Minimal Clinically Important Differences of the Childhood Health Assessment Questionnaire

HERMINE I. BRUNNER, MARISA S. KLEIN-GITELMAN, MICHAEL J. MILLER, ANDREA BARRON, NICOLE BALDWIN, MICHAEL TROMBLEY, ANNE L. JOHNSON, ANGIE KRESS, DANIEL J. LOVELL and EDWARD H. GIANNINI

ABSTRACT. ***Objective.*** The Childhood Health Assessment Questionnaire (CHAQ) is a commonly used measure of disability and physical function for children with juvenile rheumatoid arthritis (JRA), whose scores range between 0 (no disability) and 3 (very severe disability), with a smallest potential difference in the CHAQ score of individuals at 0.125. We estimated minimal clinically important differences (MCID) of the CHAQ for worsening and improvement that were actually experienced by children with JRA using patient, parent, and clinical perspectives.

Methods. Changes in CHAQ scores were calculated for parent (n = 92) and patient ratings (children age ≥ 8 yrs only; n = 67) between subsequent clinic visits. Changes in patient well being and disease activity and the occurrence of flare or important improvement between visits served as external standards for the MCID. MCID were defined as the median changes of the CHAQ scores of individual patients who had a minimal important improvement or worsening between visits.

Results. The median change in CHAQ scores of patients who rated themselves or were rated by others as unchanged was often 0. Depending on the external standard used, the MCID for improvement of the CHAQ was –0.188 at most, while the MCID for worsening was at most +0.125.

Conclusion. The MCID of the CHAQ for both improvement and worsening are often at or close to the level of the smallest potential difference, suggesting that the CHAQ is relatively insensitive to important short term changes in children with JRA. This may warrant a change in the calculation of the global CHAQ score, or the development of more sensitive functional measures. (J Rheumatol 2005;32:150–61)

Key Indexing Terms:
CHILDHOOD HEALTH ASSESSMENT QUESTIONNAIRE FUNCTION
ARTHRITIS MINIMAL CLINICALLY IMPORTANT DIFFERENCES
JUVENILE RHEUMATOID ARTHRITIS JUVENILE IDIOPATHIC ARTHRITIS

The Childhood Health Assessment Questionnaire (CHAQ)[1] has been developed to measure physical function and disability of children with juvenile rheumatoid arthritis (JRA), but it can also be used for children with other chronic musculoskeletal (MSK) diseases[2,3]. Since its initial publication, the CHAQ has been translated into many languages[4] and is used worldwide for assessing children with chronic MSK diseases. Although other measures of physical function have been developed for children with arthritis[5-8], the CHAQ is the most commonly used. In addition, the CHAQ is often chosen to serve as core response variable (CRV) when assessing clinically relevant changes[9,10] of patients with JRA[11] and juvenile idiopathic arthritis (JIA)[12] in clinical practice and research. Despite a large number of studies on the test-retest reliability, construct validity, and quality of the parent-proxy report of the CHAQ[2-4,13,14], the minimal clinically important differences (MCID) of the CHAQ have not been well examined.

Initially, the MCID has been defined by Jaeschke as "the smallest (absolute) difference in score which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management"[15,16]. Thus differences in scores smaller than the MCID are considered as not important, regardless of whether statistical significance is reached or not. A recently developed taxonomy by the Outcome Measures in Rheumatology (OMERACT) group[17] suggests that there are multiple types of MCID for a given measure,

From the Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; Division of Pediatric Immunology and Rheumatology, Children's Memorial Hospital, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; and the School of Medicine, University of Cincinnati, Cincinnati, Ohio, USA.

Supported by a Clinical Research Grant of the Arthritis Foundation of America and NIAMS P60 AR47784.

H.I. Brunner, MD, MSc; A.L. Johnson, BS; A. Barron, BA; D.J. Lovell, MD, MPH; E.H. Giannini, MSc, DrPH, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati; M.S. Klein-Gitelman, MD, MPH; M.J. Miller, MD, Division of Pediatric Immunology and Rheumatology, Children's Memorial Hospital, Feinberg School of Medicine, Northwestern University; N. Baldwin, BS; M. Trombley, BS, School of Medicine, University of Cincinnati.

Address reprint requests to Dr. H. Brunner, Division of Rheumatology, E 4010, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039.
E-mail: hermine.brunner@cchmc.org

Submitted November 6, 2003; revision accepted July 23, 2004.

depending on the external standards considered and the characteristics of the population assessed.

There is only one study[18] using primarily hypothetical scenarios to determine the MCID of the CHAQ. Findings of this study support that the MCID depend on whether improvement or worsening is being considered and also how disabled patients are at baseline. For hypothetical changes in disability, the MCID of the CHAQ for improvement was often around –0.13 and for worsening at roughly +0.75. However, the MCID of hypothetical changes are different from and often larger than the MCID of changes actually experienced[19-23]. Good knowledge about the MCID of the CHAQ is essential for the interpretation of changes of CHAQ scores in clinical practice and research. For example, therapeutic interventions targeted to improve patient physical function that fail to result in a minimal important improvement of the CHAQ, especially if expensive, should be abandoned, given the lack of a relevant benefit to the child.

The objective of this study was to estimate the MCID of the CHAQ for children who are actually experiencing changes in their health and well being, using the taxonomy proposed by OMERACT[17].

MATERIALS AND METHODS

Study subjects. A convenience sample of families of children with JRA was recruited during routine clinic visits if the patient was between one and 18 years of age and had symptoms of chronic arthritis for at least 2 months. The primary caretaker and children aged 8 years and older completed the questionnaires. Each child's rheumatologist was asked to assess the patient's disease activity and change of arthritis since the preceding (study) visit after they had performed the routine clinic examination.

Interview process. Study instruments were completed under direct supervision in random sequence. Parents and patients completed the questionnaires independently from each other. Each family was interviewed during 2 sequential visits to the rheumatology clinics.

Childhood Health Assessment Questionnaire. The CHAQ consists of 2 components: disability and discomfort. Disability is assessed using 30 questions in 8 domains covering major aspects of daily living over a one-week period: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities. If aids or devices are used or assistance is required, the minimal score for the corresponding domain is 2. Each domain contains at least one item that is developmentally appropriate for children according to their age. Items are rated on a 4-point Likert scale (no difficulty, some difficulty, much difficulty, unable to do), with the option to mark "not applicable" if a child cannot be expected to perform a certain maneuver because of young age. The disability index is calculated as the unweighted average of the 8 domain scores and yields a disability score between 0 (no disability) and 3 (most severe disability). Because of the algorithm underlying the calculation of the CHAQ disability score, the smallest potential difference in CHAQ scores, i.e., the smallest possible incremental change of the CHAQ in an individual who changed by one category on the Likert scale in only one of the 30 items is 0.125. The CHAQ, when used in children with arthritis, has an important flooring effect, but generally good internal consistency and test-retest reliability[1,24].

Additional patient outcomes. For comparison and to serve as external standards, information on other patient-related outcomes was collected, as follows.

The Juvenile Arthritis Quality of Life Questionnaire (JAQQ)[5,6] is a disease-specific measure of health related quality of life (HRQOL). This measure was included in the study to serve as an alternative core response variable of functional ability in the JRA core set (see below) to assess the MCID of the CHAQ. The JAQQ consists of 74 items grouped into 4 domains, and patients are requested to consider the preceding 2 weeks: gross motor function, fine motor function, psychosocial function, and systemic symptoms. Each item is scored on a 6-point Likert scale (none of the time – never; hardly any time – 10% of the time; some of the time – 25% of the time; half of the time – 50% of the time; most of the time – 75% of the time; almost all of the time – 90% of the time; all the time – always) with the option to answer "does not apply to me/my child" for items that are not expected from children because of their young age. Domain scores are calculated based on the unweighted average of the 5 highest scored items in the domain. The summary JAQQ score corresponds to the unweighted average of the 4 domain scores[25]. The JAQQ has excellent reliability, construct validity, and responsiveness to change when used in children with JRA and other MSK diseases. To increase the comparability with other patient outcomes in this study, the final scores of the JAQQ were rescaled to range between 0 and 1, 1 being the best possible score, signifying the highest HRQOL.

Patient well being. For this study, parents and patients were asked to rate the patient's well being on 3 different scales. (1) We measured patient well being for the preceding one week using a 100 mm double-anchored linear analog scale presented with the sentence stem, "My/my child's overall well being is..." The lower endpoint of the scale was marked with "extremely bad" and a sad-face, whereas the upper endpoint of the scale was defined as "excellent" and a happy-face was presented ($Well_{linear}$). (2) In addition, an 11-point Likert scale (range 0–10) of patient well being ($Well_{Cat\ num}$) was completed, using the same sentence stem and endpoints as for $Well_{linear}$. (3) As a third approach to measuring well being, a 5-point Likert scale was completed (much worse, worse, same, better, much better), which was presented with the sentence stem, "Relative to the last assessment do you feel you/your child is..." ($Well_{Category}$).

Disease activity. The treating physician rated the patient disease activity on 2 scales. (1) Physicians completed a double-anchored linear analog scale of 100 mm length to assess global disease activity (DA_{linear}). The lower endpoint was marked "0 = very mild," while the upper endpoint was marked "10 = very severe." The scale was presented with the word stem, "The patient's disease activity is..." (2) In addition, physicians completed an 11-point Likert scale (range 0–10), which was presented with the same sentence stem as $DA_{Cat\ num}$. The endpoints of this scale were marked identically to those of the DA_{linear}.

Clinically important improvement of JRA. Standard definitions of patient improvement have been developed in the past[9] and are based on changes of the so-called JRA core response variables (CRV). The core set of CRV are the patient global assessment of well being, the physician assessment of disease activity, the number of joints with active arthritis (AJC), the number of joints with limited range of motion (LROM), a laboratory marker of inflammation [erythrocyte sedimentation rate (ESR) or C-reactive protein], and a measure of patient function. Patients with JRA are considered to be clinically improved if a minimum of 3 of the 6 CRV improve by at least 30% without more than one of the remaining CRV being worse by 30% or more.

Although not strictly defined, the CHAQ, a measure of physical function, has often been used as a functional measure in the core set. For determining the MCID of the CHAQ based on relative changes of the CRV, the following measures were used: the parent rating of the JAQQ as a measure of overall function, the AJC, the number of joints with LROM, the parent rating on $Well_{linear}$ to measure well being, DA_{linear} for measuring disease activity, and the ESR as a laboratory marker of disease.

Clinically important worsening of JRA (flare). The preliminary criteria of clinically important worsening in JRA are based on percentage changes of the CRV[10]. Patients are considered as having a disease flare, i.e., experiencing a clinically important worsening of JRA, if at least 2 of 6 CRV wors-

en by at least 40% without more than one of the remaining CRV improving by 30% or more. For the determination of disease flare in this study, the same measures were used as CRV as for assessing clinically important improvement.

MCID taxonomy of the OMERACT. The MCID of a measure is a special parameter of its responsiveness. The OMERACT presented a taxonomy describing the types of discrimination that could be considered when assessing the MCID of a measure[17,26]. According to this taxonomy, there are different types of MCID for a given instrument, which can all be classified based on 3 mutually independent key categories, called axes (Figure 1), as follows.

Axis 1, "Setting." This axis describes the study setting and whether the results are geared toward interpreting the scores of individual patients or the scores of a group of patients. Thus the "Setting" axis has 2 basic features: MCID for assessing changes of individuals and MCID for assessing changes of groups. This distinction is important, because it has been shown that clinical experts demand more change in an individual's score before they confidently consider an important change to have occurred than in a group of patients' scores[27].

Axis 2, "Which." The "Which" axis defines which scores are being contrasted in a study. The bottom row of Figure 1 represents the discrimination at *one point* in time between persons as the scores between patients are contrasted. This contrast is then defined as the MCID[28]. Most studies providing information on MCID look at *within-person* change *over time*, say in a group of patients undergoing physiotherapy for arthritis[29]. The same can be done on a group level, i.e., whether one patient group is different (better or worse) from another group over time[30-33]. Another possible approach is to focus on the relative change seen in a treatment group versus a control group[34,35].

Axis 3, "Type of change/ difference." The third axis defines the type of change or difference that is being quantified in a study of responsiveness (horizontal side of cube in Figure 1). The *smallest potential difference* is one type of MCID and is for the CHAQ on an individual level at 0.125. The *smallest detectable difference* (SDD; also referred to as minimum detectable difference) is yet another MCID type. It is based on the standard error mean and is linked to the test-retest reliability of a measure[36]. The SDD can serve as an anchor for describing the MCID of a measure, and it has been proposed that it is difficult to interpret change scores of a measure smaller than its SDD[17]. Other approaches to determining the MCID have been put forward[29,36,37], such as defining the MCID as the mean change scores of those patients who experienced a small but important change. Because this method possibly misclassifies patients whose change scores are close to but not at the mean, it has been proposed to define the 25th or even the 5th percentiles of the change score as MCID to ensure greater sensitivity[27,28]. The remaining categories on this axis are all dependent on the different external standards considered when determining the MCID, such as patient self-reports, parent or physician proxy-reports, criteria of important improvement, or flare.

Types of MCID assessed in this study. MCID assessments for this study were done to contrast "changes within subjects over time" (axis 2) using the "individual" setting (axis 1) only. Multiple approaches were taken to calculate different types of MCID values of the CHAQ (axis 3). The types of

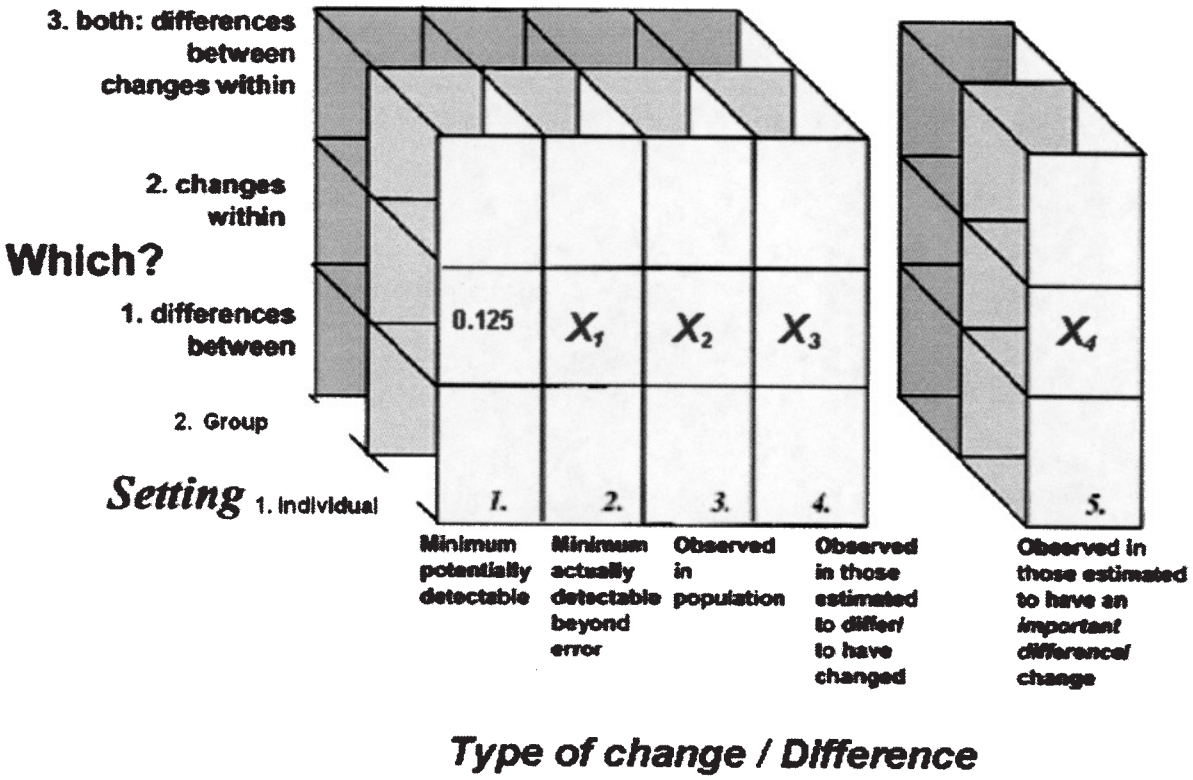
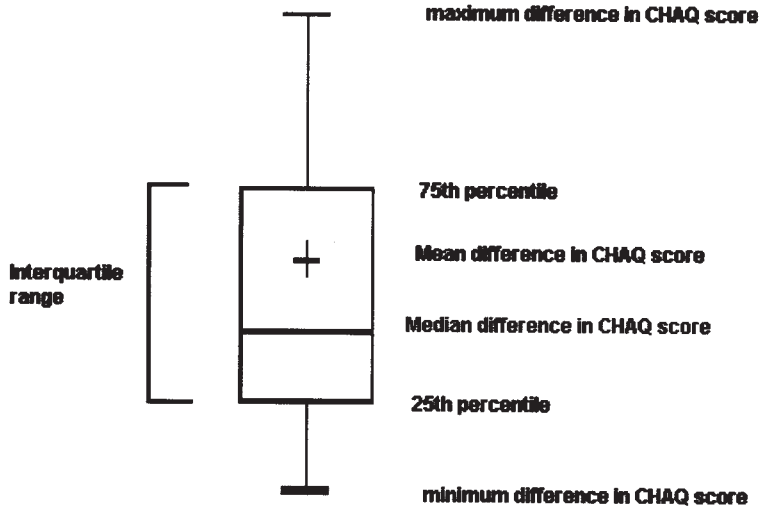


Figure 1. The different types of minimal clinically important differences (MCID) are depicted on the so-called MCID cube. These types of MCID have to be considered for improvement and worsening separately. Each type of MCID can be described on the basis of 3 key categories, called axes. They are the "Which" axis, the "Setting" axis, and the axis "Type of change/ Difference." Based on the algorithm used to calculate disability summary score, the smallest minimum potentially detectable difference in CHAQ score of a child at different points in time is 0.125 (Type of change/Difference axis: category 1; Setting: 1. Individual; Which: 2. changes within). Only MCID of the CHAQ marked X_1, X_2, X_3, and X_4 were assessed in the study, using various other patient-related outcomes as external standards. Because MCID for worsening and improvement of the CHAQ may differ X_1, X_2, X_3, and X_4 have to be assessed for worsening and also improvement separately (adapted with permission from J Rheumatol 2001;28:400-5).



X-axis descriptors – External Standards:

Abbreviation in Figures 2 - 4	Description of scale / External standard	MCID – type as depicted in figure 1
1	Child report of Well-being on scale Well $_{linear}$	X_2
2	Child report of Well-being on scale Well $_{Cat\ num}$	X_2
3	Child report of Well-being : "somewhat worse"	X_2
4	Parent report of Well-being on Well $_{linear}$	X_3
5	Parent report of Well-being on Well $_{Cat\ num}$	X_3
6	Parent report of Well-being : "somewhat worse"	X_3
7	Disease activity on scale DA $_{linear}$	X_3
8	Disease activity on scale DA $_{Cat\ num}$	X_3
9	Important change in disease based on core response variables	X_4

Legend: See Figures 2, 3, and 4.

MCID assessed in this study are marked "$X_{1–4}$" in Figure 1. MCID for improvement and worsening were assessed separately. The SDD corresponds to the area marked X_1 in Figure 1. Patient self-report of minimal important changes in health and well being were used to determine MCID, marked X_2 in Figure 1. Proxy-ratings of patient change in disease activity, health, and well being by parents and physicians yielded MCID marked X_3 in Figure 1. Important changes of patients were based on the criteria for improvement and flare, and are marked X_4.

Definition of external standards used in the study to measure the MCID of the CHAQ

SDD (Figure 1, X_1). The CHAQ scores of patients who were considered unchanged between visits were measured to determine the SDD using a data-driven approach. The SDD at a 95% confidence level (SDD_{95}) corresponds to the standard error mean (SEM) multiplied by the square root of 2 and multiplied by 1.96 ($SDD_{95} = SEM \times \sqrt{2 \times 1.96}$)[31]. Calculations of the SEM were based on the CHAQ scores of patients who were considered as not having changed (well being, disease activity, flare, and improvement criteria) between visits.

Definition of minimal important change in the external standards used to assess the MCID (Figure 1, X_{2+3}). The MCID values were based on median changes of the CHAQ scores, given the negative skewing of CHAQ scores in children with arthritis[18].

Important minimal absolute changes on linear analog scales ($Well_{linear}$, DA_{linear}) were defined as 10 mm to 30 mm differences in the ratings between visits. Patients with less than 10 mm changes on linear analog

scales between visits were rated as unchanged, while patients with changes beyond 30 mm were rated as extremely changed and excluded from the analysis for the MCID.

Similarly, changes on categorical numeric scales ($Well_{Cat\ num}$, $DA_{Cat\ num}$) of 1 or 2 categories were used to define patients with minimal important changes. Patients with more than a 2-category change between visits were rated as extremely changed and excluded from the subsequent analysis of the MCID.

When using the scale $Well_{Category}$, patients who rated themselves or were rated by others as "better," "worse," or "same" between visits were regarded as minimally improved, minimally worse, or unchanged, while patients rated as "much improved" or "much worse" were excluded from the analysis for the MCID of the CHAQ.

Assessment of CHAQ flooring effect on MCID. To assess whether patient disability influences the MCID values of the CHAQ, patients were categorized into disability groups according to their baseline CHAQ scores (parent-proxy ratings). Children were regarded as having *no disability* (CHAQ = 0), *mild disability* (0 < CHAQ < 0.250), *mild/moderate disability* (0.250 < CHAQ < 1.25), and *moderate disability* (1.25 < CHAQ), as suggested[18]. The MCID values were compared for significant differences between disability groups using Kruskal-Wallis nonparametric ANOVA and parametric ANOVA with post-hoc testing, as suggested[38].

RESULTS

Baseline patient characteristics. Ninety-two families were interviewed twice at the time of routine clinic visits. The mean age of the children with JRA with available parent interviews was 8.7 years (range 1–18) with an average disease duration of 5.3 years (range 0.5–16). All parents and also children aged 8 years and older (n = 67, mean age 12.5 yrs, SD 3.5) completed the study questionnaires (Table 1A). Arthritis of the studied patients was relatively well controlled, and parents rated 36% (n = 33) of them as having *no disability* (CHAQ = 0) at the time of enrollment (Table 1B). *Changes of patients between visits.* The average time between the 2 interviews was 3.5 months (SD 2.3). The proportion of patients categorized as having minimally worsened, minimally improved, remaining clinically unchanged, or experiencing larger than minimal changes was dependent on the scales used (Table 2). When considering the criteria for improvement and flare[9,10], 17% improved importantly, 15% experienced a disease flare, while the remaining 68% of patients were clinically unchanged. Of note is that patient self-ratings of well being yielded a similar categorization of the changes between visits.

Table 1B. Disability of the patients at baseline.

Disability Categories	No. of Children ≥ 8 yrs (%)	No. of Parents (%)
Moderate disability[1]	6 (10)	16 (17)
Mild/moderate disability[2]	25 (37)	29 (32)
Mild disability[3]	15 (22)	14 (15)
No disability[4]	21 (31)	33 (36)
Total	67 (100)	92 (100)

[1] 1.25 < CHAQ score ≤ 2.00. [2] 0.25 < CHAQ score ≤ 1.25. [3] 0 < CHAQ score ≤ 0.250. [4] CHAQ score = 0.

MCID of the CHAQ

Overview. Given the design of the study, MCID for the *individual setting* (axis 1) and *within-patient changes over time* (axis 2) were determined, while several external standards were used to obtain reference values for the different categories contained on axis 3. The various types of MCID determined in this study are described below and referred to as $X_{1–4}$ in Figure 1 to facilitate the comparison with the current OMERACT taxonomy. Based on minimal important worsening of the different external standards, the various MCID values for *worsening* ($X_{2–4}$) are summarized in Table

Table 1A. Outcome parameters of the cohort at the time of enrollment.

Outcome Measure	n	Observed Range	Median (IQR)	Mean (SD)
Childhood Health Assessment Questionnaire				
Child report (≥ 8 yrs)	67	0–2.25	0.25 (0–0.66)	0.46 (0.56)
Parent report	92	0–2.00	0.25 (0–0.91)	0.53 (0.61)
No. of joints with active arthritis		0–34	1 (0–4)	3.6 (6.6)
No. of joints with limited range of motion		0–27	0 (0–2)	2.2 (4.5)
ESR, mm/h		1–93	15 (7–25)	18 (16)
Juvenile Arthritis Quality of Life Questionnaire				
Child report (≥ 8 yrs)	66	0.30–1	0.84 (0.67–0.92)	0.78 (0.21)
Parent report	91	0.34–1	0.75 (0.64–0.88)	0.74 (0.17)
Linear analog scale of patient well being (0–100 mm; $Well_{linear}$)				
Child report (≥ 8 yrs)	67	21–100	78 (57–88)	73 (20)
Parent report	92	20–100	80 (68–89)	76 (18)
Categorical scale of patient well being (0–10; $Well_{Cat\ num}$)				
Child report (≥ 8 yrs)	67	2–10	5 (4–7)	5.4 (2.3)
Parent report	92		5 (4–8)	5.6 (2.2)
Linear analog scale of disease activity by physician (0–100 mm; DA_{linear})	92	0–95	28 (13–52)	34 (26)
Categorical scale of disease activity by physician (0–10; $DA_{Cat\ num}$)	92	0–10	2 (2–4)	2.8 (2.0)

Table 2. Changes in patients between visits based on parent and physician reports.

Measure/External Standard	Minimal Important Improvement, %	Minimal Important Worsening, %	Unchanged, %	Very Much Improved or Very Much Worse, %
Child report				
1. Well being on scale $Well_{linear}$*	14	9	59	18
2. Well being on scale $Well_{Cat\ num}$**	25	16	49	10
3. Rating on scale $Well_{category}$***	31	10	41	18
Parent report				
4. Well being on scale $Well_{linear}$*	13	26	59	2
5. Well being on scale $Well_{Cat\ num}$**	27	12	59	2
6. Rating on scale $Well_{category}$***	23	21	38	18
Physician report and CRV				
7. Disease activity scale DA_{linear}*	25	13	44	18
8. Disease activity on scale $DA_{Cat\ num}$**	29	16	43	12
9. Important change of disease based on current criteria of flare and improvement using percentage changes of CRV†	17	15	68	

* Minimal important change is defined as an absolute change on the linear analog scale of well being by 10 to 30 mm between visits. Patients with < 10 mm on the scale between visits were rated as unchanged, while patients with changes > 30 mm were rated as extremely changed (very much improved or very much worse) and excluded from the analysis for the MCID. ** Minimal important change is defined as an absolute change on the numeric Likert scale (0–10) by 1 or 2 categories between visits. Patients rated with the same numeric category were regarded as unchanged between visits, while patients with changes > 2 categories were regarded as largely changed (very much improved or very much worse) and excluded from the analysis for the MCID. *** Patients rated "better" were considered minimally improved, those rated "worse" were considered minimally worse, while patients rated "same" were regarded as unchanged between visits. Patients rated "much better" or "much worse" were excluded from the analysis of the MCID. † *Clinically important improvement* is present if a minimum of 3 of the 6 core response variables (CRV) improve by at least 30% without more than one of the remaining CRV being worse by 30% or more; *clinically important worsening* is present if a minimum of 2 of the 6 CRV worsen by at least 40% without more than one of the remaining CRV being improved by 30% or more.

3A and depicted in Figure 2. The various types of MCID for *improvement* ($X_{2–4}$) are summarized in Table 3B and depicted in Figure 3. Figure 4 depicts the MCID of the CHAQ for *improvement under consideration of the flooring effect of the CHAQ* (exploratory analysis). The data-driven approach for determining the MCID of the CHAQ (SDD_{95};X_1) is presented in Table 3C together with the changes of CHAQ scores in stable patients.

MCID of the CHAQ for worsening (Figure 1, $X_{2–4}$). The MCID of the CHAQ for worsening corresponds to the smallest decrease in CHAQ scores of patients who have worsened based on relevant changes of external standards. As shown in Table 3A, the MCID values differed depending on the external standard considered. Of note is that physicians often rated their patients as worse, even if patient physical function (CHAQ) had improved. Similarly, disease flare was often not accompanied by changes in CHAQ score.

Irrespective of the external standard used, the MCID of the CHAQ for worsening was small and close to or at the level of the smallest possible difference (Figure 2).

MCID of the CHAQ for improvement (Figure 1, $X_{2–4}$). Similar to worsening, the MCID of the CHAQ for improvement was often small and did not exceed –0.188 (Table 3B).

Table 3A. Minimal clinically important differences (MCID) of the CHAQ for worsening. See Table 2 legend for definitions of minimal worsening of the external standards 1–9 and Figure 1 for the different MCID types.

	Rater	External Standard	Median (IQR)	Mean (SD)
Observed in population (X_2)	Child	1. Worsening of well being on scale $Well_{linear}$ by 10–30 mm	+ 0.125 (0.25)	+ 0.063 (0.217)
		2. Worsening of well being on scale $Well_{Cat\ num}$ by 1–2 categories	0 (0.375)	–0.167 (0.395)
		3. Rating as "somewhat worsened" on scale $Well_{category}$	+ 0.25 (0.75)	+ 0.25 (0.375)
Observed in those estimated to differ /to have changed (X_3)	Parent	4. Worsening of well being on scale $Well_{linear}$ by 10–30 mm	0 (0.25)	–0.113 (0.579)
		5. Worsening of well being on scale $Well_{Cat\ num}$ by 1–2 categories	+ 0.063 (0.5)	0 (0.702)
		6. Rating as "somewhat worsened" on scale $Well_{category}$	+ 0.125 (0.75)	+ 0.238 (0.419)
	Physician	7. Worsening of disease activity on DA_{linear} by 10–30 mm	– 0.125 (0.375)	–0.2 (0.736)
		8. Worsening of disease activity on $DA_{Cat\ num}$ by 1–2 categories	–0.125 (0.5)	–0.231 (0.618)
Observed in those *important* change (X_4)	CRV	9. Observed in patients who fulfill criteria of important disease worsening based on changes in CRV	0 (0.250)	–0.102 (0.530)

Table 3B. Types of MCID for improvement assessed in the study. See Table 2 legend for definitions of minimal worsening of the external standards 1–9 and Figure 1 for the different MCID types.

	Rater	External Standard	Median (IQR)	Mean (SD)
Observed in population (X_2)	Child	1. Improvement of well being on scale $Well_{linear}$ by 10–30 mm	–0.063 (1.188)	–0.016 (0.766)
		2. Improvement of well being on scale $Well_{Cat\ num}$ by 1–2 categories	0 (0.875)	–0.205 (0.642)
		3. Rating as "somewhat improved" on scale $Well_{category}$	–0.188 (0.5)	–0.188 (0.418)
Observed in those estimated to differ/to have changed (X_3)	Parent	4. Improvement of well being on scale $Well_{linear}$ by 10–30 mm	0 (0.875)	+ 0.125 (0.407)
		5. Improvement of well being on scale $Well_{Cat\ num}$ by 1–2 categories	0 (1.0)	–0.272 (0.688)
		6. Rating as "somewhat improved" on scale $Well_{category}$	0 (0.125)	–0.023 (0.273)
	Physician	7. Improvement of disease activity on scale DA_{linear} by 10–30 mm	0 (0.125)	–0.118 (0.381)
		8. Improvement of disease activity on scale $DA_{Cat\ num}$ by 1–2 categories	0 (0.375)	–0.170 (0.459)
Observed in those *important* change (X_4)	CRV	9. Observed in patients who fulfill criteria of important disease improvement based on changes in CRV	+ 0.125 (0.875)	–0.115 (0.662)

Table 3C. Changes of the CHAQ scores of stable, unchanged patients and smallest detectable differences (SDD). See Table 2 legend for definitions of minimal worsening of the external standards 1–9 and Figure 1 for the different MCID types.

	Rater	External Standard	Median (IQR)	Mean (SD)
Observed in population (X_2)	Child	1. Well being on Well $being_{linear}$ changes < 10 mm	0 (0.375)	–0.117 (0.390)
		2. Well being on Well $being_{num}$ is the same	–0.063 (0.375)	–0.125 (0.374)
		3. Rating "unchanged" on Well $being_{category}$	0 (0)	–0.087 (0.324)
Observed in those estimated to differ/to have changed (X_3)	Parent	4. Well being on Well $being_{linear}$ changes < 10 mm	0 (0.250)	+ 0.003 (0.417)
		5. Well being on Well $being_{num}$ is the same	0 (0.125)	–0.005 (0.354)
		6. Rating "unchanged" on Well $being_{category}$	–0.125 (0.250)	-0.139 (0.273)
	Physician	7. Change of disease activity on DA_{linear} by < 10 mm	0 (0.125)	0.015 (0.446)
		8. Change of disease activity on DA_{Likert} is the same	0 (0.094)	0.022 (0.480)
Observed in those with *important* change (X_4)	CRV	9. Observed in patients who did not fulfill criteria of important disease worsening or improvement based on changes in CRV	0 (0.250)	–0.036 (0.517)
Minimum actually detectable difference (X_1)	Data-driven	SDD at a 95% confidence level (SDD_{95}) based on external standards 1–9	Range 0.136–0.211	

Patients who had experienced a clinically important improvement (external standard 9) between visits often did not experience the expected decrease in CHAQ scores.

Changes of CHAQ scores in stable patients and the smallest detectable difference, SDD_{95} (Figure 1, X_1). The CHAQ scores of stable or clinically unchanged patients were often at a median of 0 (Table 3C). Nonetheless, some of the stable patients experienced unexpected large changes in their CHAQ scores. Changes of CHAQ score observed in stable patients were used to determine the SDD_{95}[36] (Figure 1, X_1).

Depending on the external standard considered, the SDD_{95} of the CHAQ was between 0.136 (Parent $Well_{Categorical}$) and 0.211 ($DA_{Cat\ num}$) (Table 3C).

Exploratory analysis: MCID for improvement after correction for the flooring effect of the CHAQ (Figure 1, $X_{2–4}$). Patients with a CHAQ score of 0 at baseline have no option to further improve the rating of their physical function, even if other outcomes (external standards) suggest that the patients have improved. Given the profound flooring effect of the CHAQ, this may have an important negative influence on the absolute size of MCID for improvement. Exploratory analysis was performed and the MCID of the CHAQ for improvement was determined after exclusion of all families who improved based on the external standards, but had a CHAQ score of 0 at baseline (Figure 4). Due to small numbers, ratings from children were excluded from the exploratory analysis. The results support that the responsiveness, i.e., the size of the MCID of the CHAQ, improves with correction of the flooring effect. However, the MCID remains close to or at the level of the SDD_{95}.

Dependency of the MCID of the CHAQ on the baseline disability of the patients. Based on the results of nonparametric Kruskal-Wallis ANOVA, there were no important differences in the MCID of the CHAQ between patients of different disability groups, irrespective of the external standard used in this study (for all, chi-square < 5.8, p = nonsignificant). A similar result was obtained when repeating the analysis using a parametric ANOVA.

DISCUSSION

Knowledge of the MCID of a clinical measure is essential for the interpretation of changes in its scores. Only changes in scores beyond the MCID of a measure constitute relevant changes, while smaller changes are of minimal or no clini-

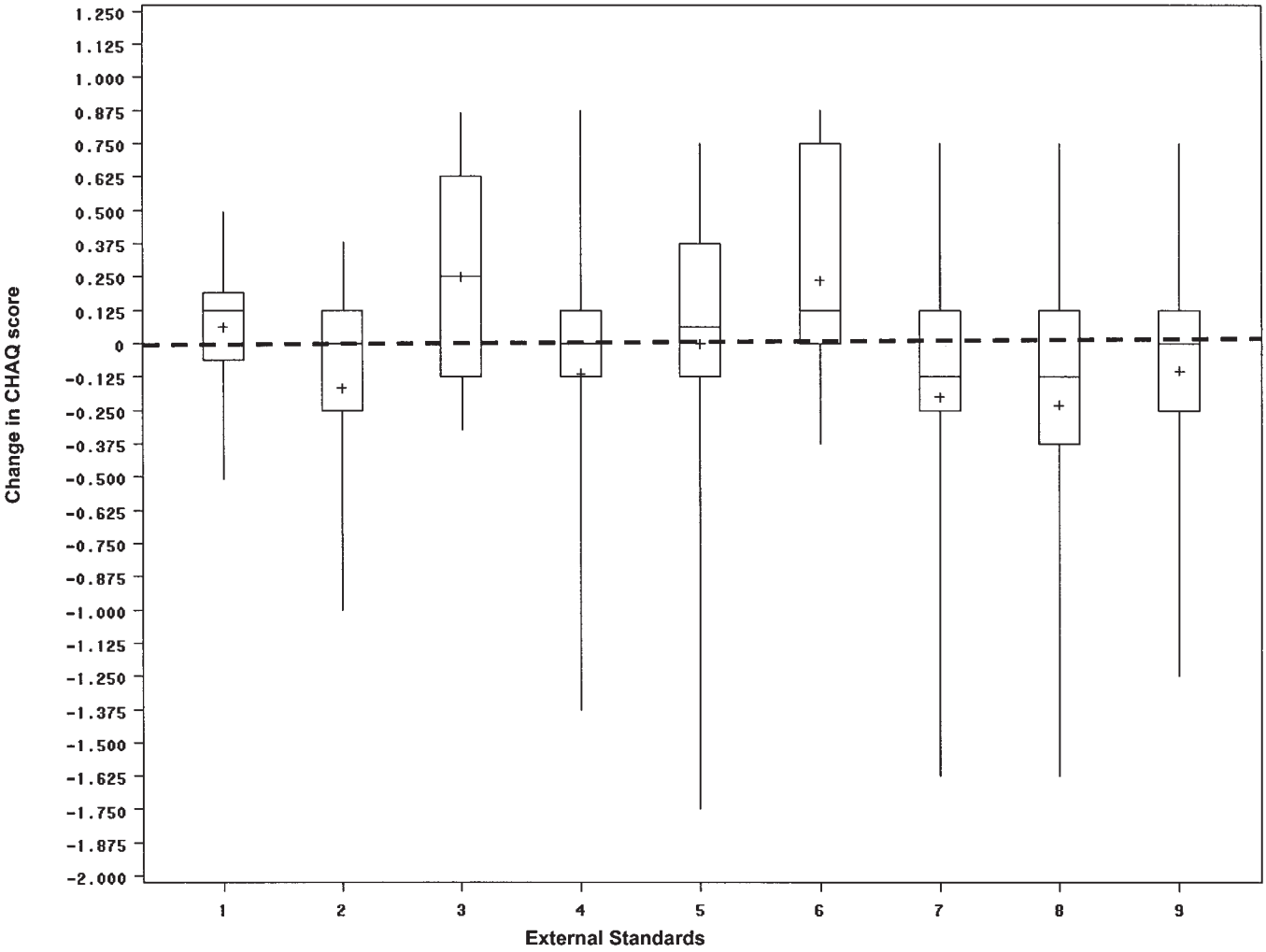


Figure 2. Minimal clinically important differences (MCID) of the CHAQ for worsening (see details of legend, page 153).

cal relevance. In recent years a series of outcome measures have been developed for JRA to help with the description and interpretation of changes in disease. The MCID of the CHAQ, one of the most commonly used of these outcome measures, have not been well examined. Therefore, it is currently unclear which changes in CHAQ scores of patients with JRA should prompt medical interventions and influence medical decision-making. In order to increase the usefulness of the CHAQ for clinicians and researchers, we examined the MCID of the CHAQ in children who actually experienced changes of their health and well being. We found that the MCID of the CHAQ for both worsening and improvement are different from each other. There appears to be no single MCID of the CHAQ, but rather a series of MCID depending on the external standards taken into account.

The results of this study suggest that the MCID of the CHAQ for worsening and improvement of individuals are generally small, and that CHAQ scores often do not change even though a minimal important change in the well being of children with JRA has occurred. Patients often experienced important changes in well being and disease activity without concomitant changes in their CHAQ scores. Thus the CHAQ is unlikely to be a very useful measure to help with short term medical decision-making for individual patients. Conversely, patients whose disability levels change, i.e., whose CHAQ scores decrease or increase, have probably experienced an important improvement or worsening of their health and disease (data not shown).

When using a data-driven approach, the MCID of the CHAQ is somewhat larger than when using the other approaches in our study, with a SDD_{95} of up to 0.211. The relatively large SDD_{95} is related to the large changes in CHAQ scores of some patients who considered themselves or were considered by others as clinically unchanged. However, different from other measures, such as the Short Form-36[39], there is a rather large difference between the SDD_{95} and the MCID of the CHAQ using other external

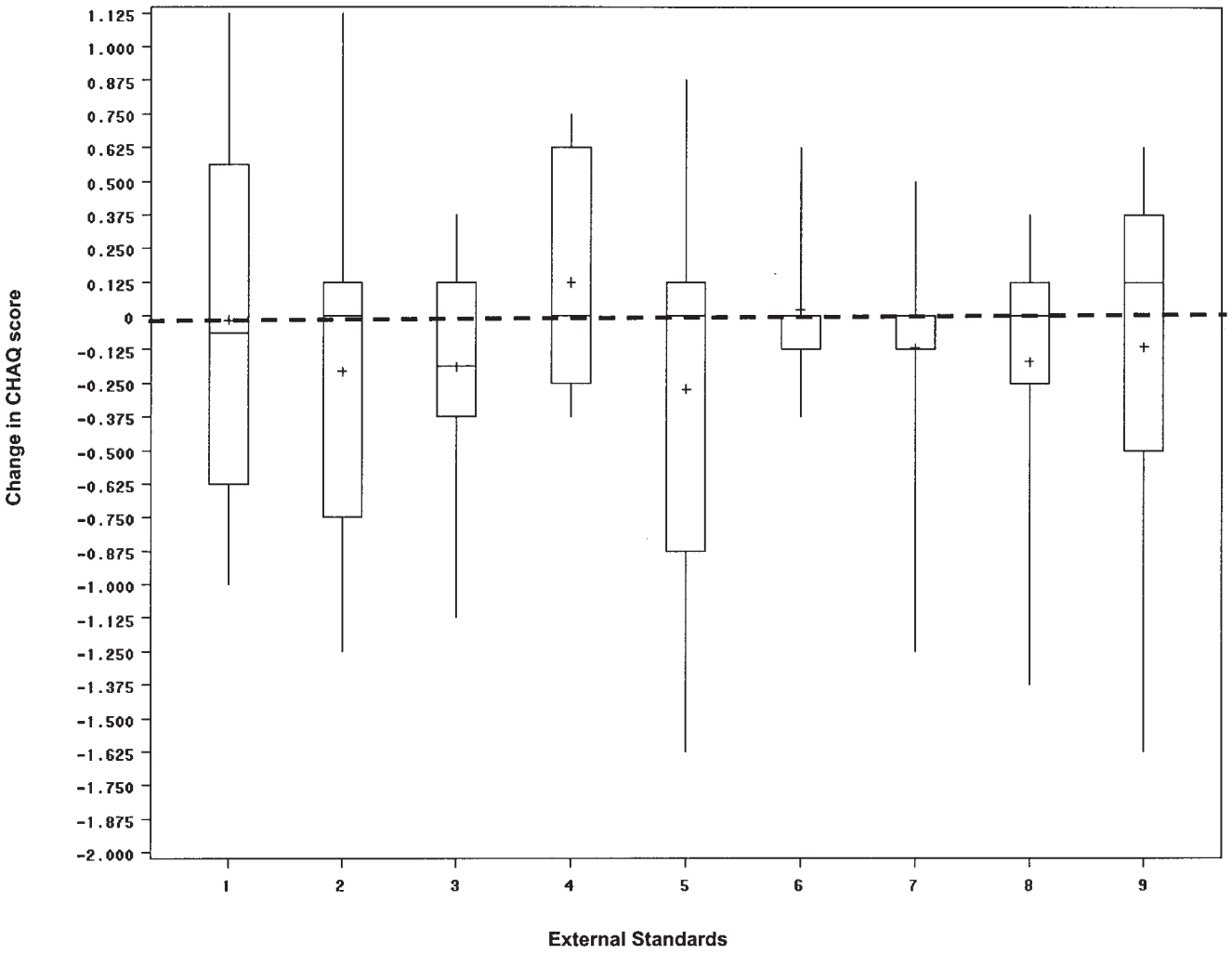


Figure 3. Minimal clinically important differences (MCID) of the CHAQ for improvement (see details of legend, page 153).

standards. The reasons for this are not clear. The SDD_{95} is known to be increased in measures with poor test-retest reliability. However, previous research suggests that the CHAQ has a high test-retest reliability, especially in patients used to the completion of the measure[1-3,40], like the subjects evaluated for this study. Other possibilities are that there were problems with the recall of symptoms or clinical findings of the families and physicians since the last visit. Similarly, physicians might have rated patients' disease activity as unchanged although an actual change in the patient had occurred that had led to a change in CHAQ score. Another explanation for the discrepancy between SSD_{95} and the other MCID values could be that a so-called "response shift"[19,22] has occurred. For example, a response shift has occurred if patients, although somewhat worsened or improved, have become used to the altered health state and rate themselves as unchanged, even though an actual change in their health had taken place. The assessment of response shift phenomena in pediatrics is still under development and was not the focus of our study. Thus further research is required to assess the effect of response shifts on the MCID and the SDD of the CHAQ.

In this study, we provided a conservative estimate for the MCID of the CHAQ by defining them as median change scores of those patients who changed. Even for these conservative estimates, the MCID of the CHAQ were small. If we had followed previously suggested more stringent standards and had defined the MCID as the 5th or 25th percentile of the change scores, then the MCID of the CHAQ would not have exceeded 0, irrespective of the external standard used in this study (Figures 2 and 3).

We speculate that, if used as CRV, the CHAQ may lead to an underestimation of clinically important improvement of patients, given the small MCID of the CHAQ for improvement. This is supported by our finding that patients who had improved importantly or worsened based on their changes of the other CRV often did not manifest the expected improvement or worsening in the CHAQ scores. The

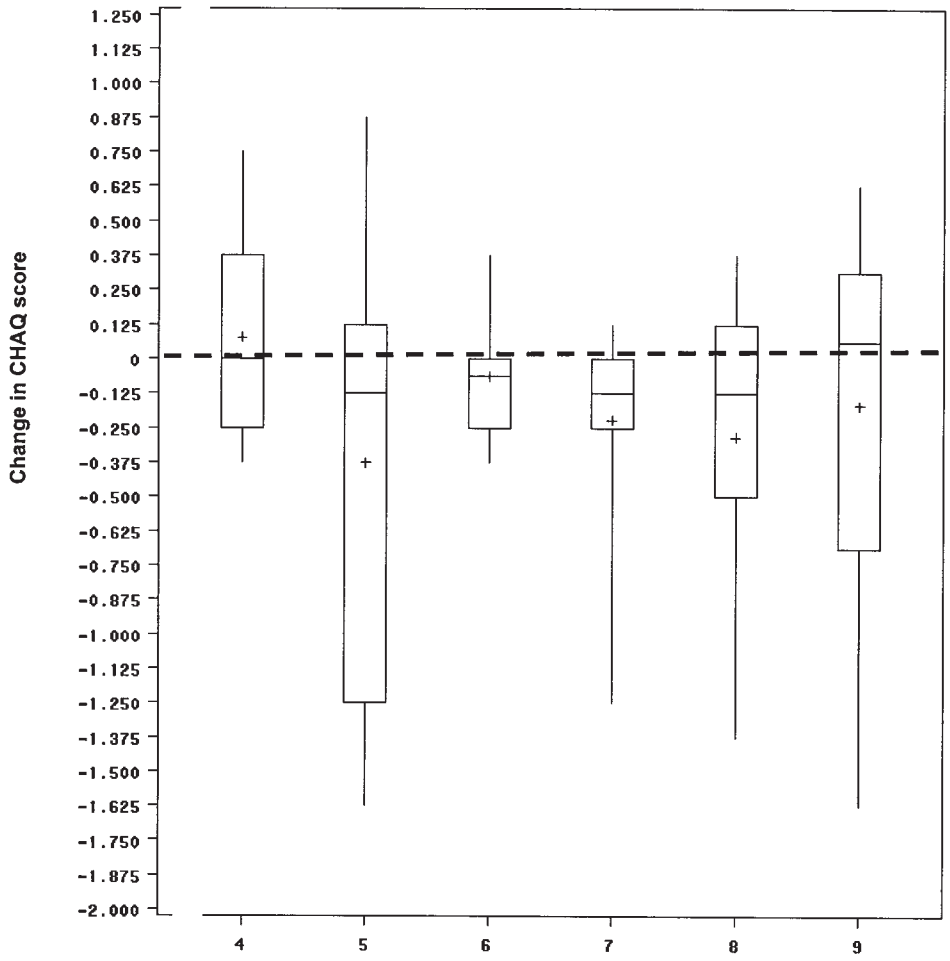


Figure 4. Minimal clinically important differences (MCID) of the CHAQ for improvement after correction for the CHAQ floor effect (see details of legend, page153).

observation that CHAQ scores of patients are less responsive to important changes compared to the other CRV during a 3-month time interval has been made previously[41].

Our results do not support some earlier studies' finding that the CHAQ is a very responsive measure in JRA[1,2]. The small MCID of the CHAQ in our study supports reports that the measure has, at most, moderate responsiveness to short term changes in JRA[14]. We hypothesize that this may be at least partially due to a change of the "phenotype" of the average patient with JRA seen in clinical practice. The physical function (CHAQ score) of the patients examined in this study was very similar to that of other recently reported JRA cohorts[4]. Due to improved treatments for JRA, the average number of joints with active arthritis or with limited range of motion observed in recently reported cohorts has decreased, and the disability level of patients appears to have improved[18,42,43]. Based on parent reports, about one-third of the patients studied had *no disability* (CHAQ = 0; mean AJC = 1.3) at baseline. Because the CHAQ was unable to determine potential improvement in these patients, the MCID of the CHAQ for improvement may decrease. To test the hypothesis that the flooring effect of the CHAQ negatively influences the MCID of the measure, we performed exploratory analyses examining the MCID of the CHAQ for improvement after excluding patients who had a CHAQ score of 0 at baseline. We were able to confirm that the MCID of the CHAQ increased only if patients with preexisting disability (CHAQ > 0) are considered.

Our results support previous reports that the MCID of a measure for changes actually experienced in disease are often smaller that those for hypothesized ones[23,44-46]. The patients with important changes in our study (improvement, worsening) had much smaller MCID than similarly affected patients who were asked to provide the MCID of the CHAQ for hypothetical changes in health[18]. We were unable to confirm the proposed differences in the MCID of the CHAQ depending on the degree of disability of patients with JRA at baseline. This could be due to the limited number of subjects with higher degrees of disability in this study. Another possibility is that differences in MCID values for actual

changes in disease of patients with different levels of disability are too small to be detectable using the current number of subjects.

Given the widespread use and the cross-cultural validation of the CHAQ[4], every effort should be made to improve the usefulness of its current items. It would be especially important to increase the responsiveness of the CHAQ and its MCID. Although several such attempts have been made[47,48], Rasch analysis[49] of the CHAQ could be done to improve the measurement properties of the measure by identifying a better approach to calculating the CHAQ disability summary score. Additional studies are necessary to assess the longterm responsiveness/MCID of the measure. It is possible that the CHAQ remains a good measure of changes in health and disease of patients with JRA over a longer time period. Research is also required to examine the MCID of the CHAQ for comparing groups of patients. This would be especially relevant for the interpretation of clinical trial data of JRA patients receiving different treatments. It is possible that, on a group level, the CHAQ has a sufficiently large MCID to allow a reasonable interpretation of changes in CHAQ scores.

We have determined the MCID of the CHAQ to enhance the interpretability of changes in CHAQ scores over time. Based on the small MCID for changes actually experienced by children with arthritis, the CHAQ in its current form may be too insensitive to determine important short term changes in health and disease for a given patient.

REFERENCES

1. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum 1994; 37:1761-9.
2. Feldman BM, Ayling-Campos A, Luy L, Stevens D, Silverman ED, Laxer RM. Measuring disability in juvenile dermatomyositis: validity of the childhood health assessment questionnaire. J Rheumatol 1995; 22:326-31.
3. Huber AM, Hicks JE, Lachenbruch PA, et al. Validation of the Childhood Health Assessment Questionnaire in the juvenile idiopathic myopathies. Juvenile Dermatomyositis Disease Activity Collaborative Study Group. J Rheumatol 2001; 28:1106-11.
4. Ruperto N, Ravelli A, Pistorio A, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. Clin Exp Rheumatol 2001; 19:S1-9.
5. Duffy CM, Arsenault L, Duffy KN. Level of agreement between parents and children in rating dysfunction in juvenile rheumatoid arthritis and juvenile spondyloarthritides. J Rheumatol 1993; 20:2134-9.
6. Duffy CM, Arsenault L, Duffy KN, Paquin JD, Strawczynski H. The Juvenile Arthritis Quality of Life Questionnaire—development of a new responsive index for juvenile rheumatoid arthritis and juvenile spondyloarthritides. J Rheumatol 1997; 24:738-46.
7. Howe S, Levinson J, Shear E, et al. Development of a disability measurement tool for juvenile rheumatoid arthritis. The Juvenile Arthritis Functional Assessment Report for Children and their Parents. Arthritis Rheum 1991; 34:873-80.
8. Wright FV, Kimber JL, Law M, Goldsmith CH, Crombie V, Dent P. The Juvenile Arthritis Functional Status Index (JASI): a validation study. J Rheumatol 1996; 23:1066-79.
9. 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.
10. Brunner HI, Lovell DJ, Finck BK, Giannini EH. Preliminary definition of disease flare in juvenile rheumatoid arthritis. J Rheumatol 2002; 29:1058-64.
11. Petty RE, Southwood TR, Baum J, et al. Revision of the proposed classification criteria for juvenile idiopathic arthritis: Durban, 1997. J Rheumatol 1998; 25:1991-4.
12. Brewer EJ, Jr., Bass J, Baum J, et al. Current proposed revision of JRA Criteria. JRA Criteria Subcommittee of the Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Section of The Arthritis Foundation. Arthritis Rheum 1977; 20:195-9.
13. Andersson Gare B, Ruperto N, Berg S, et al. The Swedish version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). Clin Exp Rheumatol 2001; 19:S146-50.
14. Ravelli A, Viola S, Migliavacca D, Pistorio A, Ruperto N, Martini A. Discordance between proxy-reported and observed assessment of functional ability of children with juvenile idiopathic arthritis. Rheumatology (Oxford) 2001; 40:914-9.
15. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989; 10:407-15.
16. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40:171-8.
17. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol 2001; 54:1204-17.
18. Dempster H, Porepa M, Young N, Feldman BM. The clinical meaning of functional outcome scores in children with juvenile arthritis. Arthritis Rheum 2001; 44:1768-74.
19. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. Qual Life Res 2003; 12:239-49.
20. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol 2002; 14:109-14.
21. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999; 48:1531-48.
22. Postulart D, Adang EM. Response shift and adaptation in chronically ill patients. Med Decis Making 2000; 20:186-93.
23. Llewellyn-Thomas HA, Thiel EC, McGreal MJ. Cancer patients' evaluations of their current health states: the influences of expectations, comparisons, actual health status, and mood. Med Decis Making 1992; 12:115-22.
24. Norrby U, Nordholm L, Fasth A. Reliability and validity of the Swedish version of child health questionnaire. Scand J Rheumatol 2003; 32:101-7.
25. Duffy CM, Tucker L, Burgos-Vargas R. Update on functional assessment tools. J Rheumatol 2000; 27 Suppl 58:11-4.
26. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. J Rheumatol 2001; 28:400-5.
27. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993; 20:561-5.
28. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Arch

Intern Med 1993; 153:1337-42.
29. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. Phys Ther 1998; 78:1186-96.
30. Terwee CB, Dekker FW, Mourits MP, et al. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. Clin Endocrinol (Oxf) 2001; 54:391-8.
31. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993; 20:557-60.
32. Redelmeier DA, Goldstein RS, Min ST, Hyland RH. Spirometry and dyspnea in patients with COPD. When small differences mean little. Chest 1996; 109:1163-8.
33. Redelmeier DA, Bloch DA, Hickam DH. Assessing predictive accuracy: how to compare Brier scores. J Clin Epidemiol 1991; 44:1141-6.
34. Hemingway H, Stafford M, Stansfeld S, Shipley M, Marmot M. Is the SF-36 a valid measure of change in population health? Results from the Whitehall II Study. BMJ 1997; 315:1273-9.
35. Ross M. Relationship of implicit theories to the construction of personal histories. Psychol Rev 1989; 96:341-57.
36. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999; 37:469-78.
37. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. Phys Ther 1998; 78:1197-207.
38. Zar JH. Biostatistical analysis. Upper Saddle River, N.J.: Prentice-Hall, 1996:1 v. (various pagings).
39. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999; 52:861-73.
40. 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:S91-5.
41. Lovell DJ, Giannini EH, Reiff A, et al. Etanercept in children with polyarticular juvenile rheumatoid arthritis. Pediatric Rheumatology Collaborative Study Group. N Engl J Med 2000; 342:763-9.
42. Oen K, Malleson PN, Cabral DA, Rosenberg AM, Petty RE, Cheang M. Disease course and outcome of juvenile rheumatoid arthritis in a multicenter cohort. J Rheumatol 2002; 29:1989-99.
43. Brunner HI, Kim KN, Ballinger SH, et al. Medication choices in juvenile rheumatoid arthritis. J Clin Rheumatol 2001;7:295-300.
44. Llewellyn-Thomas HA, Schwartz CE. Response shift effects in patients' evaluations of health states: The influences of expectations, comparisons, actual health status, and mood. Amer Psychol Assoc Press. Washington, D.C., 2000:109-122.
45. Llewellyn-Thomas HA, Sutherland HJ, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? Med Care 1993; 31:1002-12.
46. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996; 66:3-8.
47. Feldman B, Lam C, J M, Young NL. Ability vs. Disability: a re-scaled Childhood Health Assessment Questionnaire (CHAQ) performs better and is more sensitive [abstract]. Arthritis Rheum 2001;44 Suppl: S381.
48. Whitney-Mahoney KJ, Young NL, Feldman BM, Badely EM. Alternative scoring methods improve the responsiveness of the Childhood Health Assessment Questionnaire [abstract]. Arthritis Rheum 2001;44 Suppl:S381.
49. Rasch G. An item analysis which takes individual differences into account. Br J Math Stat Psychol 1966; 19:49-57.