

Inter-observer agreement of the physician's global assessment of disease activity in children with juvenile idiopathic arthritis

A. Falcone, R. Cassone,
F. Rossi, A. Pistorio¹,
A. Martini, A. Ravelli

Dipartimento di Pediatria, Università di Genova, Pediatria II, Istituto di Ricovero e Cura a Carattere Scientifico G. Gaslini, Genova, Italy; and ¹Servizio di Epidemiologia Clinica e Biostatistica, Direzione Scientifica, Istituto di Ricovero e Cura a Carattere Scientifico G. Gaslini, Genova, Italy.

Alessandra Falcone, MD; Raffaella Cassone, MD; Federica Rossi, MD; Angela Pistorio, MD, PhD; Alberto Martini, MD; Angelo Ravelli, MD.

Please address correspondence to: Angelo Ravelli, MD, Pediatria II, Istituto G. Gaslini, Largo G. Gaslini no. 5, 16147 Genova, Italy.
E-mail: angeloravelli@ospedale-gaslini.ge.it

Received on April 9, 2004; accepted in revised form on November 4, 2004.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2005.

Key words: Juvenile idiopathic arthritis, physician's global assessment, visual analogue scale, disease activity.

ABSTRACT

Objective. To assess the inter-observer agreement in the physician's global assessment of overall disease activity (MD global) in a cohort of patients with juvenile idiopathic arthritis (JIA).

Methods. Forty consecutive patients with JIA, who were representative of a wide spectrum of disease activity and severity, were examined simultaneously by 4 observers. Observer 1 (who was the most experienced rheumatologist) carried out a routine rheumatologic examination of each patient including a complete articular assessment, and subsequently calculated in secrecy the MD global score on an anchored horizontal 10-cm VAS. Observers 2, 3 and 4 were present during the examination; afterwards they also scored in secrecy the MD global score for the patient. Agreement was measured by the intraclass correlation coefficient (ICC), using the score of Observer 1 as the gold standard. An ICC below 0.75 was considered unsatisfactory.

Results. The mean (SD) MD global scores for Observers 1, 2, 3, and 4 were 5.2 (3.4), 6.7 (3.9), 5.9 (3.5), and 5.6 (3.7), respectively. The level of agreement with Observer 1 in scoring was 0.83 for Observer 2, 0.88 for Observer 3, and 0.90 for Observer 4, indicating good agreement for all observers.

Conclusions. Our study shows a good inter-observer agreement in the physician's global assessment of overall disease activity in patients with JIA. Analyses involving investigators from different countries are needed to determine whether these results can be generalized.

Introduction

The physician's global assessment of overall disease activity (MD global) on a visual analogue scale (VAS) has been shown to be the most responsive measure for the detection of clinical change in patients with juvenile idiopathic arthritis (JIA) (1, 2). Based on its good performance, the MD global has been selected for inclusion in the core set of outcome variables for the assessment of clinical response in JIA clinical trials (3). Despite the widespread use of this

measure, however, a rigorous assessment of its inter-observer variation has not yet been carried out. We therefore investigated the inter-observer agreement of the MD global in a cohort of patients with JIA.

Patients and methods

All consecutive patients with JIA based on the International League of Associations for Rheumatology (ILAR) revised criteria (4) who were seen at our outpatient clinic between March 6 and April 9, 2003 were examined simultaneously by 4 observers. Observer 1 (AR) was a pediatric rheumatologist with more than 15 years of experience; Observers 2 (FR), 3 (AF), and 4 (RC) were pediatric residents who were in the 3rd, 3rd, and 5th years of their residency and had 3, 1 and 1 years of training in pediatric rheumatology, respectively. Observer 2 had been working in the study unit since the beginning of her residency, whereas Observers 3 and 4 belonged to a different university in Italy and had just started their fellowships in the study unit. Observers 2, 3, and 4 had never given an MD global before, although Observer 2 had regularly seen senior physicians (including Observer 1) to score their MD global during her residency. Prior to the study visit, Observers 3 and 4 were briefly instructed on the general clinical meaning and the scoring method of the MD global.

Observer 1 carried out a routine rheumatologic examination in each patient, including an interview regarding the patient's history, morning stiffness, fatigue, joint pain, functional limitations, and other issues relating to disease activity, and a complete joint assessment (examining a total of 67 joints). He recorded, as previously described (5), the numbers of joints with pain upon movement/tenderness, swollen joints, joints with limited range of motion (LROM) and joints with active arthritis (defined as the number of joints with swelling or, if no swelling was present, with limitation of movement and either pain upon movement or tenderness) on a specific form. He subsequently calculated (privately) the MD global score for the patient using an anchored hori-

Table I. Main clinical features of the study patients.

	No	%	Mean	Median	SD	Range
Sex (male/female)	14/26	35/65				
Onset age (years)			4.3	3.3	3.4	0.6 – 15.4
Disease duration (years)			5.1	5.0	3.9	0.2 – 13.4
Onset subtype						
Systemic	8	20				
Polyarticular	5 [#]	12				
Oligoarticular	23 [§]	58				
Enthesitis – arthritis	1	2.5				
Psoriatic	3	7.5				
Parent's global assessment [§]			3.3	2.2	3.2	0 – 10
Parent's assessment of pain [§]			3.3	2.2	3.1	0 – 9.7
CHAQ score ^{&}			0.6	0.6	0.5	0 – 1.5
No. of active joints			4.5	2.0	5.2	0 – 27
No. of swollen joints			2.9	2.0	4.0	0 – 22
No. of joints with POM/tenderness			4.7	2.5	5.0	0 – 21
No. of joints with LROM			4.8	1.0	10.6	0 – 63
Erythrocyte sedimentation rate (mm/hr)			35.8	20.5	31.5	2 – 130

[#]All rheumatoid factor-negative; [§]10 with polyarticular course; POM: pain on motion; LROM: limited range of motion.

[§]Parent global assessments could range from 0 (best) to 10 (worst); [&]CHAQ score may range from 0 (best) to 3 (worst).

zontal 10-cm VAS, with “no activity” at one end (scored as 0) and “maximum activity” at the other end (scored as 10). While Observers 2, 3 and 4 were present during the patient examination and the assessment of the articular indices, they were not allowed to discuss the findings with Observer 1 or each other, or to be present when Observer 1 was calculating the MD global score. After the examination, Observers 2, 3, and 4 each calculated a MD global score for the patient (alone, without consulting his/her colleagues) using the same 10-cm VAS as that used by Observer 1.

At the time of the study visit, one parent of each child was asked to make a global assessment of the child's overall well-being on a 10-cm VAS (0 = very good; 10 = very poor), to assess the degree of the child's pain on a 10-cm VAS (0 = no pain; 10 = very severe pain) and to complete the Italian version of the Childhood Health Assessment Questionnaire (CHAQ) (6) (0 = best; 3 = worst). The erythrocyte sedimentation rate (ESR) was determined using the Westergren method. Patient characteristics recorded retrospectively included: sex, onset age, disease duration and JIA subtype.

Table II. Agreement in MD global assessment among the 4 observers

	Observer			
	1	2	3	4
MD global				
Mean	5.2	6.7	5.9	5.6
Median	6.2	8.4	7.2	6.8
Standard deviation	3.4	3.9	3.5	3.6
Range	0-10	0-10	0-10	0-10
ICC with Observer 1	-	0.83	0.88	0.90

ICC: Intraclass correlation coefficient

Table III. Correlations between MD global assessment of Observer 1 and JIA severity parameters

	Spearman's r _s	P
Parent's global assessment [§]	0.45	0.003
Parent's assessment of pain [§]	0.56	0.0004
CHAQ score ^{&}	0.31	0.064
No. active joints	0.55	0.0003
No. swollen joints	0.64	< 0.0001
No. joints with POM/tenderness	0.45	0.004
No. joints with LROM	0.39	0.014
Erythrocyte sedimentation rate (mm/h)	0.41	0.016

POM: pain on motion; LROM: limited range of motion; [§]Parent global assessments may range from 0 (best) to 10 (worst); [&]CHAQ score may range from 0 (best) to 3 (worst)

Statistical analysis

Descriptive statistics were reported in terms of means, medians, standard de-

viations and ranges for the quantitative variables and in terms of absolute frequencies and percentages for the quali-

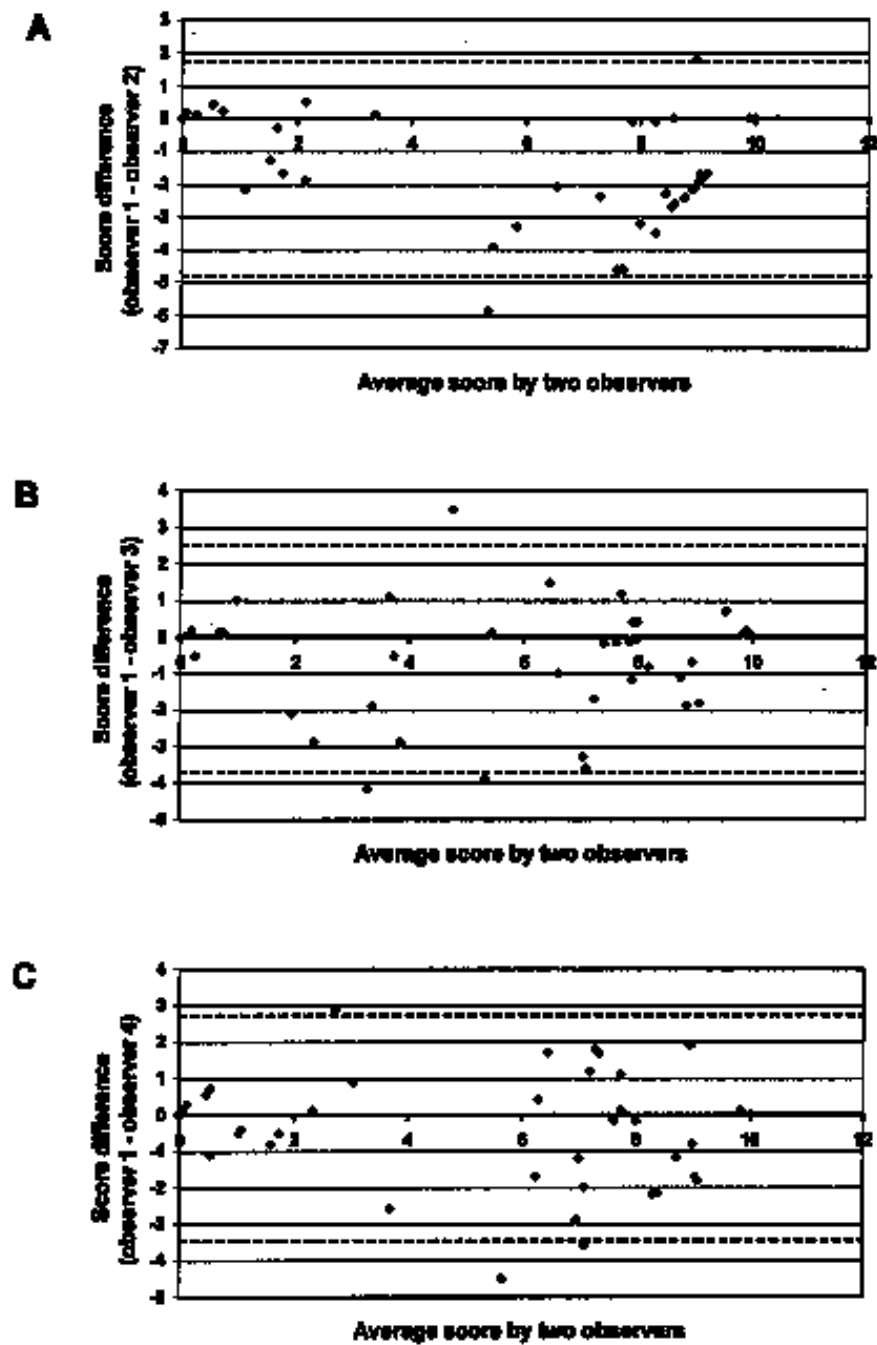


Fig. 1. Bland-Altman plots of MD global assessment by the 4 observers.

tative variables. The interobserver agreement on the MD global was measured with the intraclass correlation coefficient (ICC), using the score of Observer 1 as the gold standard. An ICC below 0.75 was considered unsatisfactory. Agreement was also evaluated by means of the Bland and Altman method (7). The bias estimate and 95% limits of agreement between pairs of

observers were calculated and reported. All tests were two-sided and a P value less than 0.05 was considered statistically significant. The correlation between the MD global of Observer 1 and JIA severity parameters was evaluated by means of the non-parametric Spearman's correlation coefficient (r_s). For the purposes of this analysis, correlations >0.7 were considered as high,

correlations ranging from 0.4 to 0.7 were considered moderate, and correlations < 0.4 were considered low. The statistical package "Statistica" (Stat-Soft Corp., Tulsa, OK) was used for all the analyses.

Results

A total of 40 JIA patients, 14 boys and 26 girls, who were representative of a wide spectrum of disease activity and severity, were examined. The JIA subtype was systemic in 8 patients, rheumatoid factor negative polyarticular in 5 patients, persistent oligoarticular in 13 patients, extended oligoarticular in 10 patients, enthesitis-arthritis in 1 patient, and psoriatic in 3 patients. The mean age of the patients at the time of the study visit was 8.7 years (range 1.2-17.8 years). Nine patients were receiving no treatment, 21 were taking nonsteroidal antiinflammatory drugs, 19 methotrexate, 7 etanercept, 4 cyclosporine A, and one hydroxychloroquine. The main clinical features of the patient population are reported in Table I.

The mean (SD) score of the MD global for Observers 1, 2, 3, and 4 was 5.2 (3.4), 6.7 (3.9), 5.9 (3.5), and 5.6 (3.7), respectively. The level of agreement with Observer 1 in scoring the MD global was 0.82 for Observer 2, 0.88 for Observer 3, and 0.90 for Observer 4, indicating good agreement for all observers. The results of the MD global assessments and the inter-observer correlation coefficients are summarized in Table II. The good agreement between observers was confirmed by the Bland-Altman plots (Fig. 1). The bias estimate and 95% limits of agreement between pairs of observers were -1.5 and -4.8-1.8, respectively, for Observers 1 and 2, -0.7 and -3.8-2.4, respectively, for Observers 1 and 3, and -0.4 and -3.5-2.7, respectively, for Observers 1 and 4. Due to the good agreement observed, we did not believe it necessary to assess factors influencing variations in agreement.

Table III presents the correlations between the MD global assessment of Observer 1 and the main JIA severity variables. All of the correlations were moderate, except those with the CHAQ score and the number of joints with LROM, which fell in the low range.

The strongest correlation with MD global was provided by the number of swollen joints. The correlation with the CHAQ score was the only one which did not prove statistically significant.

Discussion

The definition of the inter-observer variability of a clinical measure is an essential requirement before it can be widely used in clinical practice and clinical trials. Poor reliability could produce conflicting results in the assessment of disease severity and treatment outcome and make comparisons between patient groups difficult. These issues may have relevant implications in the classification of patients in different centers and in the conducting of multicenter clinical trials.

In general, there is little data on the inter-observer agreement of clinical measures in JIA. Studies on the reliability of the articular examination have shown that physicians commonly involved in the care of patients with JIA often disagreed on their assessment (8, 9). Besides the differences in judgment and examination technique, this discrepancy was found to be influenced by the severity of the illness and the type of joint involved (9).

The good inter-observer agreement observed for the MD global in our study is, perhaps, not surprising because this measure does not require knowledge of any specific assessment or scoring technique, but merely implies a subjective and global judgment of the overall level of disease activity, which is mainly based on joint symptoms (2) and, in patients with systemic JIA, also on extra-articular manifestations. It is worth noting that this satisfactory agreement was reached among rheumatologists whose specific experience was quite

different. Furthermore, only 2 of the 4 observers worked in the study unit; the other 2 had studied at a different university in Italy and were just starting their fellowship in our unit. Analyses involving investigators from different countries are needed to determine whether our results are generalizable, however.

The prominent influence of joint manifestations, particularly swelling, on the physician's subjective evaluation was confirmed by the analysis of the correlations between the MD global and the clinical variables of JIA severity. Notably, since the active joint count was found to be among the strongest determinants of functional outcome in patients with both JIA and adult rheumatoid arthritis (10, 11), this findings suggests that the physician's perception of disease severity in the earlier stages of the disease should be included in future studies aimed at identifying the prognostic factors for long-term outcome in chronic arthritides.

We must acknowledge that the fact that Observers 2, 3, and 4 did not conduct their own examinations of the patients, but merely watched while Observer 1 carried out his may have influenced the scores and partly explain their high consistency. To better simulate the conditions of the practicing rheumatologist, future studies should ask the different observers to independently examine the patients and then independently calculate the joint counts and global assessment of disease activity.

In conclusion, our study shows good inter-observer agreement in the physician's global assessment of overall disease activity in patients with JIA. This finding further demonstrates the good performance characteristics of this clinical measure.

References

1. RUPERTO N, RAVELLI A, FALCINI F *et al.*: Responsiveness of outcome measures in juvenile chronic arthritis. *Br J Rheumatol* 1999; 38: 176-80.
2. RUPERTO N, RAVELLI A, MIGLIAVACCA D *et al.*: Responsiveness of clinical measures in children with oligoarticular juvenile chronic arthritis. *J Rheumatol* 1999; 26: 1827-30.
3. GIANNINI EH, RUPERTO N, RAVELLI A, LOVELL DJ, FELSON DT, MARTINI A: Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum* 1997; 40: 1202-9.
4. PETTY RE, SOUTHWOOD TR, MANNERS P *et al.*: International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: Second revision, Edmonton, 2001. *J Rheumatol* 2004; 31: 390-2.
5. RAVELLI A, VIOLA S, RUPERTO N, CORSI B, BALLARDINI G, MARTINI A: Correlation between conventional disease activity measures in juvenile chronic arthritis. *Ann Rheum Dis* 1997; 56: 197-200.
6. RUPERTO N, RAVELLI A, PISTORIO A *et al.*: The Italian version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). *Clin Exp Rheumatol* 2001; 19 (Suppl. 23): S91-S93.
7. BLAND JM, ALTMANDG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-11.
8. GIANNINI EH, LOVELL DJ, for the PEDIATRIC RHEUMATOLOGY COLLABORATIVE STUDY GROUP: Inter-physician reliability estimates for conventional disease activity measures in JRA. *Arthritis Rheum* 1994; 37 (Suppl.): S328.
9. GUZMAN J, BURGOS-VARGAS R, DUARTE-SALAZAR C, GÓMEZ-MORA O: Reliability of the articular examination in children with juvenile rheumatoid arthritis: interobserver agreement and sources of disagreement. *J Rheumatol* 1995; 22: 2331-6.
10. RAVELLI A, MARTINI A: Early predictors of outcome in juvenile idiopathic arthritis. *Clin Exp Rheumatol* 2003; 21 (Suppl. 31): S89-S93.
11. GERBER LH, FURST G, YARBORO C, EL-GABALAWY H: Number of active joints, not diagnosis, is the primary determinant of function and performance in early synovitis. *Clin Exp Rheumatol* 2003; 21 (Suppl. 31): S65-S70.