (r_s), floor and ceiling effects by proportion of individuals who reached the minimum or maximum values, test–retest reliability and interobserver reliability by intraclass correlation coefficient (ICC), internal consistency by Cronbach's coefficient alpha (CA), and responsiveness by effect size (ES). Data are provided as correlation coefficients, means, and standard deviations.

Results: SEFAS was completed 3 times faster than AOFAS. The scores correlated with an r_s of .49 for great toe disorders and .67 for ankle/hindfoot disorders (both $P < .001$). None of the scores had any floor or ceiling effect. SEFAS test–retest ICC values measured 1 week apart were .89 for great toe and .92 for ankle/hindfoot disorders, while the corresponding ICC values for AOFAS were .57 and .75. AOFAS interobserver reliability ICC values were .70 for great toe and .81 for ankle/hindfoot disorders. SEFAS CA values were .85 for great toe and .86 for ankle/hindfoot disorders, while the corresponding CA values for AOFAS were .15 and .42. SEFAS ES values were 1.15 for great toe and 1.39 for ankle/hindfoot disorders, while the corresponding ES values for AOFAS were 1.05 and 1.73.

Conclusion: As SEFAS showed similar or better outcome in our tests and was completed 3 times faster than AOFAS, we recommend SEFAS for evaluation of patients with foot and ankle disorders.

Level of Evidence: Level II, prospective comparative study.

Keywords: ankle, AOFAS, comparison, foot, hindfoot, reliability, responsiveness, SEFAS, score, validity

Foot and ankle pain affects around 20% of the middle aged and old population.[3,4] Many of these patients also require surgery. In Sweden, with 9.5 million inhabitants, 20 000 elective operative procedures were annually performed in the foot and ankle between 2007 and 2009, without any structured evaluation of patients or results. This ought to be done through national registries by the use of foot and ankle-specific scores.[4,20,26] The Self-Reported Foot and Ankle Score (SEFAS), a patient-reported outcome measure (PROM), has been validated with good results in both patients with forefoot and ankle/hindfoot disorders.[6,7] However, the most widespread foot and ankle-specific score is the American Orthopedic Foot and Ankle Society Score (AOFAS)[16] and SEFAS ought therefore to be evaluated also in relation to AOFAS. AOFAS is, in contrast to SEFAS, only partially validated, and that is 1 of the reasons why the score has been criticized.[12,14,18,19,26,28,31] The score also demands a clinical examination, making it very

[1]Departments of Orthopedics and Clinical Sciences, Lund University, Skåne University Hospital Malmö, Sweden
[2]Hand & Foot Surgery Center, Stockholm, Sweden
[3]Departments of Rheumatology and Clinical Sciences, Lund University, Lund, Sweden
[4]Department of Clinical Physiology, Kalmar Hospital, Kalmar and Department of Medicine and Health Sciences, University Hospital Linköping, Linköping, Sweden

Corresponding Author:
Maria C. Cöster, MD, Department of Orthopedics and Clinical Sciences, Lund University, Skåne University Hospital, Malmö S-205 02, Sweden.
Email: maria.coster@med.lu.se

difficult to use in large-scale registers. The purpose of this study was to compare SEFAS and AOFAS by use of psychometric properties in terms of reliability, validity, and responsiveness following international guidelines.[24,25,32]

Methods

Self-Reported Foot and Ankle Score (SEFAS)

SEFAS (www.swedankle.se) is a foot- and ankle-specific patient-reported score based on the New Zealand Total Ankle Questionnaire.[13] The scoring system has been described in detail previously.[6,7] The score contains 12 questions, each with 5 response options that score from 0 to 4 points. A total SEFAS scoring sum of 0 represents the most severe disability and 48 normal function. The score covers different constructs such as pain and function, not reported in separate subscales. SEFAS has been validated with good psychometric properties in patients with osteoarthritis (OA) and inflammatory disease in the ankle joint,[7] and in patients with a variety of disorders in the forefoot, midfoot, ankle, and hindfoot.[6] For missing answers we used the following approach: (1) when results from 2 or more questions were missing, the questionnaire was disregarded; (2) when the result from 1 question was missing, the mean result of the remaining 11 questions was used. Out of the 312 questionnaires in this study we found 1 missing answer in 11 (3.5%) questionnaires, while none had 2 missing answers.

American Orthopedic Foot and Ankle Society Score (AOFAS)

The AOFAS was developed by a committee of the American Orthopedic Foot and Ankle Society and introduced by Kitaoka et al in 1994.[16] The AOFAS includes 4 different scores, each related to a specified anatomic region in the foot or ankle: (1) the ankle/hindfoot (A-HF) scale, (2) the midfoot (MF) scale, (3) the hallux metatarsophalangeal/interphalangeal (HMTP) scale, and (4) the lesser metatarsophalangeal/interphalangeal (LMTP) scale.[16] The scores are dependent on both patient-reported questions about pain, activity, functional limitations, and footwear and examiner-reported data about alignment, gait and motion within 3 subscales: (1) pain, (2) function, and (3) alignment. Each of the 4 scores includes a clinical examination and contains 8 to 9 different questions with 3 to 4 response options where each question is rated between 0 and a maximum that ranges from 5 and 40 depending on the specific question. Thus, the questions are weighted differently, with the subscale pain including only 1 question with a possible rating between 0 and 40 points, indicating that pain is strongly weighted in AOFAS. A total scoring sum of 0 points represents the most severe disability and 100 normal function. Even though the AOFAS is commonly used, it is only partly validated.[31]

Table 1. General and Anthropometric Data for Participants.

	Patients With Disorders in the Great Toe (n = 95)	Patients With Disorders in the Ankle/Hindfoot (n = 111)
Age (years)		
Median (range)	52 (18-78)	56 (24-81)
Gender (n, %)		
Male	15 (16)	49 (44)
Female	80 (84)	62 (56)
Height (cm)		
Mean ± SD	168 ± 8.6	171 ± 10.1
Weight (kg)		
Mean ± SD	74 ± 13.7	84 ± 15.7
Diagnosis (n, %)		
Arthritis	2 (2)	28 (25)
Achilles tendon disorders	0 (0)	14 (13)
Planovalgus	0 (0)	28 (25)
Cavovarus/neurological	0 (0)	20 (18)
Great toe disorders	89 (94)	0 (0)
Others	4 (4)	21(19)
Operative procedures (n, %)		
Arthrodesis	9 (9)	33 (30)
Calcaneal osteotomy	0 (0)	32 (29)
Tendon surgery	0 (0)	17 (15)
Osteotomy first metatarsal	70 (74)	0 (0)
Tendon transfers	0 (0)	9 (8)
Others	16 (17)	20 (18)

Furthermore, we could not find a general consensus on how to use the AOFAS when answers are incomplete. We decided to use the following approach in cases of incomplete questionnaires: (1) if the questionnaire included more than 2 missing questions, we disregarded the questionnaire; (2) if the questionnaire included 1 or 2 missing questions we rated these questions with the mean result of the remaining questions. Out of the 278 questionnaires in this study we found 1 missing answer in 13 (4.7%) questionnaires while none had 2 missing answers.

Subjects

We asked in this preplanned study all patients with no exclusion criteria with disorders in the great toe, hindfoot or ankle or, who were scheduled for foot or ankle surgery at the orthopedic departments in Kalmar or Eksjö in Sweden during the period January 1, 2011, to September 30, 2013, to participate. There were 80 women and 15 men with a median age of 52 (range, 18-78) years with disorders in the great toe, and 62 women and 49 men with a median age of 56 (range, 24-81) years with disorders in the ankle or hindfoot who accepted to participate. Patient background data are presented in Table 1. The patients completed the SEFAS score

without help and the AOFAS with help from specially trained physiotherapists. The physiotherapists also conducted the AOFAS examiner-dependent evaluation. In this study we used the HMTP score and the A-HF score due to the patient selection.[16] All tests were not conducted in all patients. In a random sample, that included 18 patients with ankle/hindfoot disorders, we registered the time to complete the SEFAS and in 22 patients time to complete the AOFAS (both the patient-reported and examiner-reported parts). In 44 patients with great toe disorders and in 62 with ankle/hindfoot disorders we provided SEFAS twice before surgery (a week apart) to test reliability and agreement. The same was done for AOFAS in 32 patients with great toe disorders and in 40 with ankle/hindfoot disorders. In 12 patients with great toe disorders and in 27 with ankle/hindfoot disorders 2 physiotherapists conducted separate AOFAS scores during the same day to evaluate interobserver reliability. In 53 patients with great toe disorders and 74 with ankle/hindfoot disorders we provided SEFAS just before surgery and 6 months after surgery to test responsiveness. The same was done for AOFAS in 49 patients with great toe disorders and in 70 patients with ankle/hindfoot disorders.

Statistics

Statistical calculations were performed with Statistical Package for the Social Sciences (SPSS) software version 17.0 (IBM Software Statistics®, Armonk, NY) and STATISTICA version 10.0 (Statsoft Inc®, Tulsa, OK).We evaluated correlations between the scores by Spearman's correlation coefficient with 95% confidence intervals (95% CIs) according to Fisher's *z*-transformation. *R* values >.60 were considered strong correlations, .30-.60 moderate, and <.30 weak.[32] We considered unfavorable floor or ceiling effects to be present if more than 15% of the individuals reached the highest or lowest score.[32,35] We evaluated test–retest reliability and for AOFAS the interobserver reliability by intraclass correlation coefficient (ICC) and considered an ICC values > .70 as acceptable.[30] We evaluated intraindividual absolute variability as a measure of agreement (measurement error) by standard error of a single determination ($S_{method} = \sqrt{(\Sigma d_i^2 / (2n))}$, where d_i is the difference between the *i*th paired measurement and n is the number of differences. S_{method} expresses the measurement error in scoring points. The relative measurement error by coefficient of variation (CV %; SD/mean) was also calculated.[8] We evaluated the internal consistency (an estimate of how the questions within a score are correlated to each other) by Cronbach's alpha (CA) and considered CA values > .70 as acceptable.[30] We also evaluated responsiveness (the ability of a score to detect changes for example after surgery) by calculating effect sizes (ESs) and standardized response means (SRMs) with 95% CIs according to the method described by Becker.[3] We considered ES values > 0.80 as strong, 0.50-0.80 as moderate, ≥0.2 and <0.5 as small, and <0.2 as trivial.[5]

Informed written consent was obtained from all participants. The study was approved by the ethics committee of Lund University, Sweden (2009/698) and was performed according to the Declaration of Helsinki.

Results

Patients completed SEFAS in mean 3 times faster than AOFAS, ranging from 1 minute and 12 seconds to 6 minutes and 51 seconds for SEFAS and 5 minutes and 37 seconds to 14 minutes and 35 seconds for AOFAS (Table 2).

The therapist first filled in the answers from the patient from the patient-reported part of the questionnaire and immediately after that made the clinical examination and filled in the results. When also these results were filled in the measured time was completed. Summarized the complete AOFAS was completed in the times above.

SEFAS and AOFAS correlated significantly with an *r* value of .49 in patients with great toe disorders and .67 in patients with hindfoot/ankle disorders (Table 2). None of the patients reached maximum or minimum numeric values in SEFAS or AOFAS, that is, there were no floor or ceiling effects (Table 2).

SEFAS had in patients with great toe disorders an ICC value of .89 (.81, .94) and in patients with ankle/hindfoot disorders a value of .92 (.87, .95), while the corresponding ICC values for AOFAS were .57 (.29, .77) and .75 (.58, .86), respectively (Table 2). The interobserver reliability for AOFAS, estimated by ICC, was in patients with great toe disorders .82 (.50, .94) and in patients with ankle/hindfoot disorders .71 (.46, .86) (Table 2).

For SEFAS the measurement error (measure of agreement) estimated by S_{method} was 2.5 scoring points and the CV 8.3% in patients with great toe disorders and in patients with hindfoot/ankle disorders the corresponding values were 2.4 and 11.7%. For AOFAS the corresponding values were 11.2 scoring points and 18.8% in patients with great toe disorders and 9.4 scoring points and 19.2% in patients with hindfoot/ankle disorders (Table 2). SEFAS had CA value of .86 in patients with great toe disorders and of .85 in patients with ankle/hindfoot disorders, while the corresponding CA values for AOFAS were .15 and .42, respectively (Table 2).

For SEFAS the ES and SRM (measure of responsiveness) values were 1.39 (1.01, 1.77) and 1.40 (1.01, 1.78) in patients with great toe disorders and 1.15 (0.85, 1.45) and 0.97 (0.69, 1.26) in patients with ankle/hindfoot disorders. For the AOFAS the corresponding values were 1.73 (1.26, 2.20) and 1.53 (1.09, 1.97) in patients with great toe disorders and 1.05 (0.76, 1.35) and 0.94 (0.65, 1.23) in patients with ankle/hindfoot disorders.

Table 2. Validity, Reliability, Agreement, and Responsiveness of Self-Reported Foot and Ankle Score (SEFAS) and American Orthopedic Foot and Ankle Society Score (AOFAS) in Patients With Great Toe Disorders and Ankle/Hindfoot Disorders.

	Patients With Disorders in the Great Toe		Patients With Disorders in the Ankle/Hindfoot	
	SEFAS	AOFAS	SEFAS	AOFAS
Time to complete the scores (n)			18	22
Seconds (mean ± SD)	—	—	160 ± 82	515 ± 138
Correlations between scores (n)	95	95	111	111
Spearman rho (95% CI)	.49 (.31, .67)		.67 (.53, .81)	
Floor and ceiling effects (n)	95	95	111	111
Proportion (%)	0	0	0	0
Test–retest reliability (n)	44	32	62	40
Test (mean ± SD)	29.7 ± 7.6	61.4 ± 17.6	20.5 ± 8.6	47.5 ± 19.2
Retest (mean ± SD)	30.3 ± 7.7	58.4 ± 16.7	21.0 ± 8.5	50.5 ± 18.2
Intraclass correlation coefficient (mean, 95% CI)	.89 (.81, .94)	.57 (.29, .77)	.92 (.87, .95)	.75 (.58, .86)
Interobserver reliability (n)		12		27
Intraclass correlation coefficient (mean, 95% CI)	—	.82 (.50, .94)	—	.71 (.46, .86)
Agreement (n)	44	32	62	40
S_{method} (scoring points)	2.5	11.2	2.4	9.4
Coefficient of variation (%)	8.3	18.8	11.7	19.2
Internal consistency (n)	92	92	107	105
Cronbach's coefficient α (mean)	.86	.15	.85	.42
Responsiveness (n)	53	49	74	70
Preoperative (mean ± SD)	27.3 ± 7.9	51.7 ± 13.8	20.3 ± 7.9	47.0 ± 17.7
Postoperative (mean ± SD)	38. 2 ± 7.6	75.5 ± 13.4	29.5 ± 9.9	65.6 ± 20.6
Effect size (mean, 95% CI)	1.39 (1.01, 1.77)	1.73 (1.26, 2.20)	1.15 (0.85, 1.45)	1.05 (0.76, 1.35)
Standardized response mean (mean, 95% CI)	1.40 (1.01, 1.78)	1.53 (1.09, 1.97)	0.97 (0.69, 1.26)	0.94 (0.65, 1.23)

Data are presented as numbers of individuals who were included in the analyses (n), means ± standard deviations, means with 95% confidence intervals (CIs), means for S_{method} and Cronbach's coefficient alpha, and proportions in percentage floor and ceiling effects and coefficient of variation.

Discussion

SEFAS had more advantageous psychometric properties than AOFAS in patients with great toe and ankle/hindfoot disorders. The most evident discrepancies were found in the reliability tests, where SEFAS showed a better repeatability and lower measurement error than AOFAS. The facts that SEFAS, in addition to being a strict PROM, could be completed in mean 3 times faster than AOFAS and that SEFAS in contrast to AOFAS does not demand any clinical examination also speak in favor of SEFAS.

Several scores are in use for evaluating the outcome of foot and ankle surgery,[4,11,14,20,21,26,31] but none has been accepted as the gold standard. Among these, AOFAS is the most widespread even though it is not a PROM and has not been sufficiently validated.[17,31] Some researchers have therefore inferred that validated scores should replace the AOFAS.[14] The beneficial evaluation in this study as well as the favorable outcomes in previous reports[6,7] indicate that SEFAS could be a candidate.

It has been suggested that all new musculoskeletal outcome scores should be validated against the Short-Form 36 questionnaire (SF-36).[36] For SEFAS we have found strong correlations with SF-36 subscales bodily pain (BP) and physical function (PF), in patients with both forefoot and ankle/hindfoot disorders.[6,7] In contrast, AOFAS had in similar analyses weaker correlations with SF-36, actually weaker compared with scores used in knee, shoulder, and upper extremity evaluations.[19,27,29] This knowledge makes the usefulness of AOFAS questionable.

Floor and ceiling effect is also an important construct, since a score must have possibility to capture changes in the clinical situation. We found in this study that neither SEFAS nor AOFAS had any floor or ceiling effects, indicating that both scores are ideal scores in this aspect.

SEFAS is, in contrast to AOFAS, a strict PROM and does therefore not require any clinical examination. Inclusion of clinician-based evaluation always introduces the risk of intra- and interobserver variability. The small-scale studies that have evaluated this for AOFAS have

provided varying results,[2,4,31] while in this study, with a larger sample size than in previous reports, we found significantly lower test–retest ICC in AOFAS than SEFAS, unacceptably low in patients with great toe disorders. This is another finding that speaks in favor of SEFAS.

SEFAS also had smaller measurement errors than AOFAS when evaluating agreement (Table 2). Even if the interobserver reliability analyses for AOFAS showed acceptable values (Table 2), we speculate that the measuring error will inevitably increase when data are provided by 2 persons (the patient and the examiner) than when, as in SEFAS, including only patient-reported data. This view is partly supported in the literature, where reports have found that the repeatability of tests deteriorates when including not only PROMs but also examiner dependent evaluations.[1,15]

An ideal score should not include several questions that capture the same deficit since this only results in unnecessary work without providing additional information. Our study showed that the internal consistency, as evaluated by CA, was greater for SEFAS than AOFAS (Table 2). AOFAS has also previously, in patients with hallux valgus, been reported with lower internal consistency than the Manchester-Oxford Foot Questionnaire and the SF-36 score.[10] Thus, the internal consistency analysis also speaks in favor of SEFAS in comparison with AOFAS. However, the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) checklist suggests that this analysis may not be relevant for all type of scores and maybe AOFAS could be one such score.[22,23,33]

Another most important construct is the ability of a score to identify if a patient improves or deteriorates after surgery. Previous studies have documented that region-specific scores provide better responsiveness than generic scores in patients with foot and ankle disorders.[9,10,18,29] This view is supported by our study with beneficial responsiveness analyses for both SEFAS and AOFAS compared to generic scores.

There are more concerns with AOFAS that we did not evaluate. Madeley et al discussed the mathematical shortcoming of a linear score that runs from 0 to 100 but without possibility to reach some specific sums such as 1, 98, or 99.[18] Madeley et al also discussed the limited precision due to the small number of response intervals available for each question while other questions, such as the only 1 dealing with pain, could contribute up to 40% of the total scoring sum.[12] These concerns could result in a skewed behavior of the score, a fact that limits the precision and makes parametric testing unreliable.[12] Another problem with AOFAS is that the score introduces a problem when evaluating some specific outcomes such as an arthrodesis due to OA. The goal of this procedure is to create a stiff joint without pain. Even though this goal is fulfilled, AOFAS will still score low values due to the objective lack of motion in the joint. Finally, the fact that 4 different scores, based on anatomical region, must be used, together with the need of an examiner, makes the use of the AOFAS more resource-demanding than SEFAS and unsuitable for large-scale evaluation for registries.

The strengths of this study include the structural evaluation of psychometric properties that follow international guidelines when evaluating both SEFAS and AOFAS, the inclusion of a variety of diagnoses and operative procedures, and the inclusion of disorders in different anatomical regions in the foot and ankle. Weaknesses include the sample size in that it had been advantageous to be able to conduct gender-specific subgroup analyses in patients with 1 diagnosis, with disorders in 1 anatomical region, and when doing 1 specific operative procedure. It would also have been advantageous to follow responsiveness for more than 6 months and include more foot and ankle scores in the validation process.

Conclusion

We conclude that both SEFAS and AOFAS could be used for evaluating patients with foot and ankle disorders, but since SEFAS is a strict PROM and thus could be used in large-scale registries, has better psychometric properties, and is time-saving and resource-saving compared to AOFAS, we prefer and recommend SEFAS for evaluating patients with foot and ankle disorders. We now plan to use and evaluate SEFAS in a national foot and ankle registry and find international partners who will translate and validate the SEFAS in other languages.

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.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Baertschi E, Swanenburg J, Brunner F, Kool J. Interrater reliability of clinical tests to evaluate scapulothoracic motion. *BMC Musculoskelet Disord*. 2013;14:315.
2. Baumhauer JF, Nawoczenski DA, DiGiovanni BF, Wilding GE. Reliability and validity of the American Orthopaedic Foot and Ankle Society Clinical Rating Scale: a pilot study for the hallux and lesser toes. *Foot Ankle Int*. 2006;27:1014-1019.
3. Becker BJ. Synthesizing standardized mean change measures. *Br J Math Stat Psychol*. 1988;41:257-278.
4. Button G, Pinney S. A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system? *Foot Ankle Int*. 2004;25:521-525.

5. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press; 1978.
6. Coster M, Bremander A, Rosengren BE, et al. Validity, reliability, and responsiveness of the Self-reported Foot and Ankle Score (SEFAS) in forefoot, hindfoot, and ankle disorders. *Acta Orthop*. 2014;85:187-194.
7. Coster M, Karlsson MK, Nilsson JA, Carlsson A. Validity, reliability, and responsiveness of a self-reported foot and ankle score (SEFAS). *Acta Orthop*. 2012;83:197-203.
8. Dahlberg G. *Statistical Methods for Medical and Biological Students*. 2nd ed. London, UK: George Allen and Unwin; 1940.
9. Dawson J, Boller I, Doll H, et al. Responsiveness of the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS, SF-36 and EQ-5D assessments following foot or ankle surgery. *J Bone Joint Surg Br*. 2012;94:215-221.
10. Dawson J, Doll H, Coffey J, Jenkinson C. Responsiveness and minimally important change for the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cartilage*. 2007;15:918-931.
11. Farrugia P, Goldstein C, Petrisor BA. Measuring foot and ankle injury outcomes: common scales and checklists. *Injury*. 2011;42:276-280.
12. Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. *Foot Ankle Int*. 2001;22:779-787.
13. Hosman AH, Mason RB, Hobbs T, Rothwell AG. A New Zealand national joint registry review of 202 total ankle replacements followed for up to 6 years. *Acta Orthop*. 2007;78:584-591.
14. Hunt KJ, Hurwit D. Use of patient-reported outcome measures in foot and ankle research. *J Bone Joint Surg Am*. 2013;95:e118(111-119).
15. Juul T, Langberg H, Enoch F, Sogaard K. The intra- and inter-rater reliability of five clinical muscle performance tests in patients with and without neck pain. *BMC Musculoskelet Disord*. 2013;14:339.
16. Kitaoka HB, Alexander IJ, Adelaar RS, et al. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int*. 1994;15:349-353.
17. Lau JT, Mahomed NM, Schon LC. Results of an Internet survey determining the most frequently used ankle scores by AOFAS members. *Foot Ankle Int*. 2005;26:479-482.
18. Madeley NJ, Wing KJ, Topliss C, et al. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. *Foot Ankle Int*. 2012;33:57-63.
19. Malviya A, Makwana N, Laing P. Correlation of the AOFAS scores with a generic health QUALY score in foot and ankle surgery. *Foot Ankle Int*. 2007;28:494-498.
20. Martin RL, Irrgang JJ, Lalonde KA, Conti S. Current concepts review: foot and ankle outcome instruments. *Foot Ankle Int*. 2006;27:383-390.
21. Mayich DJ. Life without the AOFAS score: a critical problem in a time of scarcity. *Foot Ankle Int*. 2012;33:689-691.
22. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Method*. 2010;10:22.
23. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: consensus-based standards for the selection of health measurement instruments. *BMC Med Res Method*. 2006;6:2.
24. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539-549.
25. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745.
26. Naal FD, Impellizzeri FM, Rippstein PF. Which are the most frequently used outcome instruments in studies on total ankle arthroplasty? *Clin Orthop Relat Res*. 2010;468:815-826.
27. Pena F, Agel J, Coetzee JC. Comparison of the MFA to the AOFAS outcome tool in a population undergoing total ankle replacement. *Foot Ankle Int*. 2011;28:788-793.
28. SooHoo NF, Shuler M, Fleming LL. Evaluation of the validity of the AOFAS Clinical Rating Systems by correlation to the SF-36. *Foot Ankle Int*. 2003;24:50-55.
29. SooHoo NF, Vyas R, Samimi D. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. *Foot Ankle Int*. 2006;27:930-934.
30. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. New York, NY: Oxford University Press; 2008.
31. Suk M. *Musculoskeletal Outcomes Measures and Instruments*. Davos, Switzerland: AO Foundation; 2009.
32. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42.
33. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2011;21:651-657.
34. Thomas MJ, Roddy E, Zhang W, et al. The population prevalence of foot and ankle pain in middle and old age: a systematic review. *Pain*. 2011;152:2870-2880.
35. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score: do ceiling effects limit its usefulness in orthopedics? *Acta Orthop*. 2010;81:703-707.
36. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-448.