

Responsiveness of the ABILHAND Questionnaire in Measuring Changes in Rheumatoid Arthritis Patients

CHARLES SÈBIYO BATCHO,¹ PATRICK DUREZ,² AND JEAN-LOUIS THONNARD¹

Objective. ABILHAND is a Rasch-built questionnaire that measures manual ability in rheumatoid arthritis (RA) patients. This study aimed to examine the test–retest reliability and the responsiveness of ABILHAND in RA patients.

Methods. Eighty-eight patients underwent 3 evaluations: the first evaluation was at baseline (time 1), the second was 2 weeks later (time 2), and the third was 1 year later (time 3). Disease activity was assessed using the Disease Activity Score in 28 joints using the C-reactive protein level (DAS28-CRP). Patients rated the intensity of their RA-related pain using a 100-mm visual analog scale for pain and completed questionnaires based on their activity limitations (ABILHAND and the Health Assessment Questionnaire) and quality of life.

Results. The responsiveness analyses were conducted by using global, group, and individual approaches. The global approach showed significant differences between the time 1 and time 3 scores of the DAS28-CRP ($P = 0.04$) and ABILHAND ($P = 0.04$). Based on the changes in disease activity scores and the European League Against Rheumatism response criteria, the sample was divided into 3 groups: deteriorated, stable, and improved. The mean $\pm$ SD changes in manual ability were higher in the deteriorated (-1.23 ± 1.53 logit) and in the improved (1.22 ± 2.06 logits) groups than in the stable group (0.48 ± 1.09 logit). The effect size and standardized response mean confirmed that observation. The minimal clinically important difference was assessed in each group of patients.

Conclusion. The ABILHAND questionnaire exhibited responsiveness in detecting slight changes in RA patients. Therefore, the ABILHAND tool can be used to evaluate the functional status of RA patients in clinical trials and settings.

INTRODUCTION

Rheumatoid arthritis (RA) is the most prevalent inflammatory rheumatic condition, and it leads to chronic pain and progressive disability. Due to the destruction of peripheral joints and pain in the upper extremities, RA frequently compromises manual ability, which is the capacity to manage daily activities using the upper extremities (1–3). It is important that clinical practices assess the degree of manual disability in patients with RA in order to evaluate the patients' impairments in their daily activities.

The ABILHAND questionnaire, a Rasch-built questionnaire that measures manual ability, has recently been adapted and validated for individuals with RA (4). The questionnaire consists of 27 specific items divided into 3 categories. The ABILHAND questionnaire has the required psychometric properties of modern test theory, including construct validity, test–retest reliability, linearity, and unidimensionality (4).

Self-report questionnaires that measure functional status have become important outcome measures when evaluating clinical changes in patients (5,6). Patient-reported outcomes are frequently incorporated in clinical trials to compare health interventions for chronic diseases (7). To be a useful indicator of intervention success, an outcome measure must be reliable, valid, and responsive (8). Responsiveness reflects the ability of a questionnaire to detect changes over time (9,10). It is considered an essential criterion in studies evaluating treatment effectiveness and economic feasibility (9,11).

The present study aimed to examine the test–retest reliability and the responsiveness of the ABILHAND questionnaire in a large cohort of patients experiencing chronic RA. Our investigation evaluated both the statistical and clinical significance of the ABILHAND questionnaire.

Supported by internal grants from the Université catholique de Louvain.

¹Charles Sèbiyo Batcho, MSc, Jean-Louis Thonnard, PhD: Université catholique de Louvain, Brussels, Belgium; ²Patrick Durez, MD: Cliniques universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium.

Address correspondence to Jean-Louis Thonnard, PhD, Institute of Neuroscience, Université catholique de Louvain, Avenue Mounier 53, 1200 Bruxelles, Belgium. E-mail: jean-louis.thonnard@uclouvain.be.

Submitted for publication March 30, 2010; accepted in revised form August 24, 2010.

PATIENTS AND METHODS

Population. Patients were recruited in a rheumatology outpatient clinic of the university hospital from February to May 2008. All RA patients who fulfilled the 1987 American College of Rheumatology (ACR) classification criteria for RA (12) and presented no comorbidities were approached at the time of their clinic appointment and were asked to participate in the study. A sample of 132 patients agreed and received sets of questionnaires. All participants were being treated with stable standard nonbiologic and biologic disease-modifying antirheumatic drugs. This study was approved by the ethical committee, and all of the participants gave their informed consent before their inclusion.

Patient assessment. Prior to the assessment, the patients were provided written instructions on how to complete the questionnaires. The patients were given a set of questionnaires on 3 different occasions with instructions to complete them at home and then return the questionnaires to the investigator's (CSB) address. The first evaluation was completed at time 1, the second was completed 2 weeks later (time 2), and the third evaluation was performed 1 year following the first evaluation (time 3). Clinical assessments based on the ACR classification criteria (12) and the European League Against Rheumatism (EULAR) core measurement set (13) were performed during medical visits at baseline (time 1) and 1 year later (time 3). Both assessments were performed by the same investigator (PD).

During the assessment, disease activity was measured using the Disease Activity Score in 28 joints using C-reactive protein level (DAS28-CRP). Patients rated the intensity of their RA-related pain using a 100-mm visual analog scale for pain (VAS pain). It was a double-anchored VAS consisting of a horizontal line (0–100 mm) where both anchor points represent opposite ends of a continuum. Patients were instructed to indicate the severity of their pain by a vertical mark on the line. The distance in millimeters, measured with a metric ruler, from the left anchor to the patient's mark represented the VAS pain score. Each patient filled out questionnaires that measured activity limitations (ABILHAND and the Health Assessment Questionnaire [HAQ]) and RA quality of life (RAQoL). The HAQ was scored with no adjustment for assistive devices, and patients' self-reports in ABILHAND were analyzed using online Rasch software, which allows the conversion of raw scores to interval measures. This conversion is based on a previous calibration of ABILHAND published previously (4). That study showed that the scale satisfies the Rasch model, and its 27 items fit the requirement of the model. In that same study, the fit statistic computed by chi-square test ranged from 0.16–7.11, and the associated *P* value ranged from 0.07–0.98. For each patient, sociodemographic data, such as age, sex, and disease duration, were collected during the assessment.

Data analysis. Reproducibility. The test–retest reliability of ABILHAND, HAQ, RAQoL, and VAS pain was eval-

uated to assess the stability of the responses over time, given that there was no change in the patient's condition. Data from the first and second evaluations were used for this purpose, since the time lapse between the 2 evaluations was short (2 weeks). The paired *t*-test and intraclass correlation coefficient (ICC) were used to evaluate differences between the responses (14).

Responsiveness. The primary focus of this study was to evaluate the responsiveness of the ABILHAND questionnaire to measure minimal, accurate, and clinically relevant changes in RA patients (9,10,15). This analysis was based on the most frequently cited indices in responsiveness studies (9–11,15–17), including the mean change, paired *t*-test, effect size (ES), and standardized response mean (SRM). The ES is calculated as the ratio between the mean change in score and the SD of the baseline scores (18). The SRM is calculated as the ratio between the mean change in score and the SD of the change in score (19). The magnitude of the change was interpreted using the ES and SRM following the guidelines proposed by Cohen (20). Hence, ES scores were used to define the magnitude of the change as not significant (<0.2), small ($0.2–0.49$), moderate ($0.5–0.79$), or large (≥ 0.8).

The analyses were conducted using 3 approaches: global, group, and individual. The global approach consisted of statistically comparing the baseline scores with those gathered at the 1-year followup. In the group approach, patients were categorized into 3 groups based on variations in their DAS28-CRP. Based on the EULAR response criteria (13), the difference (Δ) was calculated for each patient as follows: $\Delta = \text{DAS28-CRP}(\text{time 1}) - \text{DAS28-CRP}(\text{time 3})$. The patients were divided into 3 groups according to their variation in the DAS28-CRP: 1) stable if $-0.6 \leq \Delta \leq 0.6$; 2) improved if $\Delta > 0.6$; and 3) deteriorated if Δ less than -0.6 .

Since changes that have a meaning in groups may not be meaningful in individuals (21), we conducted a second analysis using an individual approach. Here, we evaluated changes in manual ability in each patient. The individual analysis took into account the SEs of the measurements and the mean change in each patient's measurements between the time 1 and time 3 evaluations. The following formula was applied (22):

$$T \text{ score} = \frac{m3 - m1}{\sqrt{(SE3)^2 + (SE1)^2}}$$

where *m1* and *m3* represent the patient's manual ability measurements, respectively, at baseline (time 1) and the 1-year followup (time 3), and *SE1* and *SE3* are their respective associated SEs of measurement. According to Wright and Stone, the *T* score approximately follows a standardized normal distribution, so a patient with a *T* score above 1.96 showed significant improvement, while a patient with a *T* score below -1.96 showed significant deterioration in manual activity (22). The patients were divided into 5 classes according to their *T* score significance limits: significantly deteriorated (*T* score less than -1.96), deteriorated ($-1.96 \leq T \text{ score} < 0$), no change (*T* score = 0), improved ($0 < T \text{ score} \leq 1.96$), and significantly improved (*T* score > 1.96).

Table 1. Baseline characteristics of the population*

Characteristics	Value
Age, mean $\pm$ SD (range) years	57.53 $\pm$ 11.87 (28–79)
Women, no. (%)	80 (76.19)
Men, no. (%)	25 (23.81)
RA duration, mean $\pm$ SD (range) years	13.13 $\pm$ 10.47 (0.5–59)
DAS28-CRP, median (IQR)	3.07 (2.2)
HAQ, median (IQR)	0.88 (1.25)
VAS pain, median (IQR)	36 (50)

* RA = rheumatoid arthritis; DAS28-CRP = Disease Activity Score in 28 joints using C-reactive protein level; IQR = interquartile range (difference between upper and lower quartiles; limits within which the middle 50% of observations fall); HAQ = Health Assessment Questionnaire; VAS pain = visual analog scale for pain.

Clinical significance of change. When a statistically significant effect exists, it is important to assess the clinical significance of the effect. Whereas statistical significance indicates the likelihood a difference is caused by chance, clinical significance identifies whether the difference is large enough to affect patient care (23). There are several methods available for assessing clinical significance (24,25). The 2 principal methods used in this study were the minimal clinically important difference (MCID) and the empirical rule of effect size (ERES). The MCID method was described by Sloan et al (24) for assessing the clinical significance of health-related changes in quality of life. This estimates an MCID, which corresponds to the mean change in patients who reported a small change (24). The ERES method is a direct modification of the ES approach. It defines the clinical significance of change as one-half of the SD, according to Cohen's ES classification (20). The method is based upon the fact that 99% of any normal distribution falls within ± 3 SDs of the mean, which means the measurement range of any instrument can be represented by 6 SDs. One theoretical SD can be calculated by dividing the total range by 6 ($SD = \text{total range}/6$) (24).

Statistical analysis. The analyses were conducted under nonparametric conditions, since several of the outcome measurements were ordinal. However, for responsiveness indices, such as the ES and SRM, the means $\pm$ SDs were calculated for ABILHAND, since they are measured on a linear scale. Statistical analyses were performed with SPSS, version 18. The statistical significance level was 0.05.

RESULTS

Of the 132 patients initially enrolled in the study, 105 completed and returned their questionnaires for the first evaluation, 93 returned their completed surveys for the second evaluation, and 88 patients returned completed questionnaires for the third evaluation. There was an 84% participation rate measured from baseline to the 1-year followup. The baseline characteristics of the 105 participants are shown in Table 1.

Reproducibility. The patients' HAQ, ABILHAND, RAQoL, and VAS pain scores at time 1 and time 2 were

highly correlated with ICCs of 0.97, 0.86, 0.95, and 0.78, respectively, and no statistically significant differences were observed (Table 2). The absence of significant differences and the high positive correlations between the time 1 and time 2 scores indicates that the HAQ, ABILHAND, RAQoL, and VAS pain scales have good test-retest reliabilities.

Responsiveness. Global approach. Table 3 shows the responsiveness analyses based on the global approach where the patients' measurements at time 3 and time 1 were compared. There were only significant differences observed in DAS28-CRP ($z = -2.09$, $P = 0.04$) and ABILHAND ($z = -2.02$, $P = 0.04$). The mean change in ABILHAND was 0.39 logit. The ES and SRM were 0.15 and 0.23, respectively. This indicates that, overall, there was a slight improvement in manual ability after 1 year.

Group approach. According to the disease activity scores, 18 patients showed deterioration, 41 were stable, and 29 showed improvement. Statistically significant differences between the time 3 and time 1 measurements were observed for the DAS28-CRP, HAQ, ABILHAND, RAQoL, and VAS pain scales in the deteriorated and improved groups (Table 4). No significant differences were observed in the stable group for any of the scales except the ABILHAND, which detected a significant positive change ($P = 0.03$).

The mean change, ES, and SRM responsiveness indices were calculated for ABILHAND for the 3 groups. The mean $\pm$ SD change in manual ability was higher in the deteriorated group (-1.23 ± 1.53 logits) and in the improved group (1.22 ± 2.06 logits) than in the stable group (0.48 ± 1.09 logits). The ES and SRM were -0.42 and -0.80 , respectively, in the deteriorated group, 0.52 and 0.59 in the improved group, and 0.20 and 0.44 in the stable group.

Individual approach. Of the 88 patients, 8.5% presented significant deterioration (T score < -1.96), and 15.5% showed significant improvement (T score > 1.96). The other 76% of patients presented either moderate improvement ($0 < \text{T score} \leq 1.96$), moderate deterioration ($-1.96 \leq \text{T score} < 0$), or no change (T score = 0).

Concordance between individual approach and group approach. Figure 1 shows the proportion of patients in each of the 5 classes from the individual approach accord-

Table 2. Test-retest reliability indices of the HAQ, ABILHAND, RAQoL, and VAS pain*

	Median (IQR)		Wilcoxon's signed rank test		
	Time 1	Time 2	z	P	ICC
HAQ	0.94 (1.34)	1.13 (1.25)	-1.51	0.13	0.97
ABILHAND	2.71 (4.01)	2.79 (4.65)	-0.25	0.8	0.86
RAQoL	10.5 (13)	11 (11)	-1.04	0.3	0.95
VAS pain	37 (52)	34 (52)	-0.79	0.43	0.78

* HAQ = Health Assessment Questionnaire; RAQoL = rheumatoid arthritis quality of life; VAS pain = visual analog scale for pain; IQR = interquartile range; ICC = intraclass correlation coefficient.

Table 3. Responsiveness analysis using a global approach*

	Median (IQR)		Wilcoxon's signed rank test		Change		
	Time 1	Time 3	z	P	Mean $\pm$ SD	ES	SRM
DAS28-CRP	3.04 (2.12)	2.53 (1.82)	-2.09	0.04	—	—	—
HAQ	0.88 (1.13)	0.88 (1.38)	-0.1	0.92	—	—	—
ABILHAND	2.80 (3.95)	3.09 (4.87)	-2.02	0.04	0.39 $\pm$ 1.67	0.15	0.23
RAQoL	10 (13)	10 (13)	-1.38	0.17	—	—	—
VAS pain	33.5 (50)	28 (54)	-0.31	0.75	—	—	—

* IQR = interquartile range; ES = effect size; SRM = standardized response mean; DAS28-CRP = Disease Activity Score in 28 joints using C-reactive protein level; HAQ = Health Assessment Questionnaire; RAQoL = rheumatoid arthritis quality of life; VAS pain = visual analog scale for pain.

ing to the variations in disease activity measured using an in-group approach. The proportion of patients with significant deterioration on the ABILHAND was higher in the deteriorated group than in the stable group. The proportion of patients with no change on the ABILHAND and the proportion of patients with moderate improvement or moderate deterioration were higher in the stable group than in the other groups. However, patients with significant improvements on the ABILHAND were either in the improved group or in the stable group.

Clinical significance of changes. The MCID was assessed in each of the 3 groups of patients. The mean $\pm$ SD change was -1.23 ± 1.53 logits in the deteriorated group, 1.22 ± 2.06 logits in the improved group, and 0.48 ± 1.09 logits in the stable group. These mean changes confirmed the ERES analysis. In fact, the initial range of ABILHAND for RA is 14.2 logits, and 1 theoretical SD corresponds to 2.36 logits (14.2 logits/6) (4). Therefore, the ERES defines that a clinically significant change is equal to 0.47 logit (0.2 times the theoretical SD) for a small change, 1.18 logits (0.5 times the SD) for a moderate change, and 1.89 logits (0.8 times the SD) for a large change. The comparison between the theoretical expected and observed changes indicated that the observed change was small in the stable group and moderate in the deteriorated and improved groups.

DISCUSSION

Our results demonstrate that the ABILHAND questionnaire shows responsiveness in measuring small changes in

RA patients over a 1-year period. Eighty-eight of 105 patients completed this study, which means the participation rate was 84% from baseline to followup. This is similar to the response rate of 83.6% reported by Rohekar and Pope in a previous test-retest reliability study that used the HAQ, VAS pain, and other questionnaires to evaluate 122 RA patients (26).

Checking the reproducibility of a scale is an important step before addressing questions about its responsiveness. Durez et al assessed the test-retest reliability of the ABILHAND questionnaire by comparing patients receiving stable treatments for RA and observed that the ABILHAND questionnaire was reliable over an average interval of 5.4 months (4). Our results indicate that the self-reported measurements were stable over time when no change in patient condition occurred. These results confirmed the test-retest reliability of previously used instruments (4,26,27).

The functional status of our RA cohort changed slightly between the baseline and 1-year followup. There were moderately significant ($P < 0.04$) statistical differences in the disease activity score and manual ability as measured by the DAS28-CRP and ABILHAND questionnaire. Meanwhile, the HAQ, RAQoL, and VAS pain showed no statistically significant differences, indicating, a priori, that there was no global change in functional disability, quality of life, and RA-related pain in the sample. According to Dworkin et al, trials with negligible mean benefits may be sufficiently powered for the results to be statistically significant (28). When nonsignificant results are observed using a relatively small sample size, some authors suggest checking the power of the test (29). If the power of the test is less than 0.80, nonsignificant results may be due to the

Table 4. Responsiveness analysis using a group approach*

	Deterioration (n = 18)				Stable (n = 41)				Improvement (n = 29)			
	Time 1, median (IQR)	Time 3, median (IQR)	z	P	Time 1, median (IQR)	Time 3, median (IQR)	z	P	Time 1, median (IQR)	Time 3, median (IQR)	z	P
DAS28-CRP	2.87 (2.42)	4.23 (2.44)	-3.16	0.002	2.31 (1.55)	2.45 (1.56)	-1.07	0.29	3.89 (2.05)	2.42 (1.19)	-4.70	0.000
HAQ	0.88 (1.13)	1.75 (1.13)	-2.66	0.008	0.63 (1.25)	0.63 (1.25)	-0.58	0.56	1.38 (1.22)	1 (1.19)	-2.34	0.02
ABILHAND	1.96 (4.74)	0.02 (2.99)	-2.19	0.03	3.02 (3.11)	3.65 (5.62)	-2.23	0.03	1.55 (4.45)	3.11 (5.62)	-2.46	0.01
RAQoL	13.5 (13)	16 (13)	-2.24	0.03	9 (14)	10 (14)	-0.85	0.39	12 (12)	10 (9)	-2.17	0.03
VAS pain	31 (58)	75 (54)	-3.21	0.001	20 (43)	20 (44)	-0.99	0.32	46.5 (41)	18 (37)	-2.73	0.006

* See Table 3 for abbreviations.

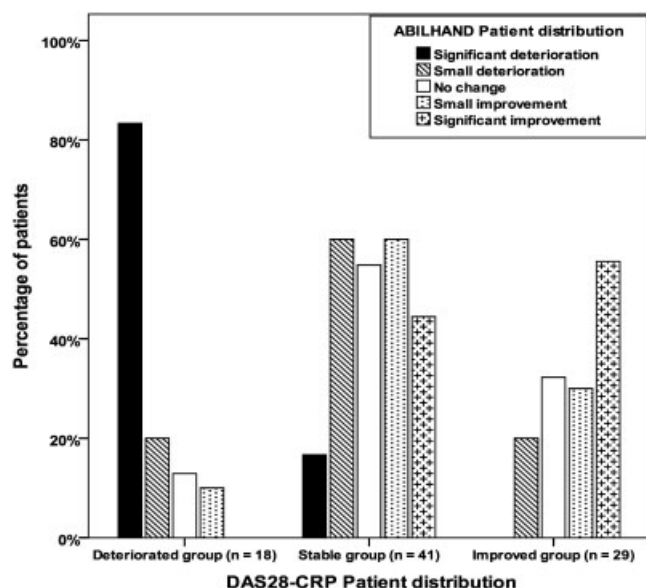


Figure 1. Concordance between patient distributions according to DAS28-CRP scores and ABILHAND measurements. The relationship between the disease activity as assessed by the DAS28-CRP clinician-based criteria and self-rating of manual ability using ABILHAND showed a good concordance. Patients with significant deterioration in manual ability were primarily in the deteriorated group. The majority of patients with small or no change were mainly in the stable group. Some patients with significant improvement were placed in the stable group; however, more than half of these patients were placed in the improved group. DAS28-CRP = Disease Activity Score in 28 joints using C-reactive protein level.

insufficient power of the test, rather than the lack of a real difference in the data. In this study, even though the sample size was quite large, the power of the test ranged from 0.26–0.52, which was less than the recommended value of 0.80. Therefore, these results had to be carefully interpreted. Another possible explanation of why the HAQ, RAQoL, and VAS pain did not show significant changes may be related to the fact that some patients deteriorated while others improved, which led to nonsignificant differences.

The responsiveness of the ABILHAND questionnaire was tested using group and individual approaches that applied anchor-based and distribution-based methods. The distribution-based method allowed for the calculation of statistical indices, such as ES, SRM, and SEs of the measurements. Calculation of these indices involves the SD, which is specific to a particular sample (30). Consequently, these indices are sensitive to the distribution of the measurements. The anchor-based method allows external criteria to be used as the standards against which the statistical or clinical importance of changes can be evaluated (28). These external criteria can be patient based, clinician based, or laboratory based. According to Dworkin et al, patient-based criteria allow for a retrospective comparison of change with an earlier time point (28). In chronic disease cases, remembering the previous functional status level can be difficult, and the patient's present condition can influence his or her perception of changes. Consequently, patient-based criteria can be more subjective

than clinician- or laboratory-based criteria when perceiving chronic diseases changes. Both distribution-based and anchor-based methods are complementary for interpreting responsiveness meaningfully. In our study, we used DAS28-CRP for the clinician-based criteria to classify each patient as “improved,” “stable,” or “deteriorated,” along with applying ES and SRM.

The scales used in this study were effective in detecting changes. They detected significant changes in the deteriorated and improved groups. The ABILHAND questionnaire noted that there was a significant improvement in manual ability in the group classified as stable using the clinician-based criteria. Considering the International Classification of Functioning, Disability, and Health model, instruments used in this study measure different variables. The DAS28-CRP and VAS pain measure impairment, while the HAQ and the ABILHAND questionnaire measure activity limitations and RAQoL evaluates quality of life. Both instruments measuring impairment showed no significant change in the stable group. The patients' functional abilities did not change as evaluated by the HAQ through its 8 categories (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and common daily activities), but the ABILHAND questionnaire pointed out that manual ability improved in the group of patients classified as stable according to their disease activity score. In chronic diseases such as stroke and RA, the actual disability is not linearly related to the impairments. The degree of disability could depend on the complex interaction between the task to be done, the patient's motivational status, and his or her compensatory behaviors, so that one could notice stability in the impairment domain with an improvement in activity limitations due to the possible adaptations that patients develop when experiencing items. Moreover, the HAQ covers several areas of functional ability, while the ABILHAND questionnaire focuses on manual ability.

Nevertheless, achieving a standard statistical significance depends on the variation of the measures and the sample size. Moreover, we cannot infer that each individual in the stable group uniformly experienced the group mean change. Therefore, the individual approach was essential to explain the significant change observed in manual ability in the stable group. The individual approach showed that some patients presented a small change, while a few patients presented significant changes in the stable group. This observation was corroborated by the ERES analysis of clinically significant change.

Our results confirmed the conclusions of a preliminary assessment of the responsiveness of the ABILHAND questionnaire. In 2007, Durez et al evaluated score changes in 6 patients before and after treatment with tumor necrosis factor blockers and observed a significant improvement in manual ability, as measured by the ABILHAND questionnaire, suggesting that the scale is responsive (4).

In the present study, there was a good concordance between the patient's individual changes and the group's changes. More than 80% of the patients whose disease was identified as significantly deteriorated on the ABILHAND questionnaire in the individual analysis were effectively detected as deteriorated according to the external DAS28-

CRP criteria in the group approach. On the other hand, most of the patients with no change or small changes in manual ability according to the individual approach were placed in the stable group. Approximately 60% of patients with individual significant improvements in manual ability were placed in the improved group. This indicates that there is a good concordance between the individual and group approaches.

This study confirmed the utility of the DAS28-CRP as a clinical criterion to gauge relevant clinical changes in RA patients, but it does not measure manual ability. It is then important to take into account the patient's perception of functional changes. Patient-reported outcomes are commonly used in clinical trials for chronic diseases (7), since the judgments of both clinicians and patients are necessary for good clinical practice. The concordance in the classifications of patients using the DAS28-CRP and the ABILHAND questionnaire indicates the usefulness of ABILHAND for evaluating functional changes in RA patients over time, especially in terms of manual ability. Our results suggest that a trial should include the DAS28-CRP for impairment, the HAQ for activity limitations, and the RAQoL for quality of life. Moreover, the ABILHAND questionnaire should also be selected to measure manual ability.

Additionally, ABILHAND scores can be analyzed using an individual approach, which integrates patient measures and associated SEs of measurement. This offers the ability to interpret results for individual patients, which is essential in clinical settings. Individual analyses are possible because, unlike other questionnaires, the ABILHAND questionnaire is measured on a linear scale and patient measurements are obtained using a Rasch analysis, which provides the associated SEs of the measurements.

As in the article by Hobart et al, this complementary investigation of the responsiveness of the ABILHAND questionnaire based on individual approach demonstrated the advantage of using Rasch analysis, which enables analysis at the individual level (by providing individual SEs for person estimates), something that is not available from classic test theory (31).

In conclusion, the ABILHAND, a simple self-report questionnaire that focuses on bimanual activities, thereby making it different from many of the functional scales currently available, can be used in clinical trials and settings to evaluate the effects of clinical interventions on manual ability in RA patients. Our results demonstrate that the ABILHAND questionnaire was responsive in detecting slight changes in RA patients. The responsiveness of this tool in individuals demonstrates that the questionnaire could be effectively used to assess the course of RA in clinical settings. The results of our study support the use of the ABILHAND questionnaire for monitoring the long-term functional status of RA patients.

ACKNOWLEDGMENTS

The authors thank the patients who participated in this study. They are also grateful to Geneviève Depresseux, Zineb Berbit, and Damienne De Limbourg for helping with administrative issues.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Thonnard had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Batcho, Durez, Thonnard.

Acquisition of data. Batcho, Durez, Thonnard.

Analysis and interpretation of data. Batcho, Durez, Thonnard.

REFERENCES

1. Wiles N, Symmons DP, Harrison B, Barrett E, Barrett JH, Scott DG, et al. Estimating the incidence of rheumatoid arthritis: trying to hit a moving target? *Arthritis Rheum* 1999;42:1339–46.
2. Walker-Bone K, Farrow S. Rheumatoid arthritis. *Clin Evid (Online)* 2007;1:1124.
3. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. *Stroke* 2001;32:1627–34.
4. Durez P, Frassel V, Houssiau F, Thonnard JL, Nielsens H, Penta M, et al. Validation of the ABILHAND questionnaire as a measure of manual ability in patients with rheumatoid arthritis. *Ann Rheum Dis* 2007;66:1098–105.
5. Cox SR, McWilliams L, Massy-Westropp N, Meads DM, McKenna SP, Proudman S. Adaptation of the RAQoL for use in Australia. *Rheumatol Int* 2007;27:661–6.
6. Grotle M, Brox JJ, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine* 2004;29:E492–501.
7. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
8. Cudkowicz ME, Schoenfeld D, Williams L. Improving the responsiveness of rating scales: the challenge of stepping twice into the same river [editorial]. *Neurology* 2004;62:1666–7.
9. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;50:239–46.
10. De Bruin AF, Diederiks JP, de Witte LP, Stevens FC, Philipsen H. Assessing the responsiveness of a functional status measure: the Sickness Impact Profile versus the SIP68. *J Clin Epidemiol* 1997;50:529–40.
11. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001;54:1204–17.
12. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
13. Van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism criteria. *Arthritis Rheum* 1996;39:34–40.
14. Shrout P, Fleiss J. Intraclass correlations: use in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
15. Portney LG, Watkins MP. Foundations of clinical research. Stamford (CT): Appleton & Lange; 1993. p. 42.
16. Matza LS, Johnston JA, Faries DE, Malley KG, Brod M. Responsiveness of the adult attention-deficit/hyperactivity disorder quality of life scale (AAQoL). *Qual Life Res* 2007;16:1511–20.
17. Abetz L, Arbuckle R, Allen RP, Mavraki E, Kirsch J. The reliability, validity and responsiveness of the Restless Legs

- Syndrome Quality of Life questionnaire (RLSQoL) in a trial population. *Health Qual Life Outcomes* 2005;3:79.
18. Middle B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002;2:e15.
 19. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985;28:542–7.
 20. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale: Laurence Erlbaum; 1988.
 21. Redelmeier DA, Tversky A. Discrepancy between medical decisions for individual patients and for groups. *N Engl J Med* 1990;322:1162–4.
 22. Wright BD, Stone MH. *Best test design*. Chicago: Mesa Press; 1979.
 23. Sloan JA, Cella D, Frost M, Guyatt GH, Sprangers M, Symonds T, and the Clinical Significance Consensus Meeting Group. Assessing clinical significance in measuring oncology patient quality of life: introduction to the symposium, content overview and definition of terms. *Mayo Clin Proc* 2002;77:367–70.
 24. Sloan JA, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health-related quality of life change within clinical trials. *Drug Inform J* 2003;37:23–31.
 25. Fulk GD, Echternach JL. Test-retest reliability and minimal detectable change of gait speed in individuals undergoing rehabilitation after stroke. *J Neurol Phys Ther* 2008;32:8–13.
 26. Rohekar G, Pope J. Test-retest reliability of patient global assessment and physician global assessment in rheumatoid arthritis. *J Rheumatol* 2009;36:2178–82.
 27. De Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. *Br J Rheumatol* 1997;36:878–83.
 28. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105–21.
 29. Pallant J. *SPSS survival manual: a step by step guide to data analysis using SPSS for windows*. 3rd edition. New York: McGraw Hill; 2007.
 30. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27:S178–89.
 31. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *J Neurol Neurosurg Psychiatry* 2010;81:1044–8.