Original article

Analysis of the psychometric properties of the American Orthopaedic Foot and Ankle Society Score (AOFAS) in rheumatoid arthritis patients: application of the Rasch model

Cristiano Sena da Conceição [a,b,*], *Mansueto Gomes Neto* [a,c], *Anolino Costa Neto* [a], *Selena M.D. Mendes* [b], *Abrahão Fontes Baptista* [d], *Kátia Nunes Sá* [b]

[a] *Biofunction Department, Physical Therapy, Universidade Federal da Bahia, Salvador, BA, Brazil*
[b] *Medicine and Health Program, Escola Bahiana de Medicina e Saúde Pública, Salvador, BA, Brazil*
[c] *Medicine and Health Program, Universidade Federal da Bahia, Salvador, Bahia, Brazil*
[d] *Biointeraction Department, Universidade Federal da Bahia, Salvador, BA, Brazil*

ARTICLE INFO

ABSTRACT

Objective: To tested the reliability and validity of Aofas in a sample of rheumatoid arthritis patients.

Methods: The scale was applicable to rheumatoid arthritis patients, twice by the interviewer 1 and once by the interviewer 2. The Aofas was subjected to test–retest reliability analysis (with 20 Rheumatoid arthritis subjects). The psychometric properties were investigated using Rasch analysis on 33 Rheumatoid arthritis patients.

Results: Intra-Class Correlation Coefficient (ICC) were ($0.90 < ICC < 0.95$; $p < 0.001$) for intra-observer reliability and ($0.75 < ICC < 0.91$; $p < 0.001$) for inter-observer reliability. Subjects separation rates were 1.9 and 4.75 for the items, showing that patients fell into three ability levels, and the items were divided into six difficulties levels. The Rasch analysis showed that eight items was satisfactory. One erroneous item have been identified, showing percentages above the 5% allowed by the statistical model. Further Rasch modeling suggested revising the original item 8.

Conclusions: The results suggest that the Brazilian versions of Aofas exhibit adequate reliability, construct validity, response stability. These findings indicate that Aofas Ankle-Hindfoot scale presents a significant potential for clinical applicability in individuals with rheumatoid arthritis. Other studies in populations with other characteristics are now underway.

© 2015 Elsevier Editora Ltda. All rights reserved.

Análise das propriedades psicométricas do American Orthopaedic Foot and Ankle Society Score (Aofas) em pacientes com artrite reumatoide: aplicação do modelo Rasch

R E S U M O

Palavras-chave:
Artrite reumatoide
Incapacidade
Escala
Avaliação
Análise Rasch

Objetivo: Testar a confiabilidade e a validade do escore Aofas em uma amostra de pacientes com artrite reumatoide.

Métodos: A escala foi aplicada a pacientes com artrite reumatoide, duas vezes pelo entrevistador 1 e uma vez pelo entrevistador 2. O Aofas foi submetido a exame de confiabilidade teste-reteste (com 20 indivíduos com artrite reumatoide). As propriedades psicométricas foram investigadas pela análise Rasch em 33 pacientes com artrite reumatoide.

Resultados: O coeficiente de correlação intraclasse (CCI) foi de 0,90 < CCI < 0,95 (*p* < 0,001) para a confiabilidade intraexaminador e 0,75 < CCI < 0,91 (*p* < 0,001) para a confiabilidade interexaminador. O índice de separação dos indivíduos foi de 1,9 e 4,75 para os itens. Isso demonstra que os pacientes se dividiam em três níveis de habilidade e os itens foram divididos em seis níveis de dificuldades. A análise Rasch mostrou que oito itens foram satisfatórios. Foi identificado um item errôneo, que mostrou percentuais acima dos 5% permitidos pelo modelo estatístico. Além disso, o modelo Rasch sugeriu a revisão do item 8 original.

Conclusões: Os resultados sugerem que a versão brasileira do Aofas apresenta confiabilidade adequada, validade de constructo e estabilidade de resposta. Esses resultados indicam que a escala de tornozelo-retropé Aofas apresenta um potencial significativo de aplicabilidade clínica em indivíduos com artrite reumatoide. Outros estudos em populações com outras características já estão em andamento.

© 2015 Elsevier Editora Ltda. Todos os direitos reservados.

Introduction

Rheumatoid Arthritis is a chronic disabling condition and may result in impairments in functions including musculoskeletal pain, joint stiffness, loss of range of motion, muscular weakness, and joint damage. The disability causes limitations of activities and restrictions in participation.[1]

The number of clinical studies addressing functioning as a study endpoint in patients with Rheumatoid Arthritis has steadily increased during the past decade.[2] It is also important to recognize that measures in Rheumatoid Arthritis have been developed to measure the disease consequences but with little attention to functional aspects.[3] Most trials involving Rheumatoid Arthritis have used the Health Assessment Questionnaire (HAQ). This questionnaire is influenced by social factors such as education level and requires a major change in score to represent a significant functioning change in the patient.[4,5] The HAQ has lower consistency compared with other measures such as C-Reactive Protein (CRP), Erythrocyte Sedimentation Rate (ESR), number of tender joints and patient/physician global assessment.[6] Furthermore, HAQ is a tool with considerable "celling-effect" and thus are unable to detect worsening after reaching a maximum score.[7]

Foot impairment occurs in 85–100% of Rheumatoid Arthritis patients and erosive synovitis is the primary reason for high levels of pain and/or disability.[8,9] Others standardized method more specifically to assess the disability caused by the dysfunction of the feet and gait in Rheumatoid Arthritis patients is necessary for practice clinical. In 1994, the American Orthopaedic Foot and Ankle Society (AOFAS) developed rating scales for the ankle-hindfoot, midfoot, hallux metatarsophalangeal–interphalangeal, and lesser metatarsophalangeal–interphalangeal sites allowing them to be applied to different kinds of injuries and treatments.[10] AOFAS clinical domains were designed to assess foot or ankle problems and are very widely used for this purposed despite the limited evidence, until now, for their reliability and validity in other circumstances.[11]

Rehabilitation programs priorities will be based increasingly on evidence of the cost-effectiveness of interventions on functioning. The reliability of such evidence is substantially dependent on the validity of the methods used to assess health and functioning status. In study of Rodrigues et al.,[12] the AOFAS was translated and culturally adapt for Brazilian Portuguese and its reproducibility and validity were tested for patients with clinical diagnostic of ankle or hindfoot injuries. Up to this date, we do not know of their validation for patients with Rheumatoid Arthritis in Brazil. In the current study, we have tested the reliability and validity of AOFAS in a sample of Rheumatoid Arthritis patients.

Materials and methods

Participants

The study comprised a convenience sample including 33 patients from Arthritis Rheumatoid Service of the Bahia School of Medicine and Public Health. Patients were eligible for inclusion if they had more than 18 years and have a diagnoses of Rheumatoid Arthritis by a rheumatologist satisfying the American Rheumatism Association revised criteria for Rheumatoid arthritis[13] and demonstrated the ability to

walking with or without assistive devices. Patients were excluded if had neurologic dysfunction, cognitive deficits on the Mini Mental State Examination,[14] skin lesions, surgery in the lower limb, pregnants or in rheumatoid arthritis remission stage (≤2.6 on DAS-28).[15] In about 25% of cases Rheumatoid Arthritis is intermittent with periods of remission and this stage may generate a confounding effect and overestimate the results.

The study was approved by the Research Ethics Committee of the Estacio of Bahia University Center, n. 657.528 and the participants signed informed consent forms.

Procedures and instrument of measure

The questionnaire AOFAS is composed of nine items, distributed over three categories: pain, functional aspects and alignment. Patients were asked to rate the pain on a scale of 0–40 points. Function was calculated as the total score of activity limitations (0–10), maximum walking distance (0–5), walking surface (0–5), gait abnormality (0–8), sagittal motion (0–8), hindfoot motion (0–6), and ankle-hindfoot stability (0–8). Alignment was rated from 0 to 10 points (good, fair, poor), giving a total score of 100 points.[10,12] Sagittal and hindfoot motion were physical examination of joint motion measured by trained professional with goniometer.

In stage one, the AOFAS were submitted to test–retest reliability analysis by two trained professionals to use the instrument. To test their reliability, the AOFAS was applied twice with a three to five-days interval to the first 20 individuals with Rheumatoid Arthritis included in the study. In stage two, the Brazilian version of AOFAS was measured using *Rasch* analysis.

Statistical analysis

Descriptive statistics were utilized to characterize the sample. Intra-class correlation coefficients (ICCs) were used to assess the intra- and inter-rater reliabilities. The ICC was chosen in preference to the Pearson correlation which may overestimate reliability.[16] All analyses were performed with SPSS for Windows (SPSS Inc., Chicago, IL, USA) and the significance level were set at 0.05.

Rasch techniques have been shown to successfully reduce the number of items on questionnaires, a particularly important issue when developing questionnaires for disabled populations.[17] *Rasch* analysis is a probabilistic model widely used in the field of rehabilitation to evaluate the psychometric properties of scales.[18] Is based on a probabilistic relationship between item difficulty and person ability, with the difference known as the functional reserve or functional ability.[19]

Item fit statistics are employed in the assessment of unidimensionality, which demonstrates whether the questionnaire or subscale is measuring a single concept. Values such as MnSq and t in two formats: infit and outfit was performed using software Winsteps.[20] The infit statistic is sensitive to the variation of scores representing the subject's abilities, and the outfit statistic reflects the occurrence of unexpected responses. These values are used to determine whether the items fit with the unidimensionality concept, with values of MnSq $= 1 \pm 0.4$ associated with $t = \pm 2$ being acceptable. The

Table 1 – Subjects demographic and clinical characteristics.

	$n = 33$
Age (years)	53.0 (10.97) [30 – 75]
Years post rheumatoid arthritis	12.21 (7.54) [5 – 19.9]
Body mass index, kg/m2	25.7 (5.52) [19.98 – 41.66]
DAS 28	5.1 (0.9) [3.5 – 7.3]
Disability (HAQ)+	1.68 (0.65)
Ethnicity	
White	4 (12.2%)
Black	15 (45.4%)
Brown	14 (42.4%)
Level of physical activity	
Sedentary	28 (84.8%)
Sporadic practitioner	5 (15.2%)
Use of antiflamatory drug	33 (100%)

Values are mean (standard deviation) [range] or frequency (percentage)

condition in which more than 5% of the items of a scale exhibit erratic scores indicate that the corresponding combination of items is not able to measure a unidimensional construct.[21,22]

The *Rasch* model further includes the person and item separation indices and the satisfactory test divides the subjects into at least three ability levels: low, medium, and high.[23] The estimated values of reliability to calibrate the measures must be >0.80.[24]

Results

Sample characterization

Thirty-three females with rheumatoid arthritis were evaluated with a mean age of 53 ± 10.9 years (ranging from 30 to 75). Demographic and clinical characteristics of the subjects are reported in Table 1.

Test–retest reliability

Table 2 gives the ICC values for the intra- and inter-rater reliabilities. As can be noted, significant and adequate values of intra- (0.90 < ICC < 0.95; $p < 0.001$) and inter-rater (0.77 < ICC < 0.91; $p < 0.001$) reliabilities were obtained. In the

Table 2 – Intraclass coefficients (ICC) Values for the intra and interrater reliability of domain and total score AOFAS.

AOFAS	ICC	
	Intra	Inter
AOFAS – pain	0.90*	0.77*
AOFAS – function	0.92*	0.81*
AOFAS – total	0.95*	0.91*

* $p \leq 0.001$.

Table 3 – Calibration of AOFAS items.

Item	Calibration	Infit		Outfit	
		MnSq	t	MnSq	t
9. Alignment	57.11	0.96	0.0	0.61	−0.9
1. Pain**	55.96	0.35	−3.0	0.48	−1.6
4. Walking surfaces	54.07	0.79	−0.8	0.94	−0.1
8. Ankle-hindfoot stability*	50.05	1.62	2.4	1.49	2.0
3. Max continuous walking distance	48.98	0.61	−2.0	0.64	−1.8
7. Hindfoot motion	48.77	0.92	−0.3	0.91	−0.4
2. Activity limitations	46.50	1.02	0.2	1.03	0.2
6. Sagittal motion	46.06	1.30	1.3	1.17	0.7
5. Gait abnormality	42.48	1.31	1.2	1.34	1.4

* Erratic item: MnSq > 1.4 and t > 2.
** Predictable item: MnSq < 0.6.

present study, the application of the AOFAS lasted 5.5 min on average, ranging from 4 to 7 min.

Rasch analysis

In the AOFAS, the calibration stability of the items and individuals was 0.96 and 0.80, respectively. Results of *Rasch's* analysis are presented in Table 3, in which the values of the calibration or difficulty of items, MnSq and t (infit and outfit) are individually reported. The items are contained in decreasing order of difficulty. Item 9 "Alignment" was the most difficult, and item 5, "Gait abnormality", was the easiest one.

Out of the 9 items of the questionnaire, 1 (11.1%) did not fulfill the expectations of the model: numbers 8 (ankle-hindfoot stability [anterior drawer, varus-valgus stress]) demonstrated MnSq > 1.4 and t > 2.

The individual separation index was 1.9, which indicates that the items distributed the subjects among three ability levels. The item separation index was 4.75, which corresponds to approximately six levels of difficulties.

Fig. 1 displays a map that depicts the continuum of the difficulty of the items on the left and the sample ability continuum on the right. Most items had medium difficulty, whereas items allowing the assessment of individuals with very high or very low ability were lacking.

Discussion

The AOFAS clinical rating system consists of four site-specific scales that enable a focus on target sites or diseases.[10,12] Since the AOFAS was developed in English, it was necessary to carry out a transcultural adaptation of the questionnaire (i.e., a literal translation is not enough), as well as an evaluation of the psychometric properties of the population where it was supposed to be used.[24]

The *Rasch* analysis performed in the present study showed a calibration stability of the items and individuals in both investigated tests, indicating that the measures were stable and reproducible. The value found for the index of separation of the sample's individuals (1.9) indicates that they were divided into three levels of ability: low, intermediate, and high.

The *Rasch* analysis detected one item (11.1%) with erratic behaviors (number 8) which was superior to what is
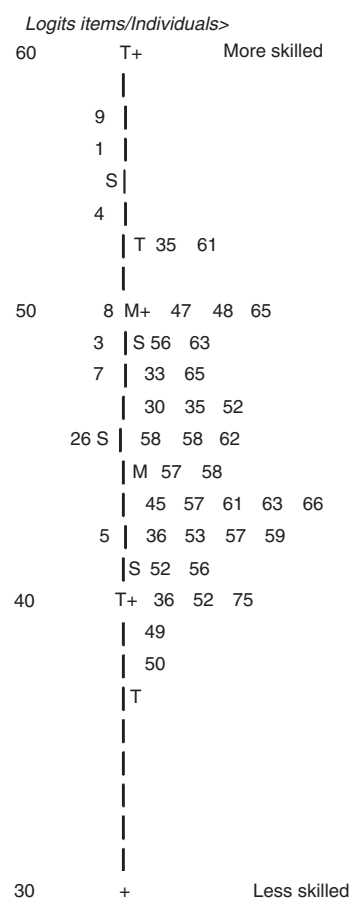

Fig. 1 – Map representing the distribution of subjects and items from equilibrium measured by the AOFAS. The numbers on the left represent the test items, and the subjects are on the right (the number identifies the age).

recommended (5%) to indicate that the instrument measured a unidimensional concept. In this item, "Ankle-hindfoot stability", the subject responds if considering your ankle stable or unstable, then there are only two score possibilities: 8 (stable) and 0 (unstable). While the further items of AOFAS have at least 03 possible answers, a fact that stratifies better and leaves clearer the possibilities for response: In addition, Rheumatoid

Arthritis is a disease that affects the feet of patients in 85 a 100% of cases; recurrent synovitis damages the tissues and the stability of the ankle-hindfoot, that worsens with advancing disease.[8] Thus it proved possible to note that individuals with longer disease claimed his ankle unstable and those with shorter disease qualified as your ankles stable.

The results of the present study show that the AOFAS exhibit a satisfactory ability to detect subtle disability deficits, thus making the distinction among persons with Rheumatoid Arthritis disability. The presence of very easy or very difficult items constitutes an advantage for the instrument. Fig. 1 items/map shows the continuous of disability by questionnaire items. This is a representation of the relationship between examined individuals Rheumatoid Arthritis severity with disability levels discriminated by scales items. We can see some items at the top, with no alignment to any individual, which means that these items measure a very high disability degree and no individuals existed in this sample with a high level of disability, the lack of individuals with high level of disability must be due to the established criterion of patients walk independently.

In this paper, we have been able to demonstrate good and excellent values of inter and intra-rater reliabilities and validity of the AOFAS. These findings provide strong support for the reliability and validity of the AOFAS measures, suggesting that they perform well in patients undergoing Rheumatoid Arthritis. The ICCs are shown as the preferred and most adequate indices for reliability analyses of interval ratio data, since they represent both correlations and levels of agreement.[23,24] According to Portney and Watkins,[25] ICC values of ≥ 0.75 are indicative of acceptable reliability and those below 0.75 are considered poor to moderate.

As found in the present results, for both intra- and inter-rater reliabilities, the ICC values were considered to be acceptable. These results are consistent with those of the study by Rodrigues et al.,[12] which assessed the values of intra and inter-rater reliabilities of the AOFAS in patients with clinical diagnostic of ankle or hindfoot injuries and found values of ICC greater than 0.9.

Although our sample shows ability to read, since most of the volunteers had completed high school education, we believe that because the subjective component of the Brazilian version of the AOFAS ankle-hindfoot scale is administered as interviews, potential interpretation errors are minimized.[26]

For further studies, it is important that the AOFAS be applied with other samples, so that the validity of the instrument can be more broadly examined. If in other kinds of samples from a number of erratic items above 5% persists, modifications to the AOFAS scale in item 8 are warranted, for example increasing the number of possible answers followed by studies on the modified version.

Conclusion

The instrument was shown to be clinically useful for the Rheumatoid arthritis sample assessed in this study. However caution must be exercised when interpreting the results, and the answers pattern must be observed, especially for items 8 regarded as erratic item.

Conflicts of interest

The authors declare no conflicts of interests.

REFERENCES

1. Stucki G, Cieza A, Geyh S, Battistella L, Lloyd J, Symmons D, et al. ICF Core Sets for rheumatoid arthritis. J Rehabil Med. 2004;44 Suppl.:87–90.
2. Cieza A, Stucki G. Understanding functioning, disability, and health in rheumatoid arthritis: the basis for rehabilitation care. Curr Opin Rheumatol. 2005;17:183–9.
3. Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. Disabil Rehabil. 2002;24:932–8.
4. Corbacho MI, Dapueto JJ. Avaliação da capacidade funcional e da qualidade de vida de pacientes com artrite reumatoide. Rev Bras Reumatol. 2010;50:31–43.
5. Leeb BF, Andel I, Leder S, Leeb BA, Rintelen B. The patient's perspective and rheumatoid arthritis disease activity indexes. Rheumatology. 2005;44:360–5.
6. Aletaha D, Machold KP, Nell VPK, Smolen JS. The perception of rheumatoid arthritis core set measures by rheumatologists. Results of a survey. Rheumatology. 2006;45:1133–9.
7. Greenwood MC, Doyle DV, Ensor M. Does the Stanford Health Assessment Questionnaires have potential as a monitoring tool for subjects with rheumatoid arthritis? Ann Rheum Dis. 2001;60:344–8.
8. Bal A, Aydog E, Aydog ST, Cakci A. Foot deformities in rheumatoid arthritis and relevance of foot funcyion index. Clin Rheumatol. 2006;25:671–5.
9. Novak P, Burger H, Tomsic M, Marincek C, Vidmar G. Influence of foot orthoses on plantar pressures, foot pain and walking ability of rheumatoid arthritis patients: a randomized controlled study. Disabil Rehabil. 2009;31:638–45.
10. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. Foot Ankle Int. 1994;15:349–53.
11. 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;25:521e5.
12. Rodrigues RC, Masiero D, Mizusaki JM, Imoto AM, Peccin MS, Cohen M, et al. Translation, cultural adaptation, and validity of the American Orthopaedic Foot and Ankle Society (Aofas) Ankle-Hindfoot Scale. Acta Ortop Bras [Serial on the Internet]. 2008;16:107–11.
13. Arnett FC, Edworthy SM, Bloch DA, MCshane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for classification or rheumatoid arthritis. Arthritis Rheum. 1988;31:315–24.
14. Brucki S, Nitrini R, Caramelli P, Bertolucci PHF, Okamoto IH. Sugestões para o uso do Miniexame do estado mental no Brasil. Arq Neuropsiquiatr. 2003;61:777–81.
15. Scott DL, Huskisson EC. The course of rheumatoid arthritis. Baillières Clin Rheumatol. 1992;6:1–21.
16. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press; 2008.
17. Mitchell JD, O'Brien MR. Quality of life in motor neurone disease – towards a more practical assessment tool? J Neurol Neurosurg Psychiatry. 2003;74:287–8.
18. Kielhofner G, Davies FA. Research in occuptional therapy. Methods Inq Pract. 2006:184–200.

19. Andrich D. A rating scale formulation for ordered response categories. Psychometrika. 1978;43:561–73.
20. Linacre JM. Winsteps Rasch Measurement Version 3.72.3 [Software]; 2011. Available from: http://www.winsteps.com
21. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. Mahwal: Lawrence Erlbaum Associates; 2001. p. 288.
22. Silverstein BF, Fisher WP, Kilgore KM, Harley JP, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: defining interval measures. Arch Phys Med Rehabil. 1992;73:507–18.
23. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil. 2005;84:719–23.
24. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25:3186–91.
25. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 2nd ed. New Jersey: Prentice-Hall; 2000.
26. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Int Med. 1993;118:622–9.