

## EXTENDED REPORT

## Reliability of the ICF Core Set for rheumatoid arthritis

Till Uhlig, Solvår Lillemo, Rikke Helene Moe, Tanja Stamm, Alarcos Cieza, Annelies Boonen, Petter Mowinckel, Tore Kristian Kvien, Gerold Stucki

*Ann Rheum Dis* 2007;**66**:1078–1084. doi: 10.1136/ard.2006.058693

See end of article for authors' affiliations

Correspondence to:  
Till Uhlig, MD, PhD, National Resource Center for Rehabilitation in Rheumatology, Department of Rheumatology, Diakonhjemmet Hospital, Postboks 23 Vinderen, N-0319 Oslo, Norway; [till.uhlig@nrrk.no](mailto:till.uhlig@nrrk.no)

Accepted 21 December 2006  
Published Online First  
11 January 2007

**Background:** The comprehensive ICF Core Set for rheumatoid arthritis (RA) is a selection of 96 categories from the International Classification of Functioning, Disability and Health (ICF), representing relevant aspects in the functioning of RA patients.

**Objectives:** To study the reliability of the ICF Core Set for RA in rheumatological practice, and to explore the metric of the qualifiers' scale.

**Methods:** 25 RA patients from an outpatient department of rheumatology were interviewed using the ICF Core Set for RA (76% females, mean (SD) age 57.5 (12.5) years, disease duration 15.9 (14.6) years). Interviews were performed independently by both a physiotherapist and an occupational therapist on the same day and again after one week by one of them. The severity of the patients' problems was quantified on a qualifier scale ranging from 0 (no problem) to 4 (complete problem). Analyses of intra-rater and inter-rater agreement, kappa statistics, and Rasch analyses were applied.

**Results:** Mean intra-rater (inter-rater) complete agreement for all categories was seen in 59% (47%) of observations, ranging from 29% (0%) to 96% (80%) for individual categories. Weighted kappa statistics with value  $\geq 0.4$  showed reliability in 86% of categories within raters, and in 43% of categories between raters. Improved inter-rater and intra-rater reliability was observed with a reduced number of qualifiers for the categories.

**Conclusions:** Inter-rater and intra-rater reliability of the ICF Core Set of RA was low to moderate. The metric of the qualifiers' scale may be improved by reducing the number of qualifiers to three for all components.

Rheumatoid arthritis (RA) is a chronic disabling disease that is often associated with limitation in physical, mental, and social function,<sup>1,2</sup> with potential work disability.<sup>3,4</sup> In order to describe and assess daily functioning and disability from a bio-psychosocial perspective in all aspects of health, the framework of the World Health Organization International Classification of Functioning, Disability and Health (ICF) can be used. This framework provides a unified and standard language for the description of health and health related conditions and a common framework among all health professions.<sup>5</sup> The ICF structure includes the two parts: (1) functioning; and (2) disability and contextual factors. Each part has two components: (i) Body Functions and Structures, Activities and Participation; (ii) Environmental Factors and personal factors. Components are further defined by the so called ICF categories. To rate the magnitude or the severity of the problem in each of the ICF categories, WHO proposes the so called qualifiers' scale. The categories are scored by a health professional during an interview with the patient.

In order to facilitate the application of the ICF in clinical practice, specific core sets were developed for specific diseases as short lists of ICF categories that are important for patients. They serve as minimal standards for the reporting of functioning and health for clinical and epidemiological studies as well as in clinical encounters (brief ICF core set) or as standards for multiprofessional, comprehensive assessment (comprehensive ICF Core Set) under consideration of influential environmental factors. The comprehensive ICF Core Set for RA represents the typical spectrum in functioning of patients with RA with a selection of 96 categories,<sup>6</sup> in the four components—Body Functions (b), Body Structures (s), Activities and Participation (d), and Environmental Factors (e). The preliminary version of the ICF Core Set for RA<sup>6</sup> (hereafter stated when referring to the comprehensive ICF Core Set for RA) was developed by experts

consisting of rheumatology health professionals in a formal decision making and consensus process. The process included a Delphi exercise,<sup>7</sup> a systematic literature review,<sup>8</sup> and an empirical data collection with the ICF checklist.<sup>9</sup>

The ICF Core Set for RA has undergone content validation from a patient perspective,<sup>10,11</sup> but the reliability of the qualifiers' scale of the ICF Core Set for RA has not yet been studied. This study had the objective to assess the observer reliability of the ICF Core Set for RA in rheumatological practice. The specific aims were (1) to estimate the inter-rater and intra-rater reliability of the ICF Core Set for RA when applied by health professionals in a specialised rheumatology facility, and (2) to study the metric of the qualifiers' scale.

## PATIENTS AND METHODS

### Study design

We conducted a reliability study with assessments at two time points one week apart in RA patients with stable disease activity.

### Patients

A sample of 25 patients with RA<sup>12</sup> was included in the study after written informed consent. Patients were recruited consecutively from the rheumatology outpatient department at Diakonhjemmet Hospital in Oslo, Norway. The recent disease history was objectified from patients and hospital records.

### Data collection

Patients were interviewed twice, at the first time point (T1) and then 6–10 days later (second time point (T2)). Interviews at T1

**Abbreviations:** ICF, International Classification of Functioning, Disability and Health; MHAQ, Modified Health Assessment Questionnaire; RA, rheumatoid arthritis; RADAI, Rheumatoid Arthritis Disease Activity Index; SF-36, Short Form 36; VAS, visual analogue scale

were performed independently by an occupational therapist (SL) and by a physiotherapist (RHM). Patients were assessed only by one therapist assigned randomly at T2. Both health professionals were working at the department of rheumatology and were trained according to a video issued by the ICF branch of the WHO at the Ludwig-Maximilian University in Munich, Germany.

The health professionals rated the magnitude or the severity of the problem in each of the ICF categories primarily based on the information obtained during the interview and only if necessary and in very few cases from available medical information in the patient charts.

A short introduction to the concepts of the ICF was given in lay terms to all patients at the beginning of each interview. Open ended questions were used to ask patients about their level of functioning in each of the areas referred to in each of the ICF categories contained in the ICF Core Set for RA. Patients also filled in self administered questionnaires while waiting for the interview. The rheumatology professionals had no insight into the questionnaires.

## Measures

### ICF Core Set for RA

The ICF Core Set for RA includes 25 categories from the component Body functions, 18 from the component Body Structures, 32 from the component Activities and Participation, and 21 from the component Environmental Factors.<sup>6</sup> By error, one of the categories (b715 stability of joints function) was omitted. The severity of the patients' problems in each of the ICF categories is quantified with the qualifiers' scale. The qualifiers' scale of the components Body Functions and Structures, and Activities and Participation has five response levels, each ranging from 0 to 4, corresponding to no/mild/moderate/severe/complete impairment. For example, a moderate problem with walking would give a score of 2. The qualifiers' scale of the component environmental factors has nine response levels ranging from -4 to +4. A specific environmental factor can be a barrier (-4 to -1), a facilitator (1 to 4), or can have no influence (0) on the patient's functioning. If the factor has an influence, the extent of the influence (either positive or negative) can be coded with mild/moderate/severe/complete. In addition, there are the response options "8, not specified" and "9, not applicable" for each category.

## Health status

Measures to describe the study population and the stability of the disease between the two time points included visual analogue scales (0–100) for pain, fatigue, and patient global assessment of disease activity, the modified Health Assessment Questionnaire (MHAQ),<sup>13</sup> short form SF-36,<sup>14</sup> and Rheumatoid Arthritis Disease Activity Index (RADAI).<sup>15</sup>

## Data analysis and statistics

We used descriptive statistics to describe the study population. To examine the stability of patients between both assessments the limits of agreement statistics by Bland and Altman were applied.<sup>16</sup> Patients were excluded from the intra-rater reliability analysis if the difference in health status exceeded limits of agreement in two of the seven measures (pain, fatigue, patient global, RADAI, MHAQ, mental health, and physical function from SF-36) over the 6–10 day period between the assessments.

Since the response options of the qualifiers' scale "not specified"<sup>8</sup> and "not applicable"<sup>9</sup> are not part of the ordinal scale ranging from 0 to 4 for the components of functioning (-4 to 4 for environmental factors) they were not included for kappa statistics. Thus, the response options 8 and 9 were considered missing data.

To examine the inter-rater and intra-rater reliability of the ICF categories raw agreement was calculated and weighted kappa statistics were applied to investigate whether there is more agreement than might occur by chance given random guessing. The choice of this approach was based on recent suggestions by, for example, Landis and Koch,<sup>17</sup> to placing more weight on the raw data than on the kappa coefficient, in agreement statistics.<sup>18–19</sup> ICC analyses were not applicable for the categorical analyses in this study.

To explore the metric of the qualifiers' scale, the Rasch model for ordered response levels was used.<sup>20</sup> Reversed threshold estimates provide sufficient evidence to conclude that the empirical ordering is not consistent with the intended ordering. Thresholds define the boundaries between response categories. Collapsing response levels may reveal the effective number and ordering of levels post hoc.<sup>20</sup> Therefore, in case of reversed threshold estimates, response levels were collapsed to obtain an effective number and ordering of response categories post hoc.

Two different dimensions, respectively, were studied based on the Rasch model—namely, the dimension "functioning" and the dimension "environmental factors." In line with the bio-psychosocial model on which the ICF is based, the ICF categories of the components Body Functions and Structures, Activities and Participation were included in the Rasch analysis to study the dimension functioning. The environmental factors categories were included in the Rasch analysis to study the dimension environmental factors. After collapsing response categories agreement statistics were recalculated.

The data analysis regarding descriptive statistics and Bland–Altman plots was performed using SPSS 12.0. Weighted kappa values were analysed using the Statistical Analysis System (SAS version 9.1.3). Rasch analyses were performed with the program RUMM 2020.<sup>21</sup> When applicable, the level of statistical significance was set to  $p < 0.05$ . The study was approved by the regional ethics committee.

## RESULTS

Demographic data and clinical characteristics of the participants are shown in table 1. For the analyses of the intra-rater reliability, two patients were excluded because the disease was not stable between the two time points of assessment. Mean time for assessments with the ICF Core Set for RA was 34.2 (SD 9.1, range 20–75) minutes.

**Table 1** Demographic characteristics and scores for health status measures of patients (n=25) at baseline (mean (SD) for continuous variables, % for counts)

Age (years)	57.5 (12.5)
Female sex	76%
Disease duration (years)	15.9 (14.6)
Pain (100 mm VAS)	35 (23)
Fatigue (0–100 mm VAS)	43 (25)
Patient global (0–100 mm VAS)	34 (25)
MHAQ (1–4)	1.7 (0.6)
RADAI (0–10)	3.9 (2.0)
SF-36 scales (0–100)	
General health	55 (25)
Vitality	43 (21)
Social function	72 (30)
Role mental	49 (43)
Mental health	75 (17)
Physical function	47 (27)
Pain	45 (23)
Role physical	38 (39)

VAS, visual analogue scale; SF-36, Short Form 36; MHAQ, Modified Health Assessment Questionnaire; RADAI, Rheumatoid Arthritis Disease Activity Index.

**Table 2** Body functions: categories from ICF Core Set for RA with intra-rater and inter-rater reliability

ICF code	Body functions ICF category title	Intra-rater			Inter-rater		
		Agreement (%) <sup>*</sup>	Weighted kappa	Missing (%) <sup>†</sup>	Agreement (%) <sup>*</sup>	Weighted kappa	Missing (%) <sup>†</sup>
b130	Energy and drive functions	<b>68</b>	0.58	0	<b>60</b>	0.50	0
b134	Sleep function	<b>60</b>	0.63	0	<b>52</b>	0.59	0
b152	Emotional functions	<b>64</b>	0.58	0	<b>60</b>	0.58	0
	Experience of self and time functions	72	0.33	0	64	-0.01	0
b180	Body image	<b>76</b>	0.47	0	<b>50</b>	0.32	4
b1801	Body image	<b>76</b>	0.47	0	<b>50</b>	0.32	4
b280	Sensation of pain	<b>52</b>	0.32	0	44	0.02	0
b2800	Generalised pain	<b>42</b>	0.47	4	<b>40</b>	0.45	0
b2801	Pain in body part	38	0.13	4	56	0.21	0
b28010	Pain in head and neck	<b>76</b>	0.79	0	<b>40</b>	0.36	0
b28013	Pain in back	<b>56</b>	0.42	0	<b>54</b>	0.54	4
b28014	Pain in upper limb	56	0.29	0	<b>72</b>	0.48	0
b28015	Pain in lower limb	<b>48</b>	0.49	0	<b>56</b>	0.52	0
b28016	Pain in joints	52	0.26	0	<b>64</b>	0.40	0
	Haematological system function	<b>71</b>	0.69	4	50	-0.07	4
b430	Exercise tolerance functions	<b>56</b>	0.65	0	<b>52</b>	0.56	0
b455	Exercise tolerance functions	<b>56</b>	0.65	0	<b>52</b>	0.56	0
b510	Ingestion functions	96	0.00	0	80	0.16	0
b640	Sexual functions	<b>75</b>	0.74	36	60	0.40	40
b710	Mobility of joints functions	<b>56</b>	0.50	0	<b>52</b>	0.48	0
b7102	Mobility of joints generalised	<b>52</b>	0.47	0	<b>60</b>	0.56	0
b730	Muscle power functions	<b>60</b>	0.43	0	52	0.30	0
b740	Muscle endurance functions	<b>64</b>	0.57	0	<b>40</b>	0.31	0
b770	Gait pattern functions	<b>60</b>	0.61	0	<b>48</b>	0.51	0
	Sensations related to muscles and movement functions	<b>68</b>	0.51	0	56	0.26	0
b780	Sensation of muscle stiffness	<b>56</b>	0.60	0	<b>56</b>	0.51	0

\*Printed bold if weighted kappa statistics beyond chance.

†Category is not specified or not applicable.

Missing values of more than 5%, the result of response options “not specified” and “not applicable,” were present in 9/95 categories within raters (9%) and 23/95 categories (24%) between raters.

Tables 2–5 list the ICF categories in the different components Body Functions, Body Structures, Activities and Participation, and Environmental Factors and present intra-rater and inter-rater reliability with percentages for complete agreement and kappa statistics.

Mean intra-rater agreement for all ICF categories was 59% which increased to 72% after collapsing of qualifiers, ranging

from 29% (e340) to 96% (b510) before, and from 44% (e450) to 96% (b510) after collapsing of qualifiers (tables 2–5).

Mean inter-rater agreement for all ICF categories was 47% and increased to 61% after collapsing of qualifiers, ranging from 0% (e450) to 80% (d560) before, and from 8% (d415) to 88% (d560) after collapsing of qualifiers (tables 2–5).

The mean intra-rater agreement per component was 61% for Body Functions, 62% for Body Structures, 60% for Activities and Participation, and 52% in the component Environmental Factors. The mean inter-rater agreement was for Body

**Table 3** Body structures: categories from ICF Core Set for RA with intra-rater and inter-rater reliability

Category	Body structures ICF category title	Intra-rater			Inter-rater		
		Agreement (%)	Weighted kappa	Missing (%) <sup>†</sup>	Agreement (%) <sup>*</sup>	Weighted kappa	Missing (%) <sup>†</sup>
s299	Ear and related structures, specified	<b>48</b>	0.54	0	<b>46</b>	0.32	4
s710	Structure of head and neck region	<b>88</b>	0.86	4	<b>52</b>	0.32	8
s720	Structure of shoulder region	<b>64</b>	0.60	0	<b>42</b>	0.47	4
s730	Structure of upper extremity	<b>52</b>	0.31	0	38	0.22	4
s73001	Elbow joint	<b>76</b>	0.80	0	<b>44</b>	0.48	0
s73011	Wrist joint	<b>52</b>	0.56	0	<b>56</b>	0.61	0
s7302	Structure of hand	<b>48</b>	0.33	0	<b>59</b>	0.49	12
s73021	Joints of hand and fingers	<b>60</b>	0.57	0	<b>48</b>	0.42	8
s73022	Muscles of hand	<b>64</b>	0.50	0	<b>50</b>	0.39	12
s750	Structure of lower extremity	<b>50</b>	0.48	4	50	0.25	12
s75001	Hip joint	<b>72</b>	0.78	0	<b>35</b>	0.41	8
s75011	Knee joint	<b>64</b>	0.65	0	<b>43</b>	0.44	8
s7502	Structure of ankle and foot	<b>60</b>	0.74	0	<b>57</b>	0.43	8
s760	Structure of trunk	42	0.17	4	36	0.05	12
s7600	Structure of vertebral column	52	0.32	0	26	-0.38	8
s76000	Cervical vertebral column	<b>84</b>	0.79	0	<b>68</b>	0.58	0
s770	Additional musculoskeletal structures related to movement	<b>68</b>	0.61	0	17	-0.16	8
s810	Structure of areas of skin	<b>68</b>	0.68	0	<b>52</b>	0.46	0

\*Printed bold if weighted kappa statistics beyond chance.

†Category is not specified or not applicable.

**Table 4** Activities and participation: categories from ICF Core Set for RA with intra-rater and inter-rater reliability

Category	Activities and participation ICF category title	Intra-rater			Inter-rater		
		Agreement (%)	Weighted kappa	Missing (%)†	Agreement (%)*	Weighted kappa	Missing (%)†
d170	Writing	<b>60</b>	0.61	0	<b>56</b>	0.58	0
d230	Carrying out daily routine	<b>40</b>	0.52	0	<b>48</b>	0.41	0
d360	Using communication devices and techniques	<b>80</b>	0.48	0	<b>72</b>	0.26	0
d410	Changing basic body position	<b>68</b>	0.70	0	32	-0.24	0
d415	Maintaining a body position	<b>60</b>	0.65	0	8	-0.40	0
d430	Lifting and carrying objects	<b>52</b>	0.53	0	<b>40</b>	0.33	0
d440	Fine hand use	<b>48</b>	0.53	0	<b>29</b>	0.35	4
d445	Hand and arm use	40	0.27	0	<b>52</b>	0.44	0
d449	Carrying, moving and handling objects, other specified and unspecified	<b>48</b>		0	<b>71</b>		4
d450	Walking	<b>56</b>	0.31	0	<b>36</b>	0.50	0
d455	Moving around	<b>60</b>	0.65	0	24	0.42	0
d460	Moving around in different locations	<b>60</b>	0.62	0	20	0.09	0
d465	Moving around using equipment	<b>40</b>	0.59	0	20	0.00	0
d470	Using transportation	<b>40</b>	0.64	60	33	0.03	64
d475	Driving	<b>68</b>	0.61	0	52	0.28	8
d475	Driving	<b>48</b>	0.56	8	<b>57</b>	0.45	8
d510	Washing oneself	<b>64</b>	0.59	0	<b>48</b>	0.38	0
d520	Caring for body parts	<b>36</b>	0.52	0	<b>36</b>	0.37	0
d530	Toileting	<b>71</b>	0.65	4	<b>76</b>	0.74	0
d540	Dressing	<b>60</b>	0.66	0	<b>60</b>	0.53	0
d550	Eating	64	0.12	0	<b>68</b>	0.40	0
d560	Drinking	<b>76</b>	0.34	0	<b>80</b>	0.60	0
d570	Looking after one's health	72	0.35	0	<b>72</b>	0.43	0
d620	Acquisition of goods and services	<b>52</b>	0.36	0	<b>48</b>	0.37	0
d630	Preparing meals	<b>58</b>	0.68	4	<b>48</b>	0.34	8
d640	Doing housework	<b>60</b>	0.65	0	<b>50</b>	0.36	4
d660	Assisting others	<b>64</b>	0.56	0	55	0.19	12
d760	Family relationships	<b>76</b>	0.62	0	68	0.18	0
d770	Intimate relationships	<b>81</b>	0.00	16	70	-0.04	20
d850	Remunerative employment	<b>69</b>	0.75	48	<b>62</b>	0.56	48
d859	Work and employment, other specified and unspecified	<b>53</b>	0.70	32	58	0.20	52
d910	Community life	<b>68</b>	0.72	0	<b>60</b>	0.41	0
d920	Recreation and leisure	<b>52</b>	0.57	0	<b>40</b>	0.36	0

\*Printed bold if weighted kappa statistics beyond chance.

†Category is not specified or not applicable.

Functions 55%, for Body Structures 46%, for Activities and Participation 51%, and 31% in the component Environmental Factors. Between raters 52% of the ICF categories showed at least 50% agreement (78% after collapsing), and in 77% within raters (99% after collapsing).

Weighted kappa statistics showed reliability of 0.4 or higher in 82/95 ICF categories (86%) within raters, but only in 41/95 ICF categories (43%) between raters (table 6).

Rasch analyses suggested that reduction of the number of qualifiers from five to three—and from nine to three for environmental factors—improved both inter-rater and intra-rater agreement. According to these results, the response levels 1–2 and 3–4 of the ICF categories belonging to Body Functions, Body Structures, and Activities and Participation were collapsed, respectively. In the component Environmental Factors, the response levels from -4 to -1 and from 1 to 4 were collapsed.

Several considerations were thereby taken into account: Firstly, the number of response categories that does not follow the consecutive order intended was considered. Secondly, a further collapsing strategy was studied—namely, the collapsing of the response categories 3 and 4. However, this strategy did not yield satisfactory results as most of the ICF categories still presented response categories that did not have a consecutive order (results not shown). Also, owing to the low frequencies in response categories 3 and 4, no further collapsing strategies, such as collapsing response categories 1 and 2 and 2 and 3, were considered. Finally, the same response format was

intended for all ICF categories. The proposed collapsing strategy is clinically intuitive for judging the severity of a problem in the corresponding ICF categories. After collapsing the response categories, only four ICF categories in the functioning component did not follow a consecutive order and five in the component environmental factors.

## DISCUSSION

In this study the reliability of the ICF Core Set for RA demonstrated only low and at best moderate agreement. Agreement for the individual categories between different health professionals was lower than within the same person. Based on explorative analyses a reduction of the scale qualifiers from five to three (and from nine to three for environmental factors) could improve both inter-rater and intra-rater reliability.

This is the first study to explore and establish the reliability of an ICF Core Set for a specific disease. Extensive testing of the ICF is necessary as the ICF is a WHO adopted classification for global application. Reliability testing of the ICF has so far only been performed in one other study where not a core set, but specified ICF categories were tested for inter-rater reliability in geriatric patients, and a moderate reliability was reported.<sup>22</sup>

ICF categories are not self assessed and a patient's report is being interpreted and scored by an interviewer, leading necessarily to a discrepancy between scores and reduced reliability. Health professionals with different specialties—such as an occupational therapist and a physiotherapist in our study—also have different focus on and awareness for the

**Table 5** Environmental factors: categories from ICF Core Set for RA with intra-rater and inter-rater reliability

Category	Environmental factors ICF category title	Intra-rater			Inter-rater		
		Agreement (%)	Weighted kappa	Missing (%)†	Agreement (%)*	Weighted kappa	Missing (%)†
e110	Products or substances for personal consumption	<b>68</b>	0.61	0	<b>44</b>	0.34	0
e115	Products and technology for personal use in daily living	<b>48</b>	0.58	0	24	-0.12	0
e120	Products and technology for personal indoor and outdoor mobility and transportation	47	0.22	40	<b>6</b>	0.35	36
e125	Products and technology for communication	72	0.27	0	<b>56</b>	0.29	0
e135	Products and technology for employment	<b>57</b>	0.68	44	25	0.29	52
e150	Design, construction and building products and technology for buildings for public use	<b>64</b>	0.67	0	<b>56</b>	0.51	0
e155	Design, construction and building products and technology of buildings for private use	<b>56</b>	0.66	0	<b>32</b>	0.31	0
e225	Climate	<b>64</b>	0.69	0	<b>52</b>	0.53	0
e310	Immediate family	<b>52</b>	0.39	0	25	0.29	4
e320	Friends	<b>44</b>	0.44	0	32	0.07	0
e340	Personal care providers and personal assistants	29	0.15	32	19	0.21	36
e355	Health professionals	<b>46</b>	0.33	4	13	-0.04	4
e360	Other professionals	<b>46</b>	0.49	4	17	0.06	4
e410	Individual attitudes of immediate family members	<b>54</b>	0.56	4	25	0.06	4
e420	Individual attitudes of friends	<b>50</b>	0.38	4	21	-0.02	4
e425	Individual attitudes of acquaintances, peers, colleagues, neighbours and community members	<b>48</b>	0.55	0	<b>44</b>	0.36	0
e450	Individual attitudes of health professionals	<b>44</b>	0.48	0	0	-0.16	4
e460	Societal attitudes	<b>60</b>	0.54	0	<b>52</b>	0.35	0
e540	Transportation services, systems and policies	<b>58</b>	0.58	4	<b>46</b>	0.48	4
e570	Societal security services, systems and policies	<b>56</b>	0.50	0	<b>44</b>	0.49	0
e580	Health services, systems and policies	<b>40</b>	0.55	0	25	0.23	4

\*Printed bold if weighted kappa statistics beyond chance.

†Category is not specified or not applicable.

individual ICF categories. Therefore, one expects better inter-rater agreement if similar professionals have been involved in the judgments.

Within this context it is important to remember that limits of agreement in clinical and health status measures are generally quite wide. This was for example demonstrated when we examined the test-retest reliability for different data collection

methods of self reported health status in patients with RA.<sup>23</sup> In rheumatology, low reliability when grading tender joints is known,<sup>24</sup> and inter-rater agreement hardly exceeds the level of chance.<sup>25</sup> Therefore, it does not surprise us that determined ICF categories of the ICF Core Set for RA show low reliability.

In this study, low reliability was demonstrated especially for categories concerning the environmental factors, in particular

**Table 6** Frequency of observer agreement within and between raters for categories in the ICF Core Set for RA

Observed agreement	All categories (n=95)		Body functions (n=24)		Body structures (n=18)		Activities and participation (n=32)		Environmental factors (n=21)	
	Intra	Inter	Intra	Inter	Intra	Inter	Intra	Inter	Intra	Inter
Higher than 80%	4	0	1	0	2	0	1	0	0	0
51-80%	67	44	20	17	12	6	24	17	11	4
21-50%	24	43	3	7	4	11	7	13	10	12
20% and lower	0	8	0	0	0	1	0	2	0	5
Number categories (%) with kappa values 0.4 or higher	82/95 (86)	41/95 (43)	18/24 (75)	14/24 (58)	14/18 (78)	10/18 (56)	25/32 (71)	13/32 (41)	15/21 (71)	4/21 (19)

inter-rater agreement; categories from this component of the ICF are scored by attributing a positive or negative graded weight. The patient provides the information during the assessment if a factor (for example, family) is a facilitator or a barrier or both. Because an environmental factor may be expressed as both a facilitator and a barrier by the patient, the interviewer may meet the challenge to decide on one overall final judgment when scoring the individual category. In addition, environmental factors may be difficult to assess owing to difficult technical terms that do not always have associations in everyday life or do not have a clear meaning. Another problem of testing the ICF is given by the qualifiers on a mixed scale 0–4 and recording if categories are not specified or not applicable. The ICF qualifiers “8, not specified” and “9, not applicable” are very useful from a clinical point of view but they represent a “barrier” that is difficult to overcome when performing statistical analyses.

Assessment of patients with the ICF Core Set is time consuming. Lack of clarity in wording of the ICF categories may also have produced a need for explanation, leading to variance in time required to complete the assessment (range 20–75 minutes) and to a potential for variance in agreement. Increasing experience of the health professionals in this study did not considerably shorten the time (data not shown), but time use was constant at all three assessments. The reliability of the core set in RA might be improved by means of a training manual.

Our analysis suggests reducing the number of qualifiers when applying the ICF Core Set for RA, because the metric of the five qualifier scale was not sound. In addition to improved reliability, the feasibility of the ICF Core Set in RA could be improved when fewer qualifiers need to be considered by the assessor. The suggested reduction of response qualifiers needs further confirmation in other studies as our finding of improved reliability with fewer response categories was driven by our own data. In addition, the sample size of our study has to be considered. This sample size was comparable with other studies examining reliability.<sup>23–26</sup> The number of ICF categories considered in the dimension functioning and in the dimension environmental factors for the Rasch analyses, respectively, is fairly high in relation to the number of patients in our sample. In future studies, a larger sample size should be included in the Rasch analyses to augment the precision of the parameter estimations of the Rasch model.

A limitation of this study is that it was conducted at one centre and our findings may not be extrapolated to other environments of patients and health professionals. In addition, the results of the Rasch analyses were analysed to derive conclusions regarding the category response functions of the qualifiers' scale and were not further considered to obtain information regarding the fit of the ICF categories to the Rasch model. Further studies are needed to investigate whether the ICF Core Set for RA from the viewpoint of the modern test theory has contributed to building a measure.

A strength of this study is that we rigorously sought to examine reliability of the comprehensive ICF Core Set for RA for both inter-rater and intra-rater agreement. Patients in our sample represent the typical average age group of patients with RA in our outpatient clinic<sup>27</sup> with age comparable to patients in a recent validation of the patient perspective of the ICF Core Set for RA.<sup>10</sup> Both assessors in this study were experienced clinicians with interest in and knowledge of the ICF concept.

In summary, the reliability of the ICF RA core set is now established, being low to moderate in this study and varied considerably across categories and between raters. Our analysis suggests reducing the number of qualifiers when applying the ICF Core Set in RA. It remains to be shown, whether this

empirical observation of improved agreement with reduced number of three response qualifiers in all dimensions may enhance the acceptance and feasibility of the ICF Cores Set in RA. The importance of testing the ICF Core Set for RA is apparent as the ICF has been adopted by WHO, and research on the applicability of the ICF in RA and in other rheumatic diseases is warranted to confirm our findings in this new field. ICF is an ongoing process where results from evaluations and suggestions for further improvement are continuously integrated.

## ACKNOWLEDGEMENTS

This project is part of the ICF Core Sets Validation Study and was supported by a research grant from the European League Against Rheumatism (EULAR). The authors thank all patients who participated in the study.

## Authors' affiliations

**Till Uhlig, Solvår Lillemo, Rikke Helene Moe, Petter Mowinckel, Tore Kristian Kvien**, National Resource Center for Rehabilitation in Rheumatology, Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway

**Tanja Stamm**, Department of Internal Medicine III, Division of Rheumatology, Vienna Medical University, Vienna, Austria

**Alarcos Cieza, Gerold Stucki**, Physical Medicine and Rehabilitation, University of Munich, Munich, Germany

**Annelies Boonen**, Department of Rheumatology, University Hospital, Maastricht, Netherlands

## REFERENCES

- Uhlig T, Kvien TK, Glennäs A, Smedstad LM, Førre O. The incidence and severity of rheumatoid arthritis, results from a county register in Oslo, Norway. *J Rheumatol* 1998;**25**:1078–84.
- Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. *Arthritis Rheum* 2000;**43**:2751–61.
- Sokka T, Kautiainen H, Möttonen T, Hannonen P. Work disability in rheumatoid arthritis 10 years after the diagnosis. *J Rheumatol* 1999;**26**:1681–5.
- Ødegård S, Kvien TK, Finset A, Uhlig T. Physical and psychological predictors for work disability over seven years in patients with rheumatoid arthritis. *Scand J Rheumatol* 2005;**34**:441–7.
- World Health Organization. *International classification of functioning, disability and health: ICF*. Geneva: WHO, 2001.
- Stucki G, Cieza A, Geyh S, Battistella L, Lloyd J, Symmons D, et al. ICF core sets for rheumatoid arthritis. *J Rehabil Med* 2004;**44**(suppl):87–93.
- Weigl M, Cieza A, Andersen C, Kollerits B, Amann E, Stucki G. Identification of relevant ICF categories in patients with chronic health conditions: a Delphi exercise. *J Rehabil Med* 2004;**44**(suppl):12–21.
- Brockow T, Cieza A, Kuhlow H, Sigl T, Franke T, Harder M, et al. Identifying the concepts contained in outcome measures of clinical trials on musculoskeletal disorders and chronic widespread pain using the International Classification of Functioning, Disability and Health as a reference. *J Rehabil Med* 2004;**44**(suppl):30–6.
- Ewert T, Fuessl M, Cieza A, Andersen C, Chatterji S, Kostanjsek N, et al. Identification of the most common patient problems in patients with chronic conditions using the ICF checklist. *J Rehabil Med* 2004;**44**(suppl):22–9.
- Stamm TA, Cieza A, Coenen M, Machold KP, Nell VP, Smolen JS, et al. Validating the International Classification of Functioning, Disability and Health Comprehensive Core Set for Rheumatoid Arthritis from the patient perspective: a qualitative study. *Arthritis Rheum* 2005;**53**:431–9.
- Coenen M, Cieza A, Stamm TA, Amann E, Kollerits B, Stucki G. Validation of the International Classification of Functioning, Disability and Health (ICF) Core Set for rheumatoid arthritis from the patient perspective using focus groups. *Arthritis Res Ther* 2006;**8**:R84.
- 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.
- Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified stanford health assessment questionnaire. *Arthritis Rheum* 1983;**26**:1346–53.
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
- Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. *Rheumatology (Oxford)* 2000;**39**:321–7.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.

- 18 **Brennan P**, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;**304**:1491–4.
- 19 **Feinstein AR**, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;**43**:543–9.
- 20 **Andrich D**. A rating formulation for ordered response categories. *Psychometrika* 1978;**43**:581–94.
- 21 **Andrich D**, Sheridan BE, Luo G. RUMM2020: Rasch Unidimensional Models for Measurement. Perth, Western Australia: RUMM Laboratory, 2002.
- 22 **Okochi J**, Utsunomiya S, Takahashi T. Health measurement using the ICF: test-retest reliability study of ICF codes and qualifiers in geriatric care. *Health Qual Life Outcomes* 2005;**3**:46.
- 23 **Kvien TK**, Mowinckel P, Heiberg T, Dammann KL, Dale Ø, Aanerud GJ, et al. Performance of health status measures with a pen-based personal digital assistant. *Ann Rheum Dis* 2005;**64**:480–4.
- 24 **Uhlig T**, Smedstad LM, Vaglum P, Moum T, Gérard N, Kvien TK. The course of rheumatoid arthritis and predictors of psychological, physical and radiographic outcome after 5 years of follow-up. *Rheumatology (Oxford)* 2000;**39**:732–41.
- 25 **Hart LE**, Tugwell P, Buchanan WW, Norman GR, Grace EM, Southwell D. Grading of tenderness as a source of interrater error in the Ritchie articular index. *J Rheumatol* 1985;**12**:716–7.
- 26 **Guillemin F**, Billot L, Boini S, Gerald N, Øddgård S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. *J Rheumatol* 2005;**32**:766–8.
- 27 **Uhlig T**, Kvien TK, Jensen JL, Axell T. Sicca symptoms, saliva and tear production, and disease variables in 636 patients with rheumatoid arthritis. *Ann Rheum Dis* 1999;**58**:415–22.

### Access a vast information database with Toll-Free linking

“Toll-free” linking gives you immediate access to the full text of many of the cited articles in a paper’s reference list—FOR FREE. With the support of HighWire’s vast journal catalogue, a huge reference library is now open to you. If HighWire hosts the journal, you can view the full text of the referenced article, completely free of charge by following the Free Full Text links in the references.



## Reliability of the ICF Core Set for rheumatoid arthritis

Till Uhlig, Solvår Lillemo, Rikke Helene Moe, et al.

*Ann Rheum Dis* 2007 66: 1078-1084 originally published online January 12, 2007

doi: 10.1136/ard.2006.058693

---

Updated information and services can be found at:

<http://ard.bmj.com/content/66/8/1078.full.html>

---

*These include:*

### References

This article cites 25 articles, 3 of which can be accessed free at:  
<http://ard.bmj.com/content/66/8/1078.full.html#ref-list-1>

Article cited in:

<http://ard.bmj.com/content/66/8/1078.full.html#related-urls>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Topic Collections

Articles on similar topics can be found in the following collections

[Connective tissue disease](#) (3180 articles)  
[Degenerative joint disease](#) (3441 articles)  
[Immunology \(including allergy\)](#) (3761 articles)  
[Musculoskeletal syndromes](#) (3691 articles)  
[Rheumatoid arthritis](#) (2408 articles)  
[Other rehabilitative therapies](#) (9 articles)

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>