Research Report

Keitel Functional Test for Patients With Rheumatoid Arthritis: Translation, Reliability, Validity, and Responsiveness

Bente Holm, Soren Jacobsen, Henrik Skjodt, Mette Klarlund, Trine Jensen, Merete Lund Hetland, Mikkel Ostergaard

B Holm, PT, MSc, is Research Physiotherapist, Department of Physiotherapy, Copenhagen University Hospital at Hvidovre, Copenhagen, Denmark. Address all correspondence to Ms Holm at: bente.holm@hvh.regionh.dk.

S Jacobsen, MD, DMSc, is Clinical Director, Department of Rheumatology, Copenhagen University Hospital at Rigshospitalet, Copenhagen, Denmark.

H Skjodt, MD, PhD, is Consultant in Rheumatology, Department of Rheumatology, Copenhagen University Hospital at Hvidovre.

M Klarlund, MD, PhD, is Staff Physician, Department of Rheumatology, Copenhagen University Hospital at Hvidovre.

T Jensen, MD, PhD, is Staff Physician, Department of Rheumatology, Copenhagen University Hospital at Hvidovre.

ML Hetland, MD, PhD, is Consultant in Rheumatology and Associate Professor, Department of Rheumatology, Copenhagen University Hospital at Hvidovre.

M Ostergaard, MD, DMSc, PhD, is Professor, Department of Rheumatology, Copenhagen University Hospital at Hvidovre and Copenhagen University Hospital at Herlev, Copenhagen, Denmark.

[Holm B, Jacobsen S, Skjodt H, et al. Keitel Functional Test for patients with rheumatoid arthritis: translation, reliability, validity, and responsiveness. *Phys Ther.* 2008; 88:664–678.]

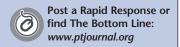
© 2008 American Physical Therapy Association

Background and Purpose. The purpose of this study was to translate the German Keitel Functional Test (KFT) into Danish and test it for reliability, concurrent and predictive validity, and responsiveness in patients with rheumatoid arthritis (RA).

Methods. Translation of the KFT was performed according to international recommendations, and the translated version was tested twice by 2 observers for intraobserver and interobserver reliability, with a 1-week interval between assessments, in 20 patients with RA with stable disease activity. Validity was investigated by studying 2 patient groups: (1) 15 patients with long-lasting (median=6 years) active RA, tested before and after 2, 6, and 14 weeks of anti-tumor necrosis factor alpha (TNF- α) inhibitor therapy, and (2) 35 patients with early (median=0.25 year) RA, tested at years 0, 0.5, 1, and 2. Twenty-three patients in the early RA group also were tested at year 7. KFT, conventional clinical and biochemical markers of disease activity, and Health Assessment Questionnaire (HAQ) were used.

Results. The translated KFT showed good intraobserver reliability (intraclass correlation coefficients [ICC]=.90 and .95, coefficient of variation [CV]=3.5%) and interobserver reliability (ICC=.99 and .92, CV=3.5%), and the KFT correlated with several measures of disease activity and, most closely, with the HAQ. The KFT was, in contrast to clinical disease activity measures, not sensitive to changes over time. Only baseline KFT data were significantly related to functional changes over a long period of time as measured by the KFT, and only in the early RA group.

Discussion and Conclusion. The Danish translation of the KFT showed good reliability, acceptable concurrent validity, very poor responsiveness, and inconclusive results concerning predictive validity. The results of this study do not support the use of the KFT for monitoring function in clinical practice, as an outcome measure in clinical trials, or as a predictor of functional changes.



heumatoid arthritis (RA) is a common inflammatory joint disease characterized by pain, swelling, and reduced mobility of the joints, followed by various degrees of functional impairment. In all stages of the disease, physical therapy intervention and medication are considered to be important cornerstones of treatment. Despite treatment, limitations in physical functions and restrictions in daily activities and social participation, including paid work, are often seen, and approximately half of the patients leave the labor force within 6 to 10 years of disease onset.1,2 To optimize interventions aimed at maintaining physical function and minimizing disability, a proper understanding of patients' functioning and health status is needed.3

The International Classification of Functioning, Disability and Health (ICF),^{4,5} a model developed by the World Health Organization (WHO), provides a useful framework for this purpose. This model refers to dimensions in a person's life at different levels (ie, participation, activity, body functions, or body structure),4,5 and it can be used to assess effects of different treatments at different levels. At the level of participation in society, health status and quality-of-life measures have been developed. At the level of activity, the Health Assessment Questionnaire (HAQ), which is a well-known, validated, self-reporting system for evaluating activities of daily living,6-8 often is used. The HAQ has been reported to have good predictive value for physical disability1,9-11 and is sensitive to changes over time,9 but it is not an appropriate instrument for assessing changes in physical impairment after short-term exercise therapy.12 At the level of body structure, several measures (ie, measurements of joint swelling, pain, and radiological changes and biochemical markers) are in use.13

In contrast, there are very few reliable and validated RA-specific measures at the level of body function, which might be useful to show the effects of physical therapy intervention on daily functioning. A German functional performance test, the Keitel Functional Test (KFT), has been developed for use in patients with RA. The KFT is based on range of motion and muscle activity, assessing 24 simple movement patterns for both upper and lower extremities. The 24 items are graded with a scoring system in which an index value of 100 points corresponds to normal functional ability. The test can be performed in 15 to 20 minutes and does not require any special instruments (Appendix).14 This RA-specific measure of impairment of body functions has for several years been used by physical therapists in Denmark, for both inpatients in hospitals and outpatients in rehabilitation clinics, without prior validation. The KFT has been described as an outcome measure15 and as a gold standard for evaluation of a new index of hand function.16 It has been shown to have good concurrent validity,¹⁷ especially when used with the HAQ,18-20 and it has been reported to be a strong predictor of mortality.²¹ Previous studies^{22,23} have shown the KFT to be sensitive for detecting changes after 0.5 and 1.5 years of treatment with disease-modifying antirheumatic drugs (DMARDs), although its sensitivity for detecting changes over shorter (weeks to months) and longer (years) periods of time has not been systematically investigated. Changes during treatment with novel RA therapies, such as use of tumor necrosis factor alpha (TNF- α) inhibitors, have not been studied.

Because the KFT provides an overall picture of functional limitation and was developed to detect functional changes over time, ¹⁴ it may be a useful measure of impairment of body functions in both clinical practice and research. However, the KFT

needs to be sufficiently validated. The purpose of this study was to validate a translated version of the KFT for reliability, concurrent and predictive validity, and responsiveness.

Method

This study of KFT involved 3 main stages: (1) translation, including field testing, (2) assessment of reliability, and (3) assessment of validity and responsiveness.

Translation, Including Field Testing

The test was translated from the original German language into Danish following international recommendations,²⁴ including a "bilingual panel," a "professional panel," a practical field test, and finally a backtranslation procedure. This was done in 1996.

Bilingual panel. The primary translation was performed separately by 3 Danish physical therapists with German-language skills (females, aged 27–48 years) recruited in Copenhagen. All panel members were bilingual and had been working in Germany. After the initial translation, the 3 physical therapists met and discussed the items until they had reached consensus. In cases where the panel members could not reach consensus by discussion, they sent the items to the professional panel for further discussion.

Professional panel. Three physical therapists (females, aged 31-45 years) who were not specialized in the German language and who had no knowledge of the KFT, but who had experience treating patients with RA, were recruited from the Department of Physiotherapy, Copenhagen University Hospital at Hvidovre. They read the text to check that the tasks in the KFT were clearly described and intelligible. The panel members then adjusted the language, if necessary.

Field testing. A pilot study with 4 patients with RA and 4 physical therapists was carried out to assess the new translation of the KFT. Four patients with RA at Copenhagen University Hospital at Hvidovre were included (1 male, aged 55 years, and 3 females, aged 35-78 years). The patients were in functional class II to III25 and showed a variation in disease activity. Four physical therapists from the Department of Physiotherapy, Copenhagen University Hospital at Hvidovre, (females, aged 26-55 years), who were not specialized in treating patients with RA and who had no knowledge of the KFT tested the 4 patients. They were introduced to the KFT by reading and performing all test items once. If doubts in interpretations occurred during testing, the test was returned to the professional panel to adjust the language to produce the final version.

Back translation. The final version was translated back to German by a Danish rheumatologist whose first language was German. The original version and the back-translated version were given to the bilingual panel, who compared the 2 texts to examine whether the meaning was identical for all items. If this was confirmed, the translated KFT was finally approved.

Assessment of Reliability

Procedure. To determine the intraobserver and interobserver reliability of the translated KFT, 20 patients with RA were tested 4 times (by 2 observers at 2 time points, with a 1-week interval between tests, randomized into 4 different sequences). This was carried out in 1997.

Patients. By review of patient files, 40 patients with RA and with unchanged medical therapy during the last 3 months were identified at the Department of Rheumatology, Copenhagen University Hospital at Hvi-

dovre. A random sample of 20 patients (17 women and 3 men, median age=64 years [range=26-78], median disease duration=6 years [range=2-48]) was included and evaluated as described above. Patients were excluded if they had changed therapy or reported a change in disease status during the 1-week interval between tests.

Observers. Two physical therapists from the Department of Physiotherapy, Copenhagen University Hospital at Hvidovre, were the observers. One physical therapist was very experienced in treating patients with RA and was familiar with the KFT. The other physical therapist had never seen or tried testing with the KFT. She first was introduced to the KFT and then administered the test, under supervision, to one patient. The 2 observers were masked to the functional level and previous test results of the patients.

Statistical analysis. The intraobserver and interobserver intraclass correlation coefficients (ICCs) were calculated. In addition, the variance between the 2 tests performed by the same observer and the 2 tests performed by 2 different observers was calculated using the coefficient of variance (CV).

For later evaluation of the ability of the KFT to show changes over time, the smallest detectable difference (SDD) was calculated. The SDD is derived from the limits of agreement method, representing the smallest change in score that can be discriminated from the measurement error of the scoring method. Use of the SDD as the threshold level for a certain increase or decrease in scores of functional changes ensures that changes are not due to random variability or measurement error. The SDD is based on the 95% limits of agreement, as described by Bland and Altman.26

Assessment of Validity and Responsiveness

Criterion validity is the agreement with concurrent and future standards, defined as the degree to which a measure truly reflects a gold standard.²⁷ There are 2 types of criterion validity: concurrent validity and predictive validity. Concurrent validity is the degree to which a measure reflects a gold standard applied at the same time (eg, pathologic evidence of joint inflammation and destruction), and predictive validity is the degree to which a measure predicts a future gold standard outcome (eg. functional impairment).27 Furthermore, we looked at responsiveness or sensitivity to change, which means the ability of a measure to detect clinically important degrees of change. Both variation in measurements over time (eg, treatment induced) and sufficient reproducibility to allow a reliable detection of this change are required.^{27,28}

Procedure. To investigate concurrent and predictive validity and responsiveness, 2 groups of patients with RA from the outpatient clinic at the Department of Rheumatology, Copenhagen University Hospital at Hvidovre, were recruited and followed.

Patients. One group (the anti-TNF- α group) consisted of 15 patients (14 women and 1 man, median age=45 years, range=23-62) with long-lasting RA (median disease duration=6 years, range=0-36). They were examined before treatment, at week 0, and during treatment with the TNF- α antagonist infliximab after 2, 6, and 14 weeks. This part of the study was carried out in 2005. Another group (the TIRA group) consisted of 35 patients (28 women and 7 men, median age=55 years, range=20-82) with early, relatively mild RA (median disease duration=0.25 years, range 0-2). They were included in the Danish TIRA Group study²⁹ and treated according to a protocol aimed at maximal inflammatory suppression with nonsteroidal anti-inflammatory drugs, DMARDs, corticosteroids, and, when available and necessary after 2 years, with biological treatment. The patients were examined before and after 0.5, 1, 2, and 7 years of therapy. The part of this study was carried out from 1998 to 2005.

Tests. The KFT was performed at all test sessions using the newly developed Danish KFT manual as described above. Furthermore, conventional clinical and biochemical measurements of disease activity were obtained, as recommended by the European League Against Rheumatism.13 These measurements consisted of the HAQ score,8 which is related to the patient's activity level according the ICF framework, and 6 parameters of disease activity: number of swollen joints, number of tender joints, patient's pain on a visual analog scale (VAS), patient's and physician's global assessments of disease activity on a VAS, and an acute phase reactant, the serum C-reactive protein (CRP). These 6 parameters of disease activity are related to the body structure level according to the ICF framework. In addition, the 28-item Disease Activity Score (DAS-28)—a composite measure combining number of tender joints, number of swollen joints, patient's global assessment of disease activity, and CRP—was calculated,13 and, the TIRA group patients, the examination program was supplemented with conventional radiography of the hands and wrists, scored using the method described by Larsen et al.30 These tests were selected because they are related to the body structure and activity levels according to the ICF framework and are most closely related to the body function level, which is measured by the KFT. Because no gold standard at this level exists, the disease activity

measures and particularly the HAQ score were the best available comparators for the KFT.

Test procedure. Except for radiographs, all tests of the individual patients were performed on the same day. The physical therapists were masked to previous KFT results and other clinical data.

Observers. Only experienced physical therapists administered the KFT. In the anti-TNF- α group, 1 physical therapist did the testing of 10 patients, and 2 physical therapists tested 5 patients each. In the TIRA group, 4 different physical therapists administered the tests during the first 2 years. The 7-year follow-up tests were administered by the same physical therapist.

Statistical analysis. Because all dependent variables were not normally distributed (Kolmogorov-Smirnov test), nonparametric tests were applied. Medians and ranges were used in the analysis. The KFT results were compared with the results of all conventional methods at all test times using the Spearman coefficient of correlation (ρ) to illustrate the concurrent validity. To assess responsiveness, the standardized response mean (SRM; small = < 0.5, medium = 0.5-0.8, $large = > 0.8)^{31}$ was used and changes from baseline were tested using the Wilcoxon signed rank test. Statistical significance was defined as $P \leq .05$. Assessment of the predictive value was applied using a forward stepwise regression analysis, with changes in functional ability, measured by the KFT and the HAQ, as the dependent variable. Baseline values for the KFT, the HAO, and the other clinical parameters were included in the regression analysis.

Ethics

The study protocol was approved by the local ethics committee, and the patients provided informed consent after receiving verbal and written information.

Results

Translation

The bilingual panel experienced only a few problems in translating the 24 functional tasks of the KFT. Only 1 task was sent to the professional panel because of alternative translations. The professional panel agreed on the text of this particular task and corrected a few minor linguistic errors.

The results of the field testing showed that 78 of the total of 86 possible answers in the KFT were used, and no problems in understanding the tasks occurred. However, the 4 physical therapists suggested that the differences between the "normal" speed and the "reduced" speed of a task (as seen in 4 tasks for the lower extremities and 2 tasks for transfer) should be explained specifically at the introduction of the test. The back translation and the original text were not identical, but no differences in task performance occurred when instructing by each of the 2 texts.

Reliability

Twenty patients with RA participated in the reliability testing. Twelve patients were classified as being in functional class II, 7 in class III, and 1 in class I.²⁵ In accordance with the inclusion criteria, no patient had changes in medication, number of painful joints (median=10, range=0-38), or morning stiffness (median=1, range=0.5-2 hours) within the 1-week interval between assessments. Results from the reliability cohort are presented in Table 1.

For intraobserver agreement, the ICCs for observer A and observer B were .95 and .90, respectively. The mean CV was 3.5% (3.4% for observer and 3.6% for observer B). For interobserver agreement, the ICCs

Table 1.Total Scores of the Keitel Functional Test (KFT) in the Reliability Cohort

	Test Time 1		Test Time 2	
	Observer A	Observer B	Observer A	Observer B
No. of patients	20	20	20	20
KFT total score, median (range)	75 (28–96)	73 (24–97)	76 (26–98)	78 (26–97)
KFT total score, mean (SD)	73 (14.0)	72 (15.3)	73 (15.5)	73 (15.8)

for test times 1 and 2 were .92 and .99, respectively. The mean CV was 3.5% (4.8% for test time 1 and 2.1% for test time 2). The intraobserver and interobserver SDDs were 9.7 and 9.3 points, respectively.

Validity and Responsiveness

In this part of the study, 2 groups of patients participated. In the anti-TNF- α group, 13 of the 15 patients completed the study. Two patients showed lack of efficacy and wanted to discontinue at the follow-up assessment. In the TIRA group, 23 of the 35 patients completed the 7-year follow-up assessment. Reasons for the 12 dropouts were that 5 patients had moved to another area of the country, 3 patients did not respond to the written request, 2 patients were living in residential homes for elderly people, and 2 patients had died. As assessed by the KFT and the HAO, the functional level of both groups of patients was reduced. Baseline values of the various parameters are represented in Table 2.

Concurrent validity. In the anti-TNF- α group, the highest correlation coefficients were generally found between the KFT and the HAQ and between the KFT and the CRP. The correlation coefficients were generally higher in the TIRA group than in the anti-TNF- α group, with the highest observed value between KFT and HAQ. At baseline, the KFT showed correlation coefficients of at least .50 for the HAQ, patient's pain, physician's global assessment of disease activity, CRP, and DAS-28 (Tab. 3).

Responsiveness. In both patient groups, all clinical parameters of disease activity and the HAQ showed, as expected, significant improvements from baseline. Most SRMs were large (>0.8), indicating good responsiveness (Tab. 2). In contrast, the KFT showed no significant changes from baseline at any time point except after 2 years in the TIRA group, and SRMs were low, except for medium SRMs (0.5-0.80) at 14 weeks in the anti-TNF- α group and at 2 years in the TIRA group, indicating poor responsiveness. Table 4 shows the proportion of patients who had an SDD of 10 or higher for change in KFT scores. In the anti-TNF- α group, the largest number of patients with an SDD of 10 or higher for change in KFT scores was 54% at week 14. In the TIRA group, the largest number of patients with an SDD of 10 or higher for change in KFT scores was 56% at year 7.

Predictive validity. In the anti-TNF- α group, the regression analysis showed no significant results (ie, no baseline parameters could predict KFT changes over time). In the TIRA group, the baseline KFT scores could explain 41% of the change in KFT scores after 2 years (P<.001) and 28% of the change in KFT scores after 7 years (P<.05). Baseline KFT scores could not predict changes measured by the HAQ in the 2 groups.

Discussion

Impairment of physical function is an important aspect in the evaluation of RA. There is a lack of observational tests of physical performance designed for use in patients with RA that have been tested for reliability, validity, and responsiveness. In the present study, the KFT, which focuses on detection of functional impairment in the trunk and the extremities, was translated into Danish, and reliability and various aspects of validity were tested. The KFT correlated with a disability measure (ie, the HAQ), with measures of disease activity, and with radiographic assessment of structural joint damage. Regression analysis showed the KFT to be only a predictor of future development of functional changes over long periods of time, but the KFT was not sufficiently sensitive to show changes over time. Thus, the study did not support the use of the KFT for monitoring function in clinical practice, as an outcome measure in clinical trials, or as a predictor of functional changes.

The translation of the KFT into Danish appeared to be successful. Both the bilingual panel and the professional panel experienced only a few problems during the translation procedure. Even though only 4 patients were assessed, the field test revealed good results, as almost all instructions were used without difficulty in performing the tasks. The translation method applied in this study is very extensive, in accordance with recommendations for health-related quality-of-life measures.²⁴

In the present study, test-retest reliability was assessed in 20 patients, with a 1-week interval between tests. This test-retest interval was used in an earlier study³² with good results and seems to be adequate for securing 2 identical groups of patients with RA to be studied. The cohort of 20 patients was smaller than co-

(Continued)

Table 2. Test Results in the Anti-Tumor Necrosis Factor Alpha (TNF- α) Inhibitor Therapy Group and the TIRA Group^{σ}

Neasure Anti-INI-α Croup	2 wk (n=15) 78 (55-97) .075 .0.45 .0.66** (0-2.0) .004	6 wk (n=15) 79 (56-97)	14 wk (n=13)	Baseline (n=35)	0.5 y (n=35)	1 y (n=35)	2 y (n=35)	7 y (n=23)
dian (range) 7 A A A Gian (range) 7 Gian (range) A f swollen rts (0–28) dian (range)	N 9	6 wk (n=15) 79 (56-97)	14 wk (n=13)	Baseline (n=35)	0.5 y (n=35)	1 y (n=35)	2 y (n=35)	7 y (n=23)
dian (range) 7 A A Oo-3.00) dian (range) A f swollen tts (0-28) dian (range)		79 (56–97)				()		
dian (range) 7 A Oo-3.00) dian (range) f swollen trs (0-28) dian (range)		79 (56–97)						
A dian (range) A f swollen tts (0–28) dian (range)		.187	78 (54–100)	82 (35–99)	81 (39–99)	88 (40–100)	85* (38–100)	83 (30–100)
Addian (range) Addian (range) A f swollen tts (0–28) dian (range)		:	.087		.161	.154	.013	.528
dian (range) A A f swollen tts (0–28) dian (range)		0.47	0.70		0.19	0.20	0.63	0.13
(ag) (ag)								
(əß	.004	0.55** (0-2.25)	0.63** (0-2.0)	0.94 (0-2.0)	0.19*** (0–2.25)	0.00*** (0-1.75)	0.00*** (0-1.25)	0.13* (0–2.5)
(alg	0.84	.004	.005		.001	.001	.001	.028
(ag		1.00	1.13		0.85	0.90	1.26	0.52
ian (range)								
Ь	3* (0–17)	3* (0–9)	1** (0–3)	6 (0–18)	0*** (0-11)	(9-0) ***0	0*** (0–21)	0*** (0-4)
	.047	.047	.003		.001	.001	.001	.001
SRM	0.48	0.32	0.90		26:0	0.93	1.07	0.93
No. of tender joints (0–28)								
Median (range) 9 (0–25)	5** (0–16)	5* (0–25)	3** (0-10)	18 (0–24)	2*** (0-24)	0*** (0-24)	0*** (0–24)	1 *** (0-7)
Р	800.	.046	.005		.001	.001	.001	.001
SRM	0.87	0.88	1.35		1.36	1.53	1.42	2.05
Patient's pain, VAS (0–100)								
Median (range) 49 (2–88)	34** (2–80)	25** (3–98)	27* (0–77)	40 (8–76)	16* (0–68)	14*** (0–65)	6*** (0–55)	14** (0–64)
р	.005	.005	.013		.022	.001	.001	.004
SRM	1.07	0.98	0.98		0.43	0.64	1.34	0.70
Patient's global assessment of disease activity, VAS (0–100)								
Median (range) 50 (2–96)	34** (2–90)	25** (5–98)	29** (0–75)	47 (3–85)	17** (0–65)	12*** (0-74)	14*** (0–58)	14* (0–85)
Р	.010	.008	.007		900.	.001	.001	.044
SRM	0.70	0.95	98.0		0.58	0.64	66.0	0.47

Table 2.Continued

Measure	Anti-TNF-α Group	dno			TIRA Group				
	Baseline (n=15)	2 wk (n=15)	6 wk (n=15)	14 wk (n=13)	Baseline (n=35)	0.5 y (n=35)	1 y (n=35)	2 y (n=35)	7 y (n=23)
Physician's global assessment of disease activity, VAS (0–100)									
Median (range)	45 (11–95)	26*** (0–75)	18** (0–54)	16** (0–33)	23 (3–79)	5** (0–52)	3*** (0-42)	0*** (0-35)	5*** (0–14)
Ь		.001	.005	.003		.002	.001	.001	.001
SRM		1.49	86:0	1.34		0.65	0.81	96:0	0.85
CRP, mg/L (<7−∞)									
Median (range)	36 (7–152)	21* (7–100)	13* (7-42)	23* (7–67)	7 (<7–128)	7 (<7-145)	7 (<7-96)	7 (<7–61)	7 (<7–48)
Ь		.017	.036	.028		959	.496	.199	.182
SRM		0.75	69:0	0.82		90.0	0.01	0.23	0.27
DAS-28 (0.0–9.0)									
Median (range)	4.7 (2.4–7.5)	3.8* (1.8–6.6)	3.5* (1.7–7.4)	3.2** (1.8–4.7)	5.3 (1.8–7.3)	3.2*** (1.7–6.9)	2.0*** (1.7–6.5)	2.4*** (1.7–6.4)	2.6*** (1.7–5.2)
Ь		.019	.046	.003		.001	.001	.001	.001
SRM		0.79	0.90	1.63		1.16	1.28	1.50	1.47
Radiological damage, Larsen score (0–140)									
Median (range)					0 (0–26)	0* (0–28)	0** (0-29)	0*** (0-30)	
Ь						.011	.003	.001	
SRM						-0.38	-0.44	-0.54	

 o KFT=Keitel Functional Test, SRM=standardized response mean, HAQ=Health Assessment Questionnaire, VAS=visual analog scale, CRP=C-reactive protein, DAS-28=28-item Disease Activity Score. o P values indicate statistically significant differences from baseline values, as assessed by the Wilcoxon signed rank test (*=P<.05, **=P<.01, ***=P<.001). Absolute values of SRM are regarded as small (<0.5), medium (0.5–0.8), or large (>0.8).

Table 3. Spearman Correlation Coefficients (rho) and *P* Values Between the Keitel Functional Test and Clinical, Biochemical, and Radiographic Parameters at Different Time Points in the Anti-Tumor Necrosis Factor Alpha (TNF- α) Inhibitor Therapy Group and the TIRA Group^{α}

Measure	Anti-TNF-α	Group			TIRA Grou	р			
	Baseline (n=15)	2 wk (n=15)	6 wk (n=15)	14 wk (n=13)	Baseline (n=35)	0.5 y (n=35)	1 y (n=35)	2 y (n=35)	7 y (n=23)
HAQ (0.00-3.00)									
rho	36	53	54	57*	50**	25	24	48**	67**
Р	.195	.054	.057	.041	.007	.186	.211	.004	.001
No. of swollen joints (0–28)									
rho	64*	39	14	42	28	38	19	16	14
Р	.011	.173	.649	.152	.143	.041	.311	.362	.524
No. of tender joints (0–28)									
rho	38	44	.10	30	35	35	19	16	29
Р	.163	.118	.738	.314	.063	.059	.305	.357	.181
Patient's pain, VAS (0–100)									
rho	.37	05	.13	41	62**	35	30	34	55**
Р	.181	.863	.666	.164	.001	.065	.115	.540	.008
Patient's global assessment of disease activity, VAS (0–100)									
rho	02	07	12	30	33	27	35	36	53*
P	.995	.805	.687	.381	.092	.164	.058	.043	.012
Physician's global assessment of disease activity, VAS (0–100)									
rho	42	34	.24	50	51**	29	09	13	51*
Р	.116	.239	.437	.099	.005	.124	.647	.472	.016
CRP, mg/L (<7–∞)									
rho	26	64*	54*	49	57**	12	29	23	40
Р	.368	.014	.046	.108	.001	.537	.124	.199	.064
DAS-28 (0.0-9.0)									
rho	41	49	02	32	53**	34	31	31	58**
Р	.146	.077	.950	.313	.004	.072	.100	.084	.005
Radiological damage, Larsen score (0–140)									
rho					28	33	32	19	
Р					.156	.076	.091	.286	

^a HAQ=Health Assessment Questionnaire, VAS=visual analog scale, CRP=C-reactive protein, DAS-28=28-item Disease Activity Score. *P* values indicate statistically significant differences from baseline values, as assessed by the Wilcoxon signed rank test (*=*P*<.05, ***P*<.01).

Table 4. Frequency of Change in Keitel Functional Test (KFT) Scores Higher Than the Smallest Detectable Difference of 10 Points in the Anti-Tumor Necrosis Factor Alpha (TNF- α) Inhibitor Therapy Group and the TIRA Group^{α}

				1
	Distribution Patients Tes	of Changes in ted)	KFT Scores (%	of All
	≥10 Points	≥+10 Points	≥-10 Points	<10 Points
Anti-TNF-α group				
Change from baseline to wk 2 (n=15)	43	29	14	57
Change from baseline to wk 6 (n=15)	36	29	7	64
Change from baseline to wk 14 (n=13)	54	39	15	46
TIRA group				
Change from baseline to 0.5 y (n=35)	32	24	8	68
Change from baseline to 1 y (n=35)	32	24	8	68
Change from baseline to 2 y (n=35)	43	43	0	57
Change from baseline to 7 y (n=23)	56	28	28	44

^a The smallest detectable difference was calculated from the reliability cohort.

horts in other studies^{8,10,32} but was favored by unchanged medication, number of painful joints, and duration of morning stiffness.

To ensure that the KFT version using the Danish manual was reproducible before using it in the validity study, a reliability study was completed beforehand. Consequently, 2 different cohorts were examined. Optimally, the validity study would have included a repetition of reliability testing.

In accordance with a previous study, ¹⁰ the present study had a very comprehensive approach, including both long-term and short-term assessments to ensure that a broad spectrum of disease severity and disease duration was included. In the TIRA group, 12 of the 35 patients dropped out between years 2 and 7. This dropout rate of 34% is higher than in an earlier study with a similar follow-up time. ¹ As the results of the 7-year follow-up assessment were

similar to the results of the 2-year follow-up assessment with all patients examined, the dropout rate is not expected to have had a major influence on the results.

In both the anti-TNF- α and the TIRA groups, the KFT was compared with internationally recommended parameters for assessment of disease activity, structural joint damage, and disability,13 but it would have been optimal if a gold standard for assessment of impairment of body function had been included. Neither the present study nor previous studies that tested the KFT^{17-20,23,33} included comparisons with other measures of functional impairment. To our knowledge, the only existing relevant functional test is the Signals of Functional Impairment (SOFI),10 which also was developed to evaluate impairment of body functions in patients with early symptoms of RA. The SOFI has been shown to correlate with the HAQ and to be sensitive to changes over time, and could have been of value for comparison with the KFT in this study. However, no Danish translation of the SOFI currently exists.

Analysis of concurrent validity in the present study showed acceptable degrees of correlation between the KFT and the HAQ, number of swollen joints, patient's pain, patient's and physician's global assessments of disease activity, CRP, and DAS-28. These results are similar to findings in a previous study¹⁷ that found the KFT to correlate with pain, patient's global assessment of disease activity, Ritchie Index, and CRP at baseline. The observed correlations possibly reflect the fact that many other conditions (eg, fatigue, stiffness, reduced range of motion and muscle strength) affect functional ability and the fact that the KFT, the HAQ, and the disease activity parameters reflect different aspects of the disease. It has been suggested that the correlation between the KFT and the HAQ could be explained by the fact that the KFT mainly reflects impairment at the body function level and that the HAO reflects disability at activity level.4,5,10,34 Despite the variation in correlations in the present study, the KFT showed a closer correlation with the HAQ than with the measures of disease activity, which is in agreement with findings of other studies. 18-20

This study was favored by correlating KFT and other parameters at different test times with several moderate correlation coefficients, contrary to previous studies, 17-20 which incorporated only baseline correlations. Analysis of responsiveness at the group level did not show significant improvements of the KFT values over time, in contrast to all other parameters. This finding indicates that the KFT cannot be used to show functional changes over short or long periods of time. These results are in accordance with findings from

an earlier study of 65 patients who showed no significant changes in KFT scores after 1 year.35 Two other studies22,23 showed good responsiveness for the KFT over periods of 0.5 to 1.5 years in patients initiating medication with DMARDs. Such therapy-induced changes were not found in this study. The cohort of patients treated with TNF- α inhibitors was smaller and was studied over a shorter period of time (ie, 3.5 months). This and different baseline functional levels may explain the different results in this study and in earlier studies.

However, the apparent unresponsiveness of the KFT could be due to lack of real change in functional ability over time in the patients in the TIRA group, as they had early, relatively mild RA and received effective therapy, which was aimed at controlling disease activity and, consequently, at reducing functional impairment. The poor responsiveness at the group level also could partly have been influenced by the fact that the functional level of the individual patient could change in both directions (improvement or deterioration), as illustrated in Table 4. At the patient level, changes greater than the SDD of 10 points occurred in 56% of the patients in the TIRA group after 7 years, suggesting that the individual patients did change. However, considered as a group, this change was smaller (between 1 and 6 points) (Tab. 2) and not detectable, because some patients improved and others deteriorated. Conclusively, the negative results of the responsiveness assessment can be explained by poor reliability, by changes in opposite directions at the level of the individual patient, and by the patient's functional stability.

The predictive value of the KFT was found only in the TIRA group for future functional changes after 2 and

7 years. The predictive value of the KFT has been analyzed in only one previous study,³³ which showed that baseline KFT scores, patient's pain, and patient's global assessment of disease activity were able to predict 25% of the functional changes in 85 patients after 1 year. Such results could not be found in our study.

Our study had some limitations. The study design should determine whether the KFT could be recommended as a test of physical performance. The KFT was compared with conventional clinical and biochemical measures of disease activity related to the body structure and activity level according to the ICF framework.4,5 However, the KFT relates to the body function level, and we could have added a measure such as the Swedish SOFI at this level as a gold standard for patients with newly diagnosed RA. The first step then should have been a translation of the Swedish SOFI into Danish.10 This would have been of value in interpreting both validity and responsiveness of the KFT, but was considered too laborious for the purpose of this work.

The numbers of patients in the anti-TNF- α group and in the 7-year follow-up assessment of the TIRA group were small. No formal sample size calculation was done. However, we considered the sample size sufficient to illustrate potential responsiveness of the KFT because available data indicated that it would be sufficient to illustrate changes in the conventionally used measures. This was confirmed.

Another methodological problem may be that we tested the reproducibility of the KFT separately and did not repeat this test in the validity part of the study. However, we considered the good results from the reliability part of the study to be largely transferable to the validity part of the

study because the 24 items of the KFT manual were described in detail and the same manual was used throughout the study.

Conclusion

A Danish translation of the KFT was successfully performed, and it showed good intraobserver and interobserver agreement, with acceptable concurrent validity, by comparison with measures of disease activity and of activities of daily living. In this study, the KFT showed very poor responsiveness and inconclusive results concerning predictive validity, indicating that the KFT is not suitable for assessing treatment efficacy, as it could not show changes over time. The present study does not support the use of the KFT for monitoring function in clinical practice, as outcome measure in clinical trials, or as a predictor of functional changes.

Ms Holm, Dr Jacobsen, and Dr Ostergaard provided concept/idea/research design. Ms Holm, Dr Jacobsen, Dr Hetland, and Dr Ostergaard provided writing. Ms Holm, Dr Skjodt, and Dr Jensen provided data collection. Ms Holm, Dr Jensen, and Dr Ostergaard provided data analysis. Ms Holm provided project management and clerical support. Ms Holm and Dr Hetland provided subjects. Ms Holm and Dr Ostergaard provided facilities/equipment. Dr Holm, Dr Jacobsen, Dr Klarlund, Dr Jensen, Dr Hetland, and Dr Ostergaard provided consultation (including review of manuscript before submission).

The data, in part, were presented at the REUMA Congress (Nordic Multidisciplinary Congress); September 14–17, 2005; Copenhagen, Denmark.

This article was submitted April 13, 2007, and was accepted January 4, 2008.

DOI: 10.2522/ptj.20070119

References

1 Lindqvist E, Saxne T, Geborek P, Eberhardt K. Ten-year outcome in a cohort of patients with early rheumatoid arthritis: health status, disease process, and damage. *Ann Rheum Dis.* 2002;61:1055–1059.

- 2 National Board of Health, Denmark. Rheumatoid Arthritis: Health Technology Assessment of Diagnosis and Treatment. Copenhagen, Denmark: Danish Center for Evaluation and Health Technology Assessment: 2002.
- 3 Kirchberger I, Glaessel A, Stucki G, Cieza A. Validation of the comprehensive International Classification of Functioning, Disability and Health core set for rheumatoid arthritis: the perspective of physical therapists. Phys Ther. 2007;87:368–384.
- 4 Stucki G, Cieza A. The *International Classification of Functioning, Disability and Health* (ICF) Core Sets for rheumatoid arthritis: a way to specify functioning. *Ann Rheum Dis.* 2004;63(suppl 2):ii40-ii45.
- 5 International Classification of Functioning, Disability and Health: (ICF). Geneva, Switzerland: World Health Organization; 2001.
- 6 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23:137-145.
- 7 Ramey DR, Raynauld JP, Fries JF. The Health Assessment Questionnaire, 1992: status and review. *Artbritis Care Res.* 1992;5:119-129.
- 8 Thorsen H, Hansen TM, McKenna SP, et al. Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the Rheumatoid Arthritis Quality of Life Scale (RAQoL). Scand J Rheumatol. 2001;30:103–109.
- 9 Drossaers-Bakker KW, Zwinderman AH, Vlieland TP, et al. Long-term outcome in rheumatoid arthritis: a simple algorithm of baseline parameters can predict radiographic damage, disability, and disease course at 12-year follow-up. Arthritis Rheum. 2002;47:383-390.
- 10 Eberhardt KB, Svensson B, Mortiz U. Functional assessment of early rheumatoid arthritis. *Br J Rheumatol*. 1988;27:364–371.
- 11 Eberhardt KB, Fex E. Functional impairment and disability in early rheumatoid arthritis: development over 5 years. *J Rheumatol.* 1995;22:1037–1042.
- 12 van den Ende CH, Breedveld FC, Dijkmans BA, Hazes JM. The limited value of the Health Assessment Questionnaire as an outcome measure in short-term exercise trials. *J Rheumatol*. 1997;24:1972–1977.
- 13 EULAR Standing Committee for International Clinical Studies Including Therapeutic Trials. EULAR Handbook of Clinical Assessment in Rheumatoid Artbritis. Leiden, the Netherlands: van Zuiden Communications BV; 2000.

- 14 Keitel W, Hoffmann H, Weber G, Krieger U. Evaluation of the percentage of functional decrease of the joints using a motor function test in rheumatology [in Dutch]. Disch Gesundheitsw. 1971;26:1901–1903.
- 15 Keysser M, Keysser C, Keitel W, Keysser G. Loss of functional capacity caused by a delayed onset of DMARD therapy in rheumatoid arthritis: long-term follow-up results of the Keitel Function Test—brief definite report. *Z Rheumatol*. 2001;60: 69-73.
- 16 Poole JL, Cordova KJ, Brower LM. Reliability and validity of a self-report of hand function in persons with rheumatoid arthritis. *J Hand Ther*. 2006;19:12-6, quiz.
- 17 Kalla AA, Kotze TJ, Meyers OL, Parkyn ND. Clinical assessment of disease activity in rheumatoid arthritis: evaluation of a functional test. *Ann Rheum Dis.* 1988;47: 773–779.
- 18 Bombardier C, Raboud J; for the Auranofin Cooperating Group. A comparison of health-related quality-of-life measures for rheumatoid arthritis research. Control Clin Trials. 1991;12:2438-2568.
- 19 Hakala M, Nieminen P, Manelius J. Joint impairment is strongly correlated with disability measured by self-report questionnaires: functional status assessment of individuals with rheumatoid arthritis in a population-based series. *J Rheumatol*. 1994;21:64-69.
- 20 Hakala M, Nieminen P. Functional status assessment of physical impairment in a community based population with rheumatoid arthritis: severely incapacitated patients are rare. *J Rheumatol.* 1996;23: 617–623.
- 21 Soderlin MK, Hakala M, Nieminen P. Anxiety and depression in a community-based rheumatoid arthritis population. *Scand J Rheumatol.* 2000;29:177–183.
- 22 Bombardier C, Ware J, Russell IJ, et al. Auranofin therapy and quality of life in patients with rheumatoid arthritis: results of a multicenter trial. Am J Med. 1986;81: 565-578.
- 23 Kalla AA, Smith PR, Brown GM, et al. Responsiveness of Keitel functional index compared with laboratory measures of disease activity in rheumatoid arthritis. Br J Rheumatol. 1995;34:141–149.
- 24 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epide*miol. 1993;46:1417–1432.

- 25 Steinbrocker O, Traeger CH, Battman RC. Therapeutic criteria in rheumatoid arthritis. *JAMA*. 1949;140:659-662.
- 26 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986:1:307-310.
- 27 Ostergaard M, Wiell C. Ultrasonography in rheumatoid arthritis: a very promising method still needing more validation. *Curr Opin Rheumatol.* 2004;16:223–230.
- 28 Felson DT, Anderson JJ. A review of evidence on the discriminant validity of outcome measures in rheumatoid arthritis. *J Rheumatol*. 2001;28:422–426.
- 29 Klarlund M, Ostergaard M, Jensen KE, et al; for the TIRA Group. Magnetic resonance imaging, radiography, and scintigraphy of the finger joints: one-year follow-up of patients with early arthritis. *Ann Rheum Dis.* 2000;59:521–528.
- 30 Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn (Stockh). 1977; 18:481–491.
- 31 Marra CA, Rashidi AA, Guh D, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Qual Life Res.* 2005;14:1333-1344.
- 32 Eberhardt K, Duckberg S, Larsson BM, et al. Measuring health-related quality of life in patients with rheumatoid arthritis: reliability, validity, and responsiveness of a Swedish version of RAQoL. *Scand J Rheumatol.* 2002;31:6–12.
- 33 Holm MB, Rogers JC, Kwoh CK. Predictors of functional disability in patients with rheumatoid arthritis. Arthritis Care Res. 1998;11:346-355.
- 34 Badley EM, Wagstaff S, Wood PH. Measures of functional ability (disability) in arthritis in relation to impairment of range of joint movement. *Ann Rheum Dis.* 1984; 43:563–569.
- 35 Ahlmen M, Bjelle A, Sullivan M. Prediction of team care effects in outpatients with rheumatoid arthritis. *J Rheumatol*. 1991; 18:1655–1661.

Appendix.

Keitel Functional Test^a

Keitel Functional Test

A test for functional mobility, evaluating functions of the hands, wrists, shoulders, trunk and lower limbs.

No.	Test Item	Grading Criteria	Res	ult
			Right	Left
1	Tip of the thumb touches the basic joint of the 5th finger	3= Test is performed fully and with no delay 2= Test is performed fully but with effort and/or delay 1= Tip of the thumb touches basic joint of 3rd or 4th finger 0= None of the items can be performed		
2	Bending of the 2nd finger (It is of no importance, if all 4 fingers are tested together or individually)	2= Finger can bend normally 1= Finger cannot bend normally, tip of the finger reaches palm 0= Finger does not reach palm		
3	Bending of the 3rd finger	2= Finger can bend normally		
		1 = Finger cannot bend normally, tip of the finger reaches palm 0 = Finger does not reach palm		
4	Bending of the 4th finger	2= Finger can bend normally 1= Finger cannot bend normally, tip of the finger reaches palm 0= Finger does not reach palm		
5	Bending of the 5th finger	2= Finger can bend normally 1= Finger cannot bend normally, tip of the finger reaches palm 0= Finger does not reach palm		
6	Forearms are placed horizontally. Press palms together with fingers pointing upwards.	 3 = Test is performed fully and with no delay 2 = Test is performed fully but with effort/or delay 1 = Dorsal flexion of the wrist to 45°, you may test one hand at a time 0 = Impossible (flexion less than 45°) 		
7	Forearms as above. Press back of the hands together with fingers pointing downwards.	 3 = Test is performed fully and with no delay. 2 = Test is performed fully but with effort and/or delay 1 = Dorsal flexion of the wrist to 45°, you may test one hand at a time 0 = Impossible (flexion less than 45°) 		
8	Place backs of both hands simultaneously on a table, with elbows bended 90°. The ulnar margin of the hands is lifted.	 2= Test is performed fully 1= Backs of the hands are lying on the table, lifting is not possible 0= Backs of the hands are not lying fully on the table 		
	(\sqrt{n})			

Appendix.

Continued

Keitel Functional Test

A test for functional mobility, evaluating functions of the hands, wrists, shoulders, trunk and lower limbs.

No.	Test Item	Grading Criteria	Res	ult
			Right	Left
9	Place radial margin of both hands simultaneously on a table. Thumbs pointing downwards in front of the table edge. Turn ulnar margin of the hands towards each other. Avoid bending of the trunk.	 2= Test is performed fully 1= Backs of the hands are standing perpendicular, but cannot turn towards each other 0= Backs of the hands are not standing perpendicular 		
10	Place both hands simultaneously on ipsilateral shoulder	2= Test is performed fully, delay permitted 1= Fingertips reach the shoulder or within a distance of 5 cm 0= Greater distance		
11	Place both hands (not only the fingertips) simultaneously behind the neck below ear level	 3= Test is performed fully and with no delay 2= Test is performed fully, but with effort and/or delay 1= Only the fingertips touch the neck. Neck is regarded as the area from earlobe to earlobe. 0= Fingertips cannot touch the neck 		
12	Rising from supine position (examination couch, turning to the side is not permitted)	6= With hands extended, performed quickly 5= With hands extended, performed with effort 4= With hand support 2= Only possible with support from a person 0= Impossible (Gradings 3 and 1 are not used)		
13	Active spreading of the legs in supine position	2= More than 50 cm condylar distance 1= More than 20 cm condylar distance 0= Less than 20 cm condylar distance		
14	Rising from a chair with no armrest	6= With hands extended, performed quickly 5= With hands extended, performed with effort 4= With hand support 2= Only possible with support from a person 0= Impossible (Gradings 3 and 1 are not used)		
	For test no. 15-21, support with one	hand is permitted!		
15	Stand on the tiptoes for 15 sec. The patient must stand up straight.	2= 15 sec is possible 1= Less than 15 sec is possible 0= Impossible		

(Continued)

Appendix.

Continued

Keitel Functional Test

A test for functional mobility, evaluating functions of the hands, wrists, shoulders, trunk and lower limbs.

No.	Test Item	Grading Criteria	Res	ult
			Right	Left
16	Stand on the heels for 15 sec. The patient must stand up straight and the forefoot must be lifted (how high is of no importance).	2= 15 sec is possible 1= Less than 15 sec is possible 0= Impossible		
17	Deep knee bending from standing position. Buttocks almost touch the heels (lifting of the heels is permitted.)	2= Test is performed normally 1= Only the beginning of the knee bending is performed 0= Impossible		
	A PERMITTED			
18	Standing on one leg for 15 sec	2= 15 sec is possible 1= Less than 15 sec is possible 0= Impossible		
19	External rotation of the hip from standing position. Place the heel of the test leg on the medial side of the foot of the standing leg. Avoid rotation of the pelvis. The angle between the feet must be higher than 90°.	2= Test is performed fully 1= Angle between the feet is 90° 0= Angle between the feet is less than 90°		
20	With bended knee place the sole of the foot on a chair. Patient must stand close to the chair.	2= Test is performed fully, delay permitted 1= The leg can be lifted from the floor 0= It is impossible to lift the leg from the floor		
21	With straight leg place the heel on a chair (a known extension defect of the knee is of no importance). Patient must stand approximately 1 meter from the chair.	2= Test is performed fully, delay permitted 1= The leg can be lifted from the floor 0= It is impossible to lift the leg		
	0			



(Continued)

Appendix.

Continued

Keitel Functional Test

A test for functional mobility, evaluating functions of the hands, wrists, shoulders, trunk and lower limbs.

No.	Test Item	Grading Criteria	Res Right	ult Left
22	Walk 30 meters in a hospital corridor. Standard time 20 sec. (It is permitted to cheer the patient and/or repeat the test.)	6= Standard time, no difficulty 5= Standard time, visible difficulty, walking aid permitted 4= 25 sec 3= 30 sec 2= 40 sec 1= Few steps, with or without personal support 0= Impossible		
23	Walk upstairs, 10 steps. Standard time 7 sec.	3= Standard time, no use of banister 2= Up to 14 sec. Use of banister permitted. 1= More than 14 sec. Only few steps are possible with effort. 0= Impossible		
24	Walk downstairs, 10 steps. Standard time 7 sec.	3= Standard time, no use of banister 2= Up to 14 sec. Use of banister permitted. 1= More than 14 sec. Only few steps are possible with effort. 0= Impossible		
		Total score	e	

^a This is an English translation performed by the primary author (BH) from the validated Danish version. This English version is not validated. The Danish version of the KFT is copyrighted by the Department of Physiotherapy, Copenhagen University Hospital at Hvidovre, Copenhagen, Denmark. The copyright allows the use of the Danish version free of charge after contacting the Department of Physiotherapy (attn: Bente Holm).