

# Patient-reported Joint Count in Juvenile Idiopathic Arthritis: The Reliability of a Manikin Format

Maryanne E. Dijkstra, Janneke Anink, Philomine A. van Pelt, Johanna M. Hazes, and Lisette W.A. van Suijlekom-Smit

**ABSTRACT. Objective.** To evaluate the reliability of a manikin format, patient-reported joint count in juvenile idiopathic arthritis (JIA), and to detect changes in agreement at a second visit.

**Methods.** Patients with JIA aged 12–21 were asked to mark joints with active arthritis on a manikin before their regular clinic visit. The physician then performed a joint count without having seen the patient's assessment. Agreement between scores of physician-reported and patient-reported joint counts was assessed using ICC. Kappa statistics were used to assess reliability of scoring individual joints.

**Results.** The study included 75 patients with JIA. In general, patients had a low number of active joints (median 1 joint, indicated by the physician). ICC was moderate (0.61) and  $\kappa$  ranged from 0.3–0.7. At the second visit,  $\kappa$  were similar; the ICC was 0.19. When a patient scored 0 joints, the physician confirmed this 93%–100% of the time. When the patient marked  $\geq 1$  joints, the physician confirmed arthritis 59%–76% of the time. Sensitivity to change was moderate.

**Conclusion.** Agreement between physician and patient on the number of joints with active arthritis was reasonable. Untrained patients tended to overestimate the presence of arthritis when they marked active joints on a manikin-format joint count. When the patient indicated absence of arthritis, the physician usually confirmed this. As the agreement did not improve at followup, future research should focus on the possibility of achieving this through training. For now, the patient-reported joint count cannot replace the physicians' joint count in clinical practice; it may be used in epidemiological studies with caution. (J Rheumatol First Release Nov 29 2014; doi:10.3899/jrheum.140073)

## Key Indexing Terms:

JUVENILE RHEUMATOID ARTHRITIS  
PHYSICAL EXAMINATION

OUTCOME ASSESSMENT (HEALTH CARE)  
SELF-EXAMINATION ADOLESCENT

Juvenile idiopathic arthritis (JIA) is one of the most common chronic autoimmune diseases in childhood, affecting 47–87 per 100,000 children; it requires regular monitoring<sup>1,2</sup>. One method of monitoring is gathering information from patients themselves through a patient-reported outcome (PRO) measure. PRO can be a useful tool to observe disease activity between clinic visits, increase patient involvement, and aid in epidemiological surveys. Therefore, they are increasingly being developed and applied<sup>3,4,5</sup>.

One of these PRO is the self-reported joint count, the use of which has extensively been investigated in studies in patients with rheumatoid arthritis (RA). Conclusions on agreement between patient-reported and physician-reported joint counts are mixed. Results vary between good and poor to moderate agreement. A manikin format, in which patients indicate which joints are inflamed on a figure, generally yields higher scores than a text format<sup>6</sup>.

Only 2 studies have been performed investigating self-reported joint counts in JIA, 1 using a text format and 1 using a manikin format<sup>7,8</sup>. With the text-format joint count agreement between physicians and patients/parents on individual joints being moderate, the investigators, therefore, concluded that this joint count could not replace the physician's assessment<sup>8</sup>. The manikin format was used in a study investigating whether patients could discriminate active from inactive disease; patients seldom missed arthritis, but frequently overestimated disease activity. The overall agreement and the agreement on individual joints were not described. Both of these studies were cross-sectional in design, only describing the first time patients were confronted with a joint count.

From the Department of Pediatrics and Pediatric Rheumatology, Sophia Children's Hospital, Erasmus University Medical Centre (Erasmus MC), Rotterdam; and the Department of Rheumatology, Erasmus MC, Rotterdam, the Netherlands.

M.E. Dijkstra, MSc; J. Anink, MD, MScE; L.W.A. van Suijlekom-Smit, MD, MScE, PhD, Department of Pediatrics and Pediatric Rheumatology, Sophia Children's Hospital, Erasmus MC; P.A. van Pelt, MD; J.M. Hazes, MD, MScE, PhD, Professor, Department of Rheumatology, Erasmus MC.

Address correspondence to Dr. J. Anink, Department of Pediatrics/Pediatric Rheumatology, Na 1723, Erasmus MC Sophia Children's Hospital, PO Box 2060, 3000 CB Rotterdam, the Netherlands.  
E-mail: j.anink@erasmusmc.nl

Accepted for publication October 15, 2014.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2014. All rights reserved.

The aims of our current study were to evaluate the agreement between physician-reported joint counts and self-reported active joint counts by patients with JIA using a manikin, and to determine whether the agreement between the physician and the patient changed over time.

## MATERIALS AND METHODS

Data were collected prospectively at the JIA outpatient clinic at the Erasmus Medical Centre, a tertiary referral center in Rotterdam, the Netherlands. All consecutive patients with JIA aged 12–21 fulfilling the International League of Associations for Rheumatology criteria who visited the clinic between February 2013 and February 2014 were invited to participate<sup>9</sup>. Our study was performed according to the regulations of the local ethics committee.

At a regular clinic visit, patients were first asked to mark on a figure the joints they felt to have active arthritis. Active arthritis was defined as swelling within a joint, or limited range of motion accompanied by joint pain or tenderness<sup>9</sup>. The definition of active arthritis was provided on the figure (Figure 1). No additional information was given. After the patient had filled out the manikin, the pediatric rheumatologist (PhvP) performed a formal joint count, without having seen the patient's assessment. This practice continued during followup, giving patients an indirect feedback moment during each following visit and minimum education to see how the agreement would change naturally over time. The following joints were taken into account: temporomandibular; cervical spine; shoulder; sterno-

clavicular; elbow; wrist; metacarpophalangeal, proximal interphalangeal, and distal interphalangeal (analyzed both separately and as a unit, hand); back; sacroiliac joints; hip; knee; ankle; and metatarsophalangeal and phalanges of the foot (analyzed both separately and as a unit, foot). The acromioclavicular and subtalar joints were not evaluated because they were judged to be too difficult to assess.

Additionally, patients were asked to fill out a Child Health Assessment Questionnaire (CHAQ)<sup>10</sup>, including visual analog scales (VAS, ranging from 0–100) for well-being and pain. Demographic data and data on disease history were collected from the charts.

*Statistical analysis.* For both the first and second visits, agreement was assessed in various ways. For the level of agreement on the number of active joints, a 2-way random single measure absolute agreement ICC was used. We used the following interpretation for Cohen  $\kappa$  and ICC:  $< 0.40$  = poor,  $\geq 0.40$ – $0.60$  = moderate,  $> 0.60$ – $0.80$  = substantial, and  $> 0.80$  = good reliability<sup>11</sup>. Kappa statistics were used for calculating agreement on individual joints. Additionally, the overall agreement and the positive/negative agreement proportion per joint were computed<sup>12</sup>. Overall agreement was the percentage of joints that were scored identically by physician and patient. The positive agreement was the number of joints that were scored as being active by both physician and patient, divided by the average number of joints scored positive. The negative agreement was the number of joints scored as inactive by both parties, divided by the average number of joints scored negative. Negative agreement was expected to be high, as most joints were expected to be scored as inactive.

We calculated that a sample size of 59 subjects with 2 observations per

Name: .....

Date: .....

**Mannequin joint count: which joints are currently inflamed?**

Please mark the joints where the arthritis is currently active (where there is inflammation) on the mannequin below.

A joint is considered to be inflamed when at least two of the following criteria are present:

- pain
- swelling
- limited motion.

Note: if a joint is always restricted, but none of the other two features are present, you don't need to mark it.

**RIGHT**
**LEFT**

I have no complaints, none of my joints are inflamed.

Figure 1. Manikin joint-count form as filled out by the patients (translated from Dutch).

subject achieved 80% power to detect an ICC of 0.6 ( $\rho_1$ ) when the ICC was assumed to be at least more than 0.35 ( $\rho_0$ ) using an F test with a significance level of 0.05<sup>13</sup>.

To evaluate whether patients could discriminate between inactive and active arthritis, taking the physician joint count as reference, the (positive and negative) predictive values of a patient scoring 0 active joints or > 0 active joints were calculated. In addition, sensitivity and specificity were calculated, enabling us to compare our results to previous research done on this subject.

For the assessment of construct validity, Spearman rho correlation coefficient was calculated to test the correlation between VAS scores and the number of affected joints indicated on the 2 joint counts. To test the difference between the various correlation coefficients over time, the Fisher's r-to-z transformation was used. Correlation coefficients were interpreted as follows:  $\leq 0.3$  = weak, 0.4–0.6 = moderate, and  $\geq 0.7$  = strong.

Absolute and proportional changes in total joint count scores between the first and the second visits were calculated for both the patient-reported and physician-reported joint counts. Consequently, sensitivity to change was assessed using 2 kinds of coefficients<sup>14</sup>. First, Pearson correlation coefficient was used to assess the correlation between absolute changes in patient-reported and physician-reported joint counts for the total group of patients. In addition, the ICC of the change scores was calculated. Second, patients were divided into 3 groups (improved, stable, and worsened) according to the change in the physician-reported joint counts. Standardized response means (SRM) were calculated for the groups that improved or worsened based on the mean proportional change in the patient-reported joint counts and its SD.

Descriptive statistics were reported as absolute number, median with interquartile range (IQR), or mean with range as appropriate. Data were analyzed using the IBM SPSS statistics for Windows package, version 21.0 (IBM Corp.).

## RESULTS

**Patient characteristics.** Of the 80 patients who agreed to participate, only 75 could be used for the analysis of agreement because in 5 patients, the physician did not perform a full formal joint count. Characteristics of all 75 consecutive patients are shown in Table 1. None of the patients refused to participate. For 53 patients, a second measurement was present. The inferences on the second visit are discussed below. Patients with a second visit did not differ from patients without a second visit with regard to disease activity at the first visit. Patients had a median age of 16 years (IQR 15–18) and an average disease duration of 3.7 years (IQR 0.9–8.7). Overall disease activity was low; both physician and patients indicated low disease activity on a VAS (median < 20). The distribution of JIA categories was representative for an outpatient clinic patient population within these age ranges.

**Agreement at the first measurement.** The median number of active joints scored by the physician was 1 joint (IQR 0–3). The patients scored a median of 2 joints (IQR 1–5). The ICC was moderate with a value of 0.61 (95% CI 0.43–74). Adolescents (aged 12–17 yrs, n = 44, ICC 0.69, 95% CI 0.48–0.83) appeared to agree with their physicians more than young adults did (aged 18–21 years, n = 31, ICC 0.45 95% CI –0.14–0.69), although not statistically significant. Comparable estimates were found for patients with short disease duration ( $\leq 1$  yr, n = 19) and patients with longer

Table 1. Patient and disease characteristics at the first visit during which patients filled out the manikin.

Characteristics	n = 75
Age, yrs, median (IQR)	16.0 (14.7–17.8)
Disease duration, yrs, median (IQR)	3.7 (0.9–8.7)
Female, n (%)	54 (72)
JIA category, n (%)	
Systemic	5 (7)
Oligoarticular persistent	17 (23)
Oligoarticular extended	7 (9)
Polyarticular RF+	12 (16)
Polyarticular RF–	18 (24)
Psoriatic arthritis	9 (12)
Enthesitis-related arthritis	6 (8)
Undifferentiated arthritis	1 (1)
Laboratory blood tests, n (%)	
ANA+	27 (36)
HLA-B27+	13 (17)
RF+	12 (16)
CCP+	11 (15)
History of uveitis, n (%)	11 (15)
Disease activity variables, median (IQR)	
ESR date visit	9 (3–23)
CHAQ date visit	0.375 (0.000–0.875)
VAS physician	8 (0–28)
VAS well-being patient	12 (0–42)
VAS pain patient	22 (1–56)
Education level, n (%)	
Primary school	3 (4)
Initial vocational education	22 (29)
Upper secondary education	22 (29)
Vocational education	10 (13)
University/college	7 (9)
Unknown	11 (15)

IQR: interquartile range; JIA: juvenile idiopathic arthritis; RF: rheumatoid factor; ANA: antinuclear antibody; CCP: cyclic citrullinated peptide; ESR: erythrocyte sedimentation rate; CHAQ: Childhood Health Assessment Questionnaire; VAS: visual analog scale.

disease duration (n = 56) with respective ICC of 0.69 (95% CI 0.35–0.87, n = 19) and 0.45 (95% CI 0.21–0.63, n = 56). Agreement between patients and physicians on individual joints is reported in Table 2.  $\kappa$  values ranged from 0.3–0.7. Overall agreement was generally around 90% or higher. The agreement for the knees was lower: 75% for the right knee and 79% for the left knee. Differences in agreement between left and right occurred in other joints too. Positive agreement was generally poor to reasonable (33–75%, lowest scores for the shoulders) whereas negative agreement was excellent (82–99%, lowest scores for the knees). This last finding was expected because disease activity was low and therefore most joints would be negatively scored.

In Table 3, scores are compared between the patients' and physicians' results of the first time they scored the manikin, depending on the number of active joints scored: 0, 1, 2–4, 5–10, or more than 10 joints. In all 5 groups, patients mostly overestimated the number of active joints. Underestimation

Table 2. Agreement on individual joints at the first visit (n = 75). Individual hand and foot joints, temporomandibular joints, cervical spine, sternoclavicular joints, and the SI joints were omitted because of distorted or negative  $\kappa$  values.

Joint	$\kappa$ Value	Overall Agreement, %	Positive Agreement, %	Negative Agreement, %	Total No. Times Joint Scored Positive by Physician, n (%)	Total No. Times Joint Scored Positive by Patients, n (%)
Shoulder left	0.38	89	43	94	4 (5)	10 (13)
Shoulder right	0.28	89	33	94	5 (7)	7 (9)
Elbow left	0.58	93	62	96	7 (9)	6 (8)
Elbow right	0.64	95	67	97	5 (7)	7 (9)
Wrist left	0.68	89	75	93	15 (20)	17 (23)
Wrist right	0.65	88	73	92	17 (23)	16 (21)
Hand left	0.44	85	48	91	9 (12)	14 (19)
Hand right	0.39	81	44	92	12 (16)	16 (21)
Hip left	0.57	95	60	97	4 (5)	6 (8)
Hip right	0.57	95	60	97	4 (5)	6 (8)
Knee left	0.51	79	65	85	18 (24)	28 (37)
Knee right	0.41	75	56	82	15 (20)	28 (37)
Ankle left	0.47	87	55	92	9 (12)	13 (17)
Ankle right	0.34	88	40	93	5 (7)	10 (13)
Foot left	0.31	95	36	99	5 (7)	6 (8)
Foot right	0.47	95	50	97	4 (5)	4 (5)

SI: sacroiliac.

Table 3. Agreement on number of joints between physician and patient at the first visit. Agreement is indicated by the shaded areas.

Patient Joint Count	Physician Joint Count					Total Joint Count Patient
	0 joints	1	2-4	5-10	More Than 10	
0 joints	18	0	0	0	0	18
1	7	7	4	0	0	18
2-4	7	5	6	1	1	20
5-10	0	2	2	3	1	8
More than 10	0	0	3	4	4	11
Total joint count physician	32	14	15	8	6	75

of the total number was less common. When over or underestimation occurred, this remained confined to the closest categories of number of joints.

The presence of arthritis was indicated by 14 of 31 patients where the physician found no arthritis. The knees were the most marked joints in this group (n = 11). The VAS pain of these patients was higher than the VAS pain of patients who agreed with the physician on inactive disease (median VAS pain 18 vs 0, p = 0.005, Mann-Whitney U test). Possible explanations for the overestimation were residual complaints after recent arthritis/structural damage in 5 patients, pain after high physical activity in 4 patients, and enthesitis/tendinitis in 2 patients. Three patients had arthritis within 2 months after the first visit, and 1 of those patients did have arthritis on ultrasound evaluation. In these 3 patients, the physician may have missed arthritis on examination.

Taking the physician's joint count as a reference, the predictive value of a patient scoring 0 active joints was

100%. This means that when a patient scored inactive disease, the physician generally indicated that there was no arthritis (negative predictive value). When a patient did score a number of active joints, only in 76% did the physician agree there was arthritis present (positive predictive value). Sensitivity and specificity were 100% and 56%, respectively. Sensitivity, specificity, and negative and positive predictive values for discriminating inactive from active disease did not change when only the most affected joints (shoulders, elbows, wrists, hands, hips, knees, ankles, and feet) were used.

*Construct validity.* We performed Spearman correlations to test the correlations between the several VAS and joint counts. The physician-reported joint count correlated very well with the patient-reported joint count and the VAS physician (both a Spearman rho of 0.80 and 0.79, respectively), but less well with the VAS well-being of the patient (Spearman rho of 0.61 and 0.65). The patient-reported joint

count correlated moderately with the VAS well-being (Spearman rho of 0.49) and with the VAS pain (Spearman rho of 0.64).

*Longitudinal agreement and sensitivity to change.* At the second visit, the median number of active joints reported was 0 joints (IQR 0–3) for physicians and 2 joints (IQR 0–6) for patients. The ICC that was estimated was 0.19 (95% CI –0.05–0.42). The CI indicated a possible negative value of the ICC, which was caused by a large variation between and within subject variability. The interpretation of this ICC can only be that there is very low agreement<sup>15</sup>. We suspected that this was caused by 4 subjects with a very large discordance between physician and manikin joint count. These patients all scored 30 or more active joints while the physician scored 0–10 joints. The ICC with these subjects removed was still low (0.30, 95% CI 0.04–0.53).  $\kappa$  values were similar during followup compared to the first time of marking active joints on the manikin. Negative predictive value for the second visit was 93% and positive predictive value was 59%. Sensitivity was 96% and specificity was 45%.

At the second visit, the physician indicated inactive disease in 29 patients, 16 of whom did not agree. This was a slightly higher percentage than at the first visit. The negative predictive value was slightly lower than at the first visit because 1 patient indicated inactive disease where the physician did find arthritis on physical examination. This patient indicated no complaints. The physician joint count did indicate improvement from the previous visit; however, the disease was not fully inactive.

At the second visit, we found the patient-reported joint count to have a stronger correlation ( $p < 0.05$ ) with VAS well-being (0.75) and to have a weaker correlation with physician joint count (0.36) compared with the first visit (0.49 and 0.8, respectively). Other correlations did not change significantly during followup.

The absolute changes in physician-reported and patient-reported joint counts were moderately correlated (Pearson rho 0.436,  $p = 0.001$ ). The ICC for the change scores was 0.31 (95% CI 0.05–0.53). The SRM for the proportional change in patient-reported joint counts was moderate (0.67) in patients who worsened according to the physician. The SRM for patients who improved according to their physician was low (0.23). Therefore, the patient-reported joint count appeared to be most sensitive to change for patients whose disease became more active over time.

## DISCUSSION

Our study is one of the first to investigate whether a patient-reported joint count based on a manikin format can be used as a PRO in JIA. The overall agreement between the physician and the patient total joint count was found to be moderate (ICC 0.61) the first time patients filled out the

manikin. Agreement on individual joints was moderate to good, depending on the joint ( $\kappa$  0.3–0.7). At the second visit, the  $\kappa$  values stayed stable; however, the ICC decreased during followup. Construct validity was high; however, the second time patients filled in the joint count, the correlation to general well-being scores and pain was higher and the correlation to the physician joint count was lower than at the first visit.

Patients tended to overestimate the presence of arthritis. A patient-reported joint count indicating full absence of arthritis nevertheless proved to be highly reliable. These results were consistent over time. Sensitivity to change over time proved moderate, and was highest in patients whose disease worsened.

Two other studies have investigated patient-reported joint counts in patients with JIA. The first one tested a text-format joint count in a very large group of patients and parents of patients with JIA<sup>8</sup>. In a conference abstract, it reported agreement on individual joints ranging from 0.15 for the shoulder to 0.69 for the cervical spine. In general, these  $\kappa$  values are comparable to the ones found in our present study. The most frequently scored joints often had the highest  $\kappa$  values.

The second study investigating patient-reported joint counts in JIA used the same manikin format as our present study<sup>7</sup>. While the study used the manikin format, it focused on the question of whether patients could distinguish between active and inactive disease, and did not evaluate agreement on individual joints or total joint scores. Additionally, it did not focus solely on patients, but investigated patient and parent assessments. Sensitivity was comparable and specificity slightly lower than the specificity we found. With regard to the ability of patients to discriminate between inactive and active disease, the authors reached the same conclusion: patients did not miss arthritis, but overestimated the presence of it frequently.

To our knowledge, overall agreement has not been investigated in studies of patients with JIA. It has been done in studies of adult patients with RA, and the ICC of patients with RA with their physicians were comparable to the ICC found in our present study<sup>6</sup>. It has to be kept in mind, however, that ICC do not generalize well from one study to another because they are strongly influenced by the variance in the population in the study.

Patients with RA and JIA have been shown to overestimate their disease activity compared to the physician's estimation on other disease activity scales as well<sup>16,17</sup>. The reason for this overestimation is not clear, but it has been suggested that high functional disability and pain might influence this discordance<sup>6,17</sup>. Also in our study, patients who agreed with their physician on the inactivity of the disease had lower pain scores than did patients who indicated disease activity where the physician did not. There was no significant difference in CHAQ scores between the

2 groups. Patients may have difficulty distinguishing pain caused by active arthritis from pain as a result of other causes. Although radiological joint damage is relatively uncommon in the pediatric population, structural damage could be a cause of pain, as could muscular strain<sup>18,19</sup>. Persistent pain and the subsequent sensitization of the central nervous system have been proposed to cause a lowering of the pain threshold and altering pain perception in patients with JIA<sup>20,21</sup>. This could be an alternative explanation for reported pain uncorrelated to disease activity indicated by the physician. It could also partly explain the high correlation between the patient-reported joint count and the VAS pain.

Before considering the implementation of a patient-reported joint count, the reasons for overestimation should be more thoroughly investigated. The purpose of the use of a patient-reported joint count in clinical practice also has to be clearly defined. Armbrust, *et al*<sup>7</sup> used the joint count as a general assessment of disease activity, which only makes the distinction between active and inactive disease. In that respect, we found that the patient-reported joint count predicts the activity as marked on the physician-reported joint count better than does the VAS for general well-being. If the patient-reported joint count would be used for this purpose, one could consider only using the most affected joints because the discriminative performance did not change when only these joints were taken into account. For this purpose, it is encouraging that even without training, patients can generally be trusted when they indicate inactive disease.

The other option is to use the joint count not only to discriminate between active and inactive disease, but to actually monitor disease activity over time. The possibility to monitor their disease activity may stimulate patients' self-management and their adherence to therapy<sup>22</sup>. For this purpose, it is reassuring that the sensitivity to change was highest in the group that worsened. In addition, although the agreement on overall and individual joints was only moderate, we have to keep in mind that 2 physicians examining the same patient agree to the same extent (moderately) on the presence of active or inactive disease<sup>23,24</sup>. Still, ideally we would like to improve the absolute agreement between physician and patient so that all changes in disease activity can be monitored accurately. This could be done by means of a training program. Training patients with RA to examine their own joints had a positive effect on the reliability of patient-reported joint counts<sup>25</sup>.

The way the manikin should be filled out is also a question to be answered. The addition of a "doubt" option did not seem to add to the discriminative power of the manikin, but this may change with training<sup>7</sup>. Also, one could question whether the patient would have to mark every joint as the patients did in the study by Armbrust, *et al* because the results with regard to predictive value and sensitivity and specificity did not differ much from those we found in

our current study. Only marking the active joints seems to be sufficient, and is less time consuming<sup>7</sup>.

For the use in epidemiological surveys, the agreement found seems to be acceptable, although one should realize that the obtained estimate of disease activity is not flawless. In this setting, a more general indication of disease activity was sufficient because the main objective is to describe disease status on a population level and no individual decisions were based on these data.

For generalization of the results from our present study, it has to be taken into account that the JIA population in our study was 12 years and older and sampled at a tertiary referral clinic. Further, consecutive (and therefore, mostly already treated) patients visiting our clinic were included, resulting in a fairly homogeneous population with regard to disease activity, which was generally low.

When interpreting the results from our study, it has to be kept in mind that the examination by the physician is not flawless either<sup>23,24</sup>. Although imaging techniques are increasingly being applied in pediatric rheumatology, for most modalities, no reference standards are available yet. So, even though it seems that ultrasound and other techniques could be of help in identifying subclinical arthritis, the scoring systems used still need more validation before they can be used extensively to guide clinical practice<sup>26</sup>. In future studies, in addition to including a training program, multiple physicians could assess the patients so that a more reliable physician joint count could be obtained.

Because they received no training, patients may have made mistakes while filling out the joint count. Despite the manikin clearly stating which side was left and which was right, some patients might have filled in the form the wrong way round, thereby causing an overall lower positive agreement. In addition, patients may not have been aware of the existence of referred pain, and may have indicated the wrong joint to be active (for instance, in the case of the hip-knee).

The ICC of the second measurement was much lower than the ICC of the first measurement. We provided an estimate for the ICC without 4 outliers because the ICC that was estimated for the whole group was unreliable. For an unknown reason, at the second measurement, there were more outlying values, causing the within-subject variability to be disproportionate to the between-subject variability. The correlation between the physician-reported joint count and the patient-reported joint count also changed. The patient-reported joint count correlated more with the VAS pain. Apparently these patients were more likely to fill out the manikin, marking the joints with pain instead of those with active arthritis.

Agreement between physician- and patient-reported joint counts was moderate. Especially a joint count of 0 by the patient was predictive of the joint count of the physician. This PRO cannot, therefore, fully replace the physician's joint count at a regular clinical visit.

The manikin joint count could be used to aid in epidemiological surveys because it gives a reasonable estimate of the true number of active joints. Before being implemented in a clinical setting, more research is needed to determine whether agreement can be improved by training and whether the patient-reported joint count is then also better able to detect changes over time.

## REFERENCES

- Andersson Gäre B. Juvenile arthritis— who gets it, where and when? A review of current data on incidence and prevalence. *Clin Exp Rheumatol* 1999;17:367-74.
- Harrold LR, Salman C, Shoor S, Curtis JR, Asgari MM, Gelfand JM, et al. Incidence and prevalence of juvenile idiopathic arthritis among children in a managed care population, 1996-2009. *J Rheumatol* 2013;40:1218-25.
- Duffy CM. Measurement of health status, functional status, and quality of life in children with juvenile idiopathic arthritis: clinical science for the pediatrician. *Pediatr Clin North Am* 2005; 52:359-72, v.
- Filocamo G, Consolaro A, Schiappapietra B, Dalprà S, Lattanzi B, Magni-Manzoni S, et al. A new approach to clinical care of juvenile idiopathic arthritis: the Juvenile Arthritis Multidimensional Assessment Report. *J Rheumatol* 2011;38:938-53.
- Ringold S, Wallace CA, Rivara FP. Health-related quality of life, physical function, fatigue, and disease activity in children with established polyarticular juvenile idiopathic arthritis. *J Rheumatol* 2009;36:1330-6.
- Barton JL, Criswell LA, Kaiser R, Chen YH, Schillinger D. Systematic review and metaanalysis of patient self-report versus trained assessor joint counts in rheumatoid arthritis. *J Rheumatol* 2009;36:2635-41.
- Armbrust W, Kaak JG, Bouma J, Lelieveld OT, Wulfraat NM, Sauer PJ, et al. Assessment of disease activity by patients with juvenile idiopathic arthritis and the parents compared to the assessment by pediatric rheumatologists. *Pediatr Rheumatol Online J* 2013;11:48.
- Filocamo G, Pistorio A, Bertamino M, Consolaro A, Ferrari C, Lanni S, et al. Are children with juvenile idiopathic arthritis and their parents reliable reporters of joint involvement? [Abstract]. *Arthritis Rheum* 2009;60 Suppl 10:239.
- Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol* 2004;31:390-2.
- Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis Rheum* 1994;37:1761-9.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull* 1971;76:365-77.
- Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-10.
- Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. *Health Qual Life Outcomes* 2005;3:23.
- Giraudeau B. Negative values of the intraclass correlation coefficient are not theoretically possible. *J Clin Epidemiol* 1996;49:1205-6.
- Janta I, Naredo E, Martínez-Estupiñán L, Nieto JC, De la Torre I, Valor L, et al. Patient self-assessment and physician's assessment of rheumatoid arthritis activity: which is more realistic in remission status? A comparison with ultrasonography. *Rheumatology* 2013;52:2243-50.
- Consolaro A, Vitale R, Pistorio A, Lattanzi B, Ruperto N, Malattia C, et al. Physicians' and parents' ratings of inactive disease are frequently discordant in juvenile idiopathic arthritis. *J Rheumatol* 2007;34:1773-6.
- Flatø B, Lien G, Smerdel A, Vinje O, Dale K, Johnston V, et al. Prognostic factors in juvenile rheumatoid arthritis: a case-control study revealing early predictors and outcome after 14.9 years. *J Rheumatol* 2003;30:386-93.
- Oen K, Reed M, Malleson PN, Cabral DA, Petty RE, Rosenberg AM, et al. Radiologic outcome and its relationship to functional disability in juvenile rheumatoid arthritis. *J Rheumatol* 2003;30:832-40.
- Leegaard A, Lomholt JJ, Thastum M, Herlin T. Decreased pain threshold in juvenile idiopathic arthritis: a cross-sectional study. *J Rheumatol* 2013;40:1212-7.
- Munro J, Singh-Grewal D. Juvenile idiopathic arthritis and pain — more than simple nociception. *J Rheumatol* 2013;40:1037-9.
- El Miedany Y, El Gaafary M, El Arousy N, Ahmed I, Youssef S, Palmer D. Arthritis education: the integration of patient-reported outcome measures and patient self-management. *Clin Exp Rheumatol* 2012;30:899-904.
- Guzmán J, Burgos-Vargas R, Duarte-Salazar C, Gómez-Mora P. Reliability of the articular examination in children with juvenile rheumatoid arthritis: interobserver agreement and sources of disagreement. *J Rheumatol* 1995;22:2331-6.
- Senocak O, Unsal E, Akalin E, Ergör G. Interobserver reliability of articular examination in juvenile idiopathic arthritis. *Turk J Pediatr* 2003;45:29-32.
- Levy G, Cheatham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in rheumatoid arthritis and the role of training. *J Rheumatol* 2007;34:1261-5.
- Lanni S, Wood M, Ravelli A, Magni Manzoni S, Emery P, Wakefield RJ. Towards a role of ultrasound in children with juvenile idiopathic arthritis. *Rheumatology* 2013;52:413-20.