

Foot & Ankle International

<http://fai.sagepub.com/>

Responsiveness and Validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in End Stage Ankle Arthritis

N. Jane Madeley, Kevin J. Wing, Claire Topliss, Murray J. Penner, Mark A. Glazebrook and Alastair SE Younger

Foot Ankle Int 2012 33: 57
DOI: 10.3113/FAI.2012.0057

The online version of this article can be found at:
<http://fai.sagepub.com/content/33/1/57>

Published by:



<http://www.sagepublications.com>

On behalf of:



[American Orthopaedic Foot & Ankle Society](#)

Additional services and information for *Foot & Ankle International* can be found at:

Email Alerts: <http://fai.sagepub.com/cgi/alerts>

Subscriptions: <http://fai.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Jan 1, 2012

[What is This?](#)

Responsiveness and Validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in End Stage Ankle Arthritis

N. Jane Madeley, FRCS(Tr&Orth); Kevin J. Wing, MD, FRCSC; Claire Topliss, FRCS(Tr&Orth); Murray J. Penner, MD, FRCSC; Mark A. Glazebrook, MD, FRCSC; Alastair SE Younger, MD, FRCSC
Glasgow, UK

ABSTRACT

Background: We examined four commonly used scores, the SF-36, the Ankle Osteoarthritis Scale (AOS), the American Orthopaedic Foot and Ankle Society (AOFAS) Ankle Hindfoot Score, and the Foot Function Index (FFI) to determine their responsiveness and validity. **Methods:** Patients with end stage ankle arthritis were recruited into a prospective multicenter cohort study and baseline and one year outcome scores were compared. The Standardized Response Mean and Effect Size for the AOS, AOFAS, and FFI were calculated and the three region- or disease-specific scores were compared with the SF-36 to determine their criterion validity. **Results:** All four scores showed acceptable responsiveness, and when using the validated SF-36 as the standard the three region or disease specific scores all showed similar criterion validity. **Conclusion:** All four scores are responsive and can be considered for use in this population. The objective component of the AOFAS Ankle Hindfoot Score may make it harder to perform than the other three scores which have subjective components only, and as yet its objective component has not been shown to demonstrate reliability. We recommend use of a purely subjective score such as the Ankle Osteoarthritis Scale or Foot Function Index as the region- or disease-specific score of choice in this population. As the SF-36 shows acceptable responsiveness, using it alone could also be considered.

Although none of the authors has received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article, benefits have been or will be received but are directed solely to a research fund, foundation, educational institution, or other nonprofit organization with which one or more of the authors is associated.

Corresponding Author:

N. Jane Madeley FRCS Orth.
Consultant Orthopaedic Surgeon
Department of Orthopaedics, Glasgow Royal Infirmary.
Castle Street
Glasgow G4 0SF
UK
E-mail: janemadeley@yahoo.co.uk

For pricing and availability of reprints, e-mail reprints@datatrace.com or call 410-494-4994 x232.

Level of Evidence: II, Prospective Comparative Study

Key Words: Endstage Ankle Arthritis; Outcome Score

INTRODUCTION

In the foot and ankle literature, at least 49 outcomes scoring tools have been reported,⁵ with 14 specific to the ankle and hindfoot. The ideal score would be relevant, reliable, valid, and responsive; it should collect information pertinent to evaluating the outcome of treatment, provide reproducible results, measure what it is designed to measure and be sensitive enough to detect clinically significant improvement or deterioration. The most commonly used outcomes scores in foot and ankle are the AOFAS clinical rating scales,¹⁶ which have not been thoroughly validated or tested for reliability.

Martin et al. conducted a survey of self-reported outcome instruments for the foot and ankle and found 14 scores.¹⁸ The Foot Function Index showed content and construct validity, reliability, and responsiveness in rheumatoid arthritis, the disease for which it was designed, but only reliability in other foot and ankle conditions. The Ankle Osteoarthritis Scale showed construct validity and reliability. A meta-analysis of foot and ankle outcome measures by Button and Pinney found that the AOFAS Ankle Hindfoot Score was neither valid, responsive, nor reliable, whereas both the SF-36 and the Foot Function Index were both valid and reliable but not proven to be responsive.⁵ Collection of multiple scores is onerous on the patients and includes repetition.

The Medical Outcomes Study Short Form 36, referred to most commonly as the SF-36, is the most widely used generic score and has been extensively tested for reliability and validity through a wide range of conditions. It comprises 11 main questions subdivided into 36 items. The responses are then used to generate eight subscales: physical functioning (PF); role-physical (RP); bodily pain (BP); and general health (GH), which are used to generate a physical component

summary scale (PCS); and vitality (VT); social functioning (SF); role-emotional (RE); and mental health (MH), which generate a mental component summary scale (MCS). A linear transformation is performed to produce norm-based scoring based on the general population in the U.S. which achieves a mean of 50 and standard deviation of 10 for all subscales and the physical and mental summary measures, with higher scores representing a better state of health.²⁴⁻²⁷ While designed as a quality of life measure, comparisons of scores before and after treatment are used as an outcome measure.⁹

In 1994, the American Orthopaedic Foot and Ankle Society (AOFAS) published a series of four foot and ankle clinical rating scores, one of which is the Ankle Hindfoot Scale (AHS).¹³ It comprises both subjective, or patient-reported items, and objective, or physician-assessed items, and is scored from 0 to 100, with a higher score representing a better outcome.

Although an Internet survey of AOFAS members by Lau et al.¹⁶ found it to be the most widely employed scoring system in the foot and ankle, it has not been satisfactorily tested for reliability or validity.

The Foot Function Index (FFI) was developed in 1991 as a patient reported instrument to measure the impact of foot pathology on function in a population with Rheumatoid Arthritis.⁴ It consists of a series of 23 visual analogue scales, nine related to pain, nine related to difficulties, and five related to patient limitations. It was fully validated by its authors, and subsequently shown to be a reliable tool.¹ Following revision, it remains valid and reliable.³

The Ankle Osteoarthritis Scale (AOS) was derived from the Foot Function Index. It maintains the visual analogue scale structure but dispenses with the five questions on limitations, and modifies anatomic descriptors for the difficulty and pain subscales.⁸ The authors who developed this scale determined that it was reliable and valid in ankle osteoarthritis patients.

These four scores were originally selected for use in this study for the following reasons: the SF-36 is a generic health assessment tool allowing comparison of the impact of end stage ankle arthrosis with other factors and pathologies expected to affect patients' quality of life, the AOFAS Scores are the most commonly used by orthopaedic surgeons and cited the most in foot and ankle literature,^{5,16,21} the FFI is a validated score,⁴ and the AOS is the only outcome score designed specifically for end stage ankle arthritis (EAA) and was included in order to compare its performance to the other three more established scores. No other score was validated at the time of initiation of the study and, hence, were not included.

The purpose of this study was to examine these four scores, compare their responsiveness, and use the SF-36 as a standard to look for criterion validity of the three disease- or region-specific scores in order to decide which score or combination of scores is optimal for assessing outcome in these patients.

MATERIALS AND METHODS

Outcome scores

The Canadian Orthopaedic Foot and Ankle Society conducts an ongoing multicenter study designed to examine the clinical outcome of the surgical treatment of EAA. Subjects were recruited through consultation with one of the nine orthopaedic surgeons. Ethics approval has been obtained at all study centers and informed consent was received from all patients prior to administration of the outcome scores. Outcome scoring was undertaken on these patients using the SF-36, the AOS, the AOFAS AHS, and the FFI. For the purposes of this study, data from two of the COFAS centers was included.

Study group

Data was collected preoperatively, and at 6 months, 1 year, 2 years, and 5 years. Following a power calculation, we included data from all EAA patients who have undergone either total ankle replacement or ankle arthrodesis at two study sites, and have completed baseline and 1-year data. The cohort of patients with complete data at 2 and 5 years was not yet sufficiently large. The purpose of this study was not to compare the outcomes of different treatments but to assess the ability of the four scores to report outcome; therefore, the only exclusion criteria was patients having incomplete data. However, a single pathology, EAA, was selected to reduce variability. The SF-36, AOS, FFI, and the subjective components of the AOFAS AHS were collected via postal questionnaire. The patients were recalled for clinical assessment within 2 weeks of the postal questionnaire by a member of the research team to collect the objective data for the AOFAS Ankle Hindfoot Score. Passive range of sagittal plane and hindfoot motion was assessed using a goniometer. Results were recorded prospectively in a computerized database.

Demographics

The study group consisted of 117 patients. All patients who reached 1-year followup with complete data were included. The mean patient age was 59.7 (range, 27 to 84) years. Sixty-six (56.4%) of the patients were male. Sixty-four patients underwent ankle fusion and 53 underwent total ankle replacement.

Responsiveness

The responsiveness of an outcome score is its ability to detect clinical change between different time points. The responsiveness of the four scores were compared with each other. Three main comparisons were made: the pain components of all four scores were compared with one another; the difficulty subscales of the AOS and FFI were compared with the physical functioning subscale of the SF-36; and the total Physical Component Summary Score of the SF-36 was compared with the total AOS, AOFAS AHS

and FFI scores. Finally, the subjective-only components of the AOFAS Ankle Hindfoot Score were compared with the four total scores as previous authors have suggested use of this score without the objective measurements.¹¹ A t-test was used to look for statistically significant change between baseline and 1-year scores. In addition, the Standardized Response Mean (SRM) and the Effect Size (ES) were calculated. These are both measures of responsiveness which allow the mean change in score to be examined in the context of the standard deviation. Comparison with the standard deviation tells us whether the change in score following surgery is larger or smaller than the variability within the group, and thus whether a difference in score can be considered a true change. While many measures of responsiveness exist, these two were selected as both were in common usage pertaining to the responsiveness of outcome scores,²³ thus allowing comparison of our results with other studies.

The SRM was calculated as the mean change in score divided by the standard deviation of the mean score and the ES was calculated as the mean change in score divided by the standard deviation of the preoperative scores. Results of ± 0.2 to 0.5, 0.5 to 0.8, and greater than 0.8 were regarded as small, moderate, and large degrees of change for both the SRM and the ES, with the positive or negative indicating the direction of change.^{6,20} As the AOS and FFI both showed a drop in score with clinical improvement, we expected to see a negative SRM and ES, whereas positive values for the SF-36 and AHS would indicate clinical improvement and a positive result in SRM and ES would be expected. The magnitude of change of score was therefore illustrated in the results to allow comparison.

Validity

A valid score is one that actually measures what it is intended to measure. To determine criterion validity, comparison must be made to an existing measure or standard already held to be valid. In this case, the SF-36, although a generic measure, was selected as the standard because of its known rigorous testing. The Pearson correlation coefficient was used to measure the strength and direction of the

relationships between the four scores and determine whether any of the three region- or disease-specific scores showed validity when compared with the physical component scale of the SF-36. The correlation coefficient can have a value from -1 to $+1$ and the further it lies from zero the stronger the correlation. Results of ± 0 to 0.09, 0.1 to 0.3, 0.3 to 0.5, and 0.5 to 1 were interpreted to represent no, small, medium, and large correlation, respectively, using the boundaries described by Cohen.⁶

Reliability

Reliability, a measure of reproducibility of the results collected, was not evaluated in the current study.

A power calculation was performed (β -error = 0.2) and revealed 44 subjects were needed for significance ($p < 0.05$).

RESULTS

Responsiveness

There were large ESs and SRMs in all the subscales and total scores examined. When the pain subscales were examined, all four scores showed ESs and SRMs above the threshold of 0.8, indicating a high level of responsiveness. The greatest ES and SRM were seen for the pain component of the AOS, 1.61 and 1.23, respectively (Table 1). Similarly a comparison of the disability subscales alone also showed a large ES and SRM for all three scores with the greatest ES and SRM seen for the disability component of the Ankle Osteoarthritis Scale, 1.39 and 1.0, respectively (Table 2). When the total scores were examined, all four scores had large ESs and SRMs indicating a high level of responsiveness; however, the AOFAS AHS showed the greatest ES and SRM, 1.69 and 1.34, respectively. When the subjective components of the AOFAS AHS were examined alone these values dropped and were closer to those of the other three scores, 1.43 and 1.1 (Table 3). While higher values indicate a higher level of responsiveness, the significance of the magnitude of the ES or SRM beyond the 0.8 threshold has not yet been defined.

In Figure 1, the SRMs for the different outcome scores are shown. While the AOFAS total score shows the greatest

Table 1: Responsiveness of Pain Subscales

Pain components	Baseline scores*	1 year scores*	Change in score*	t-test	Standardized Response Mean change	Effect Size
SF-36 BP	32.5 \pm 7.9	42.8 \pm 10.2	10.2 \pm 10.7	$p = 0.00$	0.96	1.29
AOS Pain	53.4 \pm 18.3	24.0 \pm 20.3	-29.4 \pm 23.9	$p < 0.001$	1.23	1.61
AHS Pain	11.9 \pm 11.4	26.2 \pm 10.4	35.7 \pm 14.0	$p < 0.001$	1.02	1.26
FFI Pain	52.1 \pm 20.5	25.1 \pm 21.8	-27.0 \pm 22.4	$p < 0.001$	0.94	1.32

*, Mean and standard deviation.

Table 2: Responsiveness of Disability Subscales

Function components	Baseline scores*	1-year scores*	Change in score*	t-test	Standardized Response Mean change	Effect Size
SF-36 PF	30.0 ± 7.9	40.1 ± 10.2	10.2 ± 10.7	<i>p</i> = 0.00	0.97	1.28
AOS Disability	63.1 ± 20.8	34.2 ± 24.0	-28.9 ± 28.8	<i>p</i> < 0.001	1.00	1.39
FFI Disability	60.1 ± 22.4	32.4 ± 24.3	-28.5 ± 30.7	<i>p</i> < 0.001	0.93	1.27

*, Mean and standard deviation.

Table 3: Responsiveness of Total Scores

Outcome score	Baseline scores*	1-year scores*	Change in score*	t-test	Standardized Response Mean Change	Effect Size
SF-36 PCS	30.1 ± 7.7	40.3 ± 10.7	10.2 ± 11.0	<i>p</i> < 0.001	0.93	1.32
AOS	57.6 ± 17.2	29.7 ± 23.7	-27.9 ± 25.8	<i>p</i> < 0.001	1.08	1.62
AOFAH AHS	35.6 ± 14.0	68.9 ± 19.8	33.3 ± 24.9	<i>p</i> < 0.001	1.34	1.69
AOFAH AHS (subjective only)	17.5 ± 14.3	36.9 ± 14.7	19.0 ± 18.6	<i>p</i> < 0.001	1.10	1.43
FFI	50.4 ± 18.1	25.6 ± 19.8	-24.8 ± 23.8	<i>p</i> < 0.001	1.04	1.37

*, Mean and standard deviation.

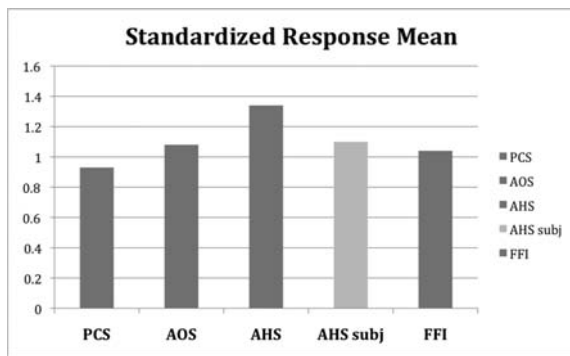


Fig. 1: SRMs of total scores.

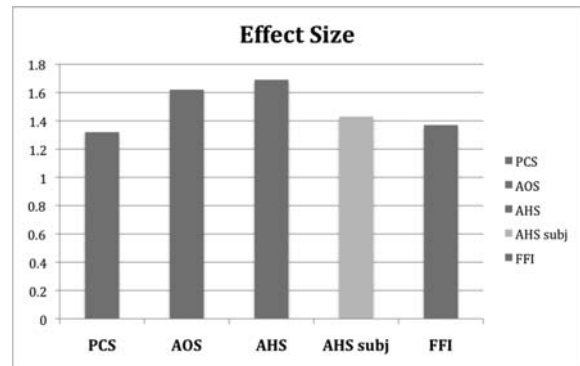


Fig. 2: ESs of total scores.

numerical value, it can be seen that each value lies above the threshold of 0.8 indicating a large SRM. Once this 0.8 threshold is reached, the significance of differences in value is undefined; however, we see that without its objective component the AOFAS AHS was very similar to the other subjective scores.

In Figure 2, the ESs for the different Outcome scores are shown. Again the AOFAS total score shows the highest numerical value, although the Ankle Osteoarthritis score has a similar value. Each score has a value above the 0.8 threshold indicating a large ES. Again, once this 0.8 threshold

reached the significance of differences in value, however, we see that without its objective component, the AOFAS AHS is very similar to the other subjective scores.

Validity

The three region- or disease-specific outcome tools all showed similar correlation coefficients when compared with the standard of the Physical Component Summary scale of the SF-36 score, the magnitude of which indicated a large correlation between scores.⁶

Table 4: Validity of Total Scores

	Correlation Coefficient	
SF-36 PCS and AOS	-0.59	$p < 0.0001$
SF-36 PCS and AHS	0.61	$p < 0.0001$
SF-36 PCS and FFI	-0.64	$p < 0.0001$

The AOS, FFI, and AHS were compared with the PCS component of the SF-36 using a correlation coefficient to examine their criterion validity.

DISCUSSION

The two main categories of outcome scoring are the generic instruments of general health status, and the disease or region specific tools. A generic health instrument measures the impact of a disease or injury on a patient's quality of life and allows comparison with other conditions or therapeutic interventions, but may not elicit small differences. A disease- or region-specific tool is intended to be more focused on the target condition, and therefore provides a more sensitive assessment of the severity of the condition being studied and the effect of any therapeutic measures employed. It is becoming accepted that a combination of both a generic health tool and a region- or disease-specific measure should be used in clinical studies.²⁸

Previous studies have used the Student t-test, SRM, and ES to measure responsiveness of outcomes scores, and correlation to examine criterion validity; therefore, we elected to use the same tools to allow comparison of our results with analyses of outcome scores already published in the literature.^{2,7,12,14,15,17,19,22,23}

SooHoo et al.²³ compared responsiveness of the FFI, AOFAS clinical rating systems, and SF-36. Their conclusions were that although the region-specific scores, FFI and AOFAS, showed higher responsiveness, the Physical Component Summary Scale and Bodily Pain subscale of the SF-36 were close enough to consider using the SF-36 alone to monitor foot and ankle outcomes. Their study included only 25 patients: 12 with forefoot pathology and 13 with ankle or hindfoot problems. They presented ES and SRM for the AOFAS of 1.12 and 1.10, and 0.68 and 0.76 for the Physical Component Summary Scale of the SF-36. They did not present the results for the total FFI for comparison but only for the three subscales. Their small numbers and mixture of disorders limited the significance of their findings.

Ibrahim et al.¹¹ compared the subjective component of the AOFAS scores to the Foot Function Index, and concluded that it had acceptable validity, reliability, and responsiveness. Again, this study included patients with a range of forefoot, midfoot, and hindfoot conditions. Our study

examined the performance of four scores in a large cohort of patients with a common pathology, EAA, allowing a clear comparison.

We found that all four scores examined showed a large ES and SRM. All of the scores exhibited acceptable responsiveness. In addition, when compared to the SF-36, a validated standard, all three of the region- or disease-specific tools showed a similar level of correlation and hence acceptable criterion validity. While a higher ES or SRM indicates a higher level of responsiveness, there is currently no data to confirm that beyond the threshold of 0.8 a higher value confirms a significant increase in responsiveness. The highest responsiveness was seen with the AOFAS AHS. This score relies on both subjective and objective components but its objective components have not been tested for intra- and inter-variability and its reliability is unknown. Without the objective component it does not show any greater responsiveness than the FFI or AOS.

The AOFAS scores are also mathematically problematic. The AHS ranges from 0 to 100 and is generally treated in the literature as a linear score; however, the scores of 1, 98, and 99 cannot be achieved. When only the subjective components are used, this is more apparent: the scores of 1,3,18,21,23,58 cannot be achieved over a range of 0 to 60. The heavy weighting towards the pain score and its large jumps in point values heavily influence the score's result. With 20 points allocated for moderate pain and 0 for severe pain, the total score is influenced heavily as the patient has to make their symptoms fit one of the suggested categories. This limits the scale's precision and may also skew results.¹⁰ In contrast, the use of visual analogue scales by the AOS and FFI do not constrain the patients' responses.

While the responsiveness and ability to report change for these scores has been calculated, the threshold at which a change in score becomes clinically significant has not yet been satisfactorily determined. It could be assumed that a change detected in a Visual Analogue Scale based outcome score, which does not constrain a patient's response, such as the FFI or AOS corresponds to clinically significant change; however, there is no data as yet to scientifically support this. The Minimal Clinically Important Difference (MCID) is one way in which this can be calculated. This correlates the patients' perception of their clinical state with the measured outcome score. There has been no reliable data published as yet regarding the MCID with regard to the AOFAS, FFI, or AOS. The SF-36 is a norm-based score and currently the SF-36 information website puts forward a change in the PCS of 2 points as a Minimal Important Difference; however, this value is not specific to conditions of the foot and ankle. A weakness of the current study is that our patients were not asked for their perception of their clinical condition at the time of scoring to allow calculation of the MCID for the outcome scores currently used, and this would be an interesting area for future research. Other areas for future research are to examine the subscales within the outcome

scores more closely to examine their responsiveness and validity and to look for floor and ceiling effects within these outcome scores.

CONCLUSION

Further research is required to improve our knowledge of the behavior of existing outcome scores in the foot and ankle in order to improve the current scores or develop new measures. However, in the current study, all four scores were responsive, as measured by large SRMs and ESs, and showed acceptable criterion validity in this patient population and use of any of these scores likely provides meaningful data. The theoretical ideal outcome score would contain both subjective and objective components;⁵ however, our experience has been that a subjective score is less time-consuming and has better patient compliance. The AOFAS score has both components, but as mentioned above, it has been criticized for its mathematical shortcomings and the reliability of the objective component has yet to be proven. As purely subjective scores, the AOS and FFI show acceptable responsiveness and less constraint of patient's response, and of the four scores studied, we recommend the use of one of these as the region- or disease-specific score of choice in conjunction with the SF-36 in the assessment of patients with EAA. As the Physical Summary Scale of the SF-36 showed high levels of responsiveness in these patients, the need for an additional region- or disease-specific score is not proven, and use of the SF-36 alone should be considered. Our study has concentrated on a patient population with arthritis of the ankle joint. Further study is needed to investigate whether our findings can be applied to the entire population of patients with foot and ankle pathology.

INVITED REVIEWER COMMENTARY

This paper takes an important look at trying to understand how we can better direct outcomes measures in foot and ankle surgery. Although the investigation would have been much more revealing had it also included nonoperative patients, addressed construct validity, and included reliability as means of determining the most appropriate foot and ankle instruments and defining their ideal applications, the information presented does serve to elevate a most apropos and as-yet-unanswered question in foot and ankle today: are we, or are we not, currently using the most accurate and scientific tool(s) necessary to assess how well we are caring for our patients? There is clearly room for improvement in the surveys and scales we now rely so heavily upon, and the authors should be applauded for their efforts in trying to move us one step closer in our endeavor to both measure things correctly and understand what we are measuring.

REFERENCES

1. Agel, J; Coetzee, JC; Sangeorzan, BJ; Roberts, MM; Hansen, ST: Reliability of the Foot Function Index: A report of the AOFAS Outcomes Committee. *Foot Ankle Int.* **26(11)**:962–7, 2005.
2. Briggs, KK; Kocher, MS; Rodkey, WG; Steadman, JR: Reliability, validity, and responsiveness of the Lysholm knee score and Tegner activity scale for patients with meniscal injury of the knee. *J Bone Joint Surg Am.* **88(4)**:698–705, 2006. <http://dx.doi.org/10.2106/JBJS.E.00339>
3. Budiman-Mak, E; Conrad, K; Stuck, R; Matters, M: Theoretical model and Rasch analysis to develop a revised Foot Function Index. *Foot Ankle Int.* **27(7)**:519–27, 2006.
4. Budiman-Mak, E; Conrad, KJ; Roach, KE: The Foot Function Index: a measure of foot pain and disability. *J Clin Epidemiol.* **44(6)**:561–70, 1991. [http://dx.doi.org/10.1016/0895-4356\(91\)90220-4](http://dx.doi.org/10.1016/0895-4356(91)90220-4)
5. 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.* **25(8)**:521–5, 2004.
6. Cohen, J: *Statistical power analysis for the behavioural sciences.* Edited, Hilldale, Lawrence Erlbaum Associates, 1988.
7. Dawson, J; Doll, H; Boller, I; et al.: Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. *Qual Life Res.* **17(10)**:1257–67, 2008. <http://dx.doi.org/10.1007/s11136-008-9409-3>
8. Domsic, RT; Saltzman, CL: Ankle osteoarthritis scale. *Foot Ankle Int.* **19(7)**:466–71, 1998.
9. Garratt, A; Schmidt, L; Mackintosh, A; Fitzpatrick, R: Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ.* **324(7351)**:1417, 2002. <http://dx.doi.org/10.1136/bmj.324.7351.1417>
10. Guyton, GP: Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. *Foot Ankle Int.* **22(10)**:779–87, 2001.
11. Ibrahim, T; Beiri, A; Azzabi, M; et al.: Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. *J Foot Ankle Surg.* **46(2)**:65–74, 2007. <http://dx.doi.org/10.1053/j.jfas.2006.12.002>
12. Jaglal, S; Lakhani, Z; Schatzker, J: Reliability, validity, and responsiveness of the lower extremity measure for patients with a hip fracture. *J Bone Joint Surg Am.* **82-A(7)**:955–62, 2000.
13. Kitaoka, HB; Alexander, IJ; Adelaar, RS; et al.: Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int.* **15(7)**:349–53, 1994.
14. Kocher, MS; Horan, MP; Briggs, KK; et al.: Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am.* **87(9)**:2006–11, 2005. <http://dx.doi.org/10.2106/JBJS.C.01624>
15. Kocher, MS; Tucker, R; Briggs, KK: Relationship between subjective and objective assessment of outcomes after anterior cruciate ligament reconstruction. *J Knee Surg.* **18(1)**:73–81, 2005.
16. Lau, JT; Mahomed, NM; Schon, LC: Results of an Internet survey determining the most frequently used ankle scores by AOFAS members. *Foot Ankle Int.* **26(6)**:479–82, 2005.
17. Lingard, EA; Katz, JN; Wright, RJ; Wright, EA; Sledge, CB: Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am.* **83-A(12)**:1856–64, 2001.
18. Martin, RL; Irrgang, JJ: A survey of self-reported outcome instruments for the foot and ankle. *J Orthop Sports Phys Ther.* **37(2)**:72–84, 2007. <http://dx.doi.org/10.2519/jospt.2007.2403>
19. McCarthy, CJ; Oldham, JA: The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee. *Rheumatology (Oxford).* **43(4)**:514–7, 2004. <http://dx.doi.org/10.1093/rheumatology/keh081>

20. **McDowell, I:** *Measuring Health: A guide to rating scales and questionnaires*. Edited, Oxford, Oxford University Press, 2006.
21. **Naal, FD; Impellizzeri, FM; Rippstein, PF:** Which are the most frequently used outcome instruments in studies on total ankle arthroplasty? *Clin Orthop Relat Res.* **468(3):**815–26. <http://dx.doi.org/10.1007/s11999-009-1036-y>
22. **O'Connor, PA; Yeap, S; Noel, J; et al.:** Lisfranc injuries: patient- and physician-based functional outcomes. *Int Orthop.* **27(2):**98–102, 2003.
23. **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.* **27(11):**930–4, 2006.
24. **Ware, J, Jr; Kosinski, M; Keller, SD:** A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* **34(3):**220–33, 1996. <http://dx.doi.org/10.1097/00005650-199603000-00003>
25. **Ware, JE, Jr; John, E; Ware, Jr:** on health status and quality of life assessment and the next generation of outcomes measurement. Interview by Marcia Stevic and Katie Berry. *J Healthc Qual.* **21(5):**12–7, 1999. <http://dx.doi.org/10.1111/j.1945-1474.1999.tb00984.x>
26. **Ware, JE, Jr; Keller, SD; Hatoum, HT; Kong, SX:** The SF-36 Arthritis-Specific Health Index (ASHI): I. Development and cross-validation of scoring algorithms. *Med Care.* **37(5 Suppl):**MS40–50, 1999. <http://dx.doi.org/10.1097/00005650-199905001-00004>
27. **Ware, JE, Jr; Sherbourne, CD:** The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* **30(6):**473–83, 1992. <http://dx.doi.org/10.1097/00005650-199206000-00002>
28. **Zlowodzki, M; Bhandari, M:** Outcome measures and implications for sample-size calculations. *J Bone Joint Surg Am.* **91 Suppl 3:**35–40, 2009. <http://dx.doi.org/10.2106/JBJS.H.01602>