Disability and Rehabilitation, 2010; 32(17): 1397–1405

RESEARCH PAPER

Validation of the International Classification of Functioning Disability and Health framework using multidimensional item response modeling

THOMAS EWERT[1], DIANE D. ALLEN[2], MARK WILSON[3], BEDIRHAN ÜSTÜN[4] & GEROLD STUCKI[1]

[1]*Department of Physical Medicine and Rehabilitation, Ludwig-Maximilian University, Munich, Germany,* [2]*Graduate Program in Physical Therapy, University of California San Francisco/San Francisco State University, San Francisco, CA, USA,* [3]*Graduate School of Education, University of California, Berkeley, CA, USA, and* [4]*World Health Organization, Classification, Assessment, Surveys and Terminology Team, Geneva, Switzerland*

Accepted January 2010

Abstract
Purpose. To examine the construct validity of the International Classification of Functioning Disability and Health (ICF) framework using multidimensional item response modelling and data collected in different regions from patients with five chronic health conditions. We assume that the ICF components should represent statistically called dimensions that are distinct although related.
Method. Retrospective validation study using the ICF Core Sets from a convenience sample of patients in an international multicentre, cross-sectional database obtained in different rehabilitation centres. Health professionals working in 89 rehabilitation centres in 32 countries collected data from 3227 rehabilitation patients using the respective ICF Core Sets. Patients included had one of the following health conditions: low back pain (LBP), rheumatoid arthritis (RA), osteoarthritis (OA), obesity (OB) or post-stroke. Data from questions regarding a patient's functioning based on body structures and functions, activities, participation along the ICF Core Sets were analysed with multidimensional item response modelling.
Results. The multidimensional models fit the data better than a model with few or no specifications regarding an underlying framework. For example, a model separating four dimensions 'body structures', 'body functions', 'activities' and 'participation' fits the data better than a model differentiating between 'body functions and structures' and 'activities and participation'. The ICF framework with its components represents underlying statistically called dimensions.
Conclusion. The results of this study support the construct validity of the functioning part of the ICF. The distinct dimensions may facilitate the alignment of ICF components with other measures used clinically and in research. Based on our results it is justifiable to construct instruments integrating ICF categories within components.

Keywords: *International Classification of Functioning Disability and Health, multidimensional item response modeling*

Introduction

Optimising functioning through rehabilitation is a major goal of society. A basis for optimal rehabilitation is the comprehensive understanding of functioning, which is based on a general valid conceptual framework and suitable measurement. With the approval of the International Classification of Functioning Disability and Health (ICF) by the 54th World Health Assembly in 2001 [1] the World Health Organization (WHO) provided a universal and globally accepted framework and classification that comprehensively addresses human functioning and disability [2,3]. The ICF is a multipurpose classification which is not restricted to a single health condition, health sector or professional discipline. The ICF classifies both functioning and contextual factors. The functioning part includes four components: body structures (S), body functions (B), activities (A) and participation (P). The contextual factors include environmental facilitators and barriers along with personal factors that may influence functioning. The ICF therefore provides a promising starting point for the integrative understanding of functioning and disability.

Correspondence: Prof. Gerold Stucki, Department of Physical Medicine and Rehabilitation, Ludwig-Maximilian University, Munich, Germany.
E-mail: gerold.stucki@med.uni-muenchen.de

ISSN 0963-8288 print/ISSN 1464-5165 online © 2010 Informa UK Ltd.
DOI: 10.3109/09638281003611037

Since the approval of the ICF there have been a wide range of implementation activities worldwide [4,5,7,8]. For example, the outcome measures in rheumatoid arthritis clinical trials have decided to use the ICF as their reference framework for the development of measures and standards [9]. Also, an increasing number of research papers are referring to the ICF when examining or reporting aspects of functioning [10,11,12].

These examples show that the ICF is not solely used for theoretical purposes but also for empirical work. The education, clinical and research applications show that the ICF is used in a different and much more comprehensive manner than the International Classification of Diseases (ICD) [13]. When data are collected and summarised using the ICF, it could be done in a qualitative way (like verbal description about relevant chapters or components) or quantitative (scoring) way. Without testing the structure of the ICF and the relationships between its components, interpretation of some research results may be questionable. To our knowledge, there are few publications dealing with the components of the ICF based on empirically derived data [14,15], and these do not include all of the components of functioning. Thus, a validation of the ICF framework could facilitate accurate interpretation of several types of disability studies and promote more meaningful research in these areas.

Analyses of data collected using the ICF categories can validate the framework of the ICF, showing that the allocation of categories to statistically called dimensions such as B, S, A and P is meaningful. This means the ICF components should represent statistically called dimensions that are distinct although related. Statistically, a model specifying the ICF framework should fit the data better than a model with few or no specifications regarding an underlying framework. The question of statistically called dimensions is important for the development of questionnaires or other measures based on the ICF [16]. If empirical data can be structured according to the components of the ICF, it means that those data could be interpreted in terms of the ICF components. Thus would facilitate the provision and the interpretation of those data.

The objective of our study was to examine the construct validity of the functioning part of the ICF using multidimensional item response modelling to analyse data collected from patients with five chronic health conditions in different regions. As a multipurpose classification we expected that the ICF components would apply to different health conditions across different regions but we also assumed that there would be some variation among groups.

Methods

Study design

In this study, we analysed data that were collected during an international multicentre, cross-sectional study aimed at testing and validating the ICF Core Sets in a convenience sample of patients. The patients included in the study had one of the following health conditions (the respective ICD-10 codes are denoted in parentheses): low back pain (LBP): (M54), rheumatoid arthritis (RA): (M05–M06), osteoarthritis (OA): (M19), obesity (OB): (E 65–E68), and stroke: (I60-I64). The study protocol and informed consent forms were approved by the local Ethics Committees. Inclusion criteria for patients were: a diagnosis of one of the five mentioned health conditions according to the ICD-10, at least 18 years of age, sufficient knowledge of the official language of the corresponding country, comprehension of the purpose of the study, and signed informed consent. Comorbidity was not an exclusion criterion if one of the above mentioned health conditions was a main diagnosis. A main diagnosis was defined as a disease clearly dominates another disease including signs and symptoms as well as the corresponding treatment. This was assessed by the health professional, conducted the data collection.

Data collection procedures

Patient recruitment and data collection were performed by health professionals at each study center. The German health professionals were trained in a structured one-day workshop by researchers of the WHO ICF Collaborating Center from the Ludwig-Maximilan University in Munich. The health professionals from other countries were trained with a training video structured similarly to the one-day workshop, showing the relevant issues for the appropriate data collection. The training involved familiarisation with the principles of the ICF as well as the practical application of the ICF Core Sets and the questionnaires used. All health professionals involved in the data collection obtained the same written training materials. As generic self-administered questionnaire the Medical Outcome Study Short Form 36 (SF 36) [17] was provided in the official language of the corresponding country. The data based on the ICF Core Sets were documented by the trained health professionals according to their judgment. The documentation sheet contained the name and the original definition of the ICF categories. All ICF Core Sets case record forms were provided in English.

Measures

The ICF functioning part comprises four components. Body functions (B) refers to functions of body systems, such as 'memory functions'; body structures (S) refers to anatomic parts such as 'structure of trunk'. Activities (A) refer to 'task or action execution by the individual' such as 'doing housework'. Participation (P) refers to 'involvement in life situations' such as 'family relationships'. There is a common list in the ICF that includes both activities and participation. Categories may be interpreted as both activities and participation items (coded both A and P), or as activities and participation items without overlap (coded either A or P).

In its original form the ICF has more than 1400 categories. This quantity is scarcely practical in clinical practice or research thus more feasible tools such as the ICF Core Sets have been developed [18–21]. ICF Core Sets are subsets of ICF categories that can serve as minimal standards for different purposes. The comprehensive ICF Core Set for a specific health condition is a list of ICF categories that includes as few categories as possible to be practical, but as many as necessary to sufficiently describe a spectrum of functioning and health in a comprehensive, multidisciplinary assessment for the given health condition [19].

Twelve tentative ICF Core Sets for persons with chronic health conditions, including RA [22] OA [23], LBP [24] stroke [25] and OB [26] have been developed. Additionally, ICF Core Sets for patients in acute care hospitals and early post-acute care rehabilitation facilities have been developed [21]. Three different ICF Core Sets can be applied for patients post-stroke, namely, the comprehensive ICF Core Set for stroke [25] and the ICF Core Sets for patients with neurological conditions in acute care hospitals [27] and early post-acute care rehabilitation facilities [28]. These ICF Core Sets were combined to create the extended ICF Core Set for stroke. This Set contains all ICF categories that have been selected for any of the three ICF Core Sets mentioned above. The extended ICF Core Set for stroke contains 166 categories of the ICF, 129 of which come from the functioning part of the ICF. For each of the four other health conditions we used the respective comprehensive ICF Core Set.

To enable the comparison between different health conditions, we identified a reduced set of categories which are shared by all five health conditions. Given that only one category from body structures was common to all five health conditions we excluded this category (s750) from the reduced set and conducted the analyses with the remaining 21 categories.

To describe the population, age, gender, years of formal education, in- and outpatient status and the SF-36 [17] were recorded. The SF-36 is a generic health status measure and enables therefore a comparison across different health conditions. We used the two summary scores, the physical component summary score (PCS) and the mental component summary score (MCS) in this study. For every country we used the validated form of the official language of the questionnaire.

Analysis

Multidimensional item response theory (MIRT) combines item response methods with the opportunity to test for different structures as is well-known in confirmatory factor analysis [29]. In contrast to one-dimensional item response analyses, MIRT uses the information from the response to any item to help estimate probabilities of item responses across all hypothesised dimensions. Because the level of measurement of the ICF qualifier is not well examined and the theory behind the ICF proposes a clear structure, the method seemed appropriate. The partial credit multidimensional random coefficients multinomial logit (MRCML) model is an extension of the Rasch family of one-parameter item response models; it is a direct extension of the model as described by Wilson et al. [30]. The MRCML model was developed to have sufficient flexibility to represent a wide range of Rasch family models, including those that apply to scales having either dichotomous or Likert-type responses. The MRCML model can also apply to scales having multiple items arising from the same stem or having more complex relationships.

To test the specific aims, we fit models to the data, using the Monte Carlo Approximation as implemented in the ConQuest software [31] with 2000 nodes. We used multidimensional between-item models, meaning that each category serves as an indicator for just one primary dimension. We used distinct sets of activities and participation categories, with no overlap of categories. More precisely, in the models with different dimensions for activities and participation, categories from chapter one (learning and applying knowledge) to six (domestic life) were assigned to the activity dimension and categories from chapter seven (interpersonal interactions and relationships) to nine (community, social and civic life) were assigned to the participation dimension.

First, we determined the model that best fit the data for a certain health condition. Then, we tested that model with a smaller set of items, for a different health condition, and between multiple regions. In the context of the validation of the ICF the correlations and therefore the differentiation between dimensions must make sense according to the purpose and characteristics of the data (health condition, environment, inpatients *vs.* outpatients etc.). Low correlation among all of the ICF dimensions is not necessary for validity because from a clinical perspective we would expect high associations between some dimensions. To determine, whether a multidimensional model fits better than a unidimensional model is only a first step towards exploring the validity of the ICF. Statistically better models do not necessarily have a clinically meaningful advantage. Therefore, correlations between statistically called dimensions provided additional information to justify whether to consider two components of the ICF as one dimension or not.

All descriptive analyses were calculated with SPSS 15.0. For all item response models the ConQuest software [31] was used. Data coding. ICF categories are scored using a generic qualifier with the following gradation: 0 = no problem (none, absent, negligible,...) 0–4%, 1 = mild problem (slight, low,...) 5–24%, 2 = moderate problem (medium, fair,...) 25–49%, 3 = severe problem (high, extreme,...) 50–95%, 4 = complete problem (total,...) 96–100%. Additionally 8 (not specified) is used when the available information does not suffice to quantify the severity of the problem, and 9 (not applicable) when a category is not applicable in that patient. The scores 'not specified' (8) and 'not applicable' (9) were recoded into missing values.

Specific aim one was to test the dimensionality of the dimensions according to the ICF and explore the differentiation between the dimensions within an extended set of 129 ICF categories as collected for persons with one health condition, post-stroke. Therefore, we compared the fit of various MIRT models of data from the extended set of ICF categories, using post-stroke as the specific health condition. In successive models we assigned categories to possible dimensions: body functions, body structures, activities, and participation. Four models were compared: unidimensional (all categories together, model B/S/A/P), 2-dimensional (body functions/structures and activities/participation, model B/S-A/P), 3-dimensional (activities and participation separated from body functions/body structures, model B/S-A-P), and 4-dimensional (all possible dimensions separated, model B-S-A-P). The model fit was compared on the basis of the differences in the deviances and the numbers of parameters by use of the G^2 likelihood ratio statistic. The difference between the deviances for nested models functions like a χ^2 distribution with the difference in the number of parameters as the degrees of freedom. Because of the large sample size we set $\alpha = 0.01$ for the G^2 test to detect differences in the model fit. We obtained the correlation between the dimensions to gather information about the differentiation between the statistically called dimensions. In addition, we report the numbers of misfitting categories (items) using two parameters of fit. First, for the 'weighted' mean square we used an upper bound of 1.33 and a lower bound of 0.75 [30]. Second, for the weighted t statistic, we used any value $<(2.00$ or > 2.00 [29]. A category was misfitting if both fit-parameters showed misfit. The reliability of the models and their dimensions was calculated based on expected *a posteriori* estimation based upon plausible values (EAP/PV). We considered reliability along with correlation, as we would not be satisfied with differentiation between dimensions (as shown by the correlations) if the reliability was judged to be too low for practical purposes.

Specific aim two was to test whether the dimensionality according to the ICF and differentiation on the level of components found in the extended set of ICF categories is retained in a reduced set of 21 ICF categories for persons post-stroke. To examine this aim, we calculated three different models for people post-stroke in a reduced set of ICF categories shared by all five health conditions and compared them with the results from the extended Core Set.

The specific aim three was to test whether the dimensionality and differentiation found in the set of 21 ICF categories shared by five health conditions can be confirmed in data from patients with OA. Therefore, we compared the same three models as in aim two (using data from the set of ICF categories shared by five health conditions as obtained in persons post-stroke) in persons with OA.

The specific aim four was to test whether the dimensionality and differentiation is similar in the reduced set of 21 ICF categories in two different regions of the world using patients from German-speaking countries *versus* patients from Serbia-Montenegro. To examine this specific aim, we compared data obtained across all health conditions in German-speaking countries and Serbia-Montenegro as two of the largest regional groups represented in our data set. With these two groups we again used the set of 21 ICF categories shared by five health conditions to determine whether the models with more dimensions fit the data better.

Results

The data analysed in this study were collected from October 2003 to December 2006 by health professionals working in 89 rehabilitation centres in 32 countries. The data were collected by physicians (2.6%), physiotherapists (4.4%), psychologists (4.8%), occupational therapists (5.6%), speech therapists

(5.8%), nurses (5.8%), social workers (5.8%) and others (10.4%). No information about the person collecting the data was given for 54.8% of the data. Of 3702 patients with data, 475 patients were excluded from the analysis because they did not fulfill the inclusion criteria. Table I shows the characteristics of patients included in the study including the SF-36 summary scores.

Table II shows the categories involved in the extended and reduced sets of ICF categories used in the analyses as well as the valid number of cases for each category. The extended data set for patients post-stroke, which includes 129 categories from the ICF functioning part, consists of all collected ICF categories used for patients post-stroke. The proportion of missing categories in this set varies from 0.7% (d330, speaking) to 67.5% (b660, procreation functions). From the 21 categories common to all five health conditions, d850 (remunerative employment) had the highest proportion of missing values (44%) and d510 (washing oneself) the lowest (0.6%).

The different models across different samples of ICF categories, health conditions or regions are shown in Table III. The reliability associated with each model varies from 0.97 to 0.65. To answer aim one, the extended ICF Core Set containing 129 categories from the patients post-stroke showed significantly better model fit for each dimension added. In general, the number of misfitting categories also decreased.

When applying the different models to the reduced set of 21 categories, the same patterns occur as for the extended set of categories: models having more dimensions fit better (see Table III). To ensure that the result was not predominately determined by the stroke data set, we re-calculated the models when including data from all health conditions except post-stroke. The results did not change (details not shown).

When comparing post-stroke and OA for the 21 categories, the better model is the model with more dimensions. Comparing models with different numbers of dimensions in data from different regions revealed that the correlations among the dimensions are higher for the Serbia-Montenegro sample for all dimensions. The three dimensional model shows more misfitting categories than the two dimensional model, but the three-dimensional model fit is better for both regional samples. The composition of the German-speaking sample differs from the sample from Serbia-Montenegro with respect to health condition and gender. In the latter sample there are 67% female (*vs.* 57% in the German-speaking sample) and the predominate health condition is OB (68% *vs.* 20%). Moreover, there were no patients post-stroke in the sample from Serbia-Montenegro (*vs.* 38% in the German speaking group).

Discussion

The results of this study support the construct validity of the functioning part of the ICF, because a model specifying the ICF framework fit the data better than a model with few or no specifications regarding an underlying framework. Based on our results it is justifiable to construct instruments integrating ICF categories within these ICF components.

This paper demonstrates a methodological approach to examine the dimensionality of the ICF based in empirical data. Our analyses show correlations between the dimensions of 0.36 and 0.93, which might suggest the merging of components for assessment of ICF categories. However, models fit significantly better when the components were analysed as separate dimensions. The separation of the dimensions provides additional information. Despite the determination of distinct dimensions within the ICF, it can also be appropriate to create one single dimension. Whether one wants to integrate B, S, A and P into one dimension of functioning or to assess the components separately

Table I. Patient characteristics.

Health condition	LBP	OA	OB	RA	Stroke	All conditions together
No. patients (*n*)	468	624	539	763	833	3227
Gender (female %)	58.80	75.50	70.10	80.70	40.10	64.30
Age (years, median)	51.54	66.83	45.12	59.84	65.21	59.70
Years of education (median)	12.00	12.00	12.00	11.00	11.00	12.00
Living alone (%)	17.50	20.20	14.10	19.70	26.80	20.40
Inpatient (%)	36.54	58.49	2.41	42.26	60.62	42.73
Outpatient (%)	37.39	30.61	95.91	49.54	27.01	46.07
SF-36 PCS (median)	30.77	28.39	45.06	28.39	33.43	30.55
SF-36 MCS (median)	46.83	49.05	45.37	46.01	47.39	47.30

LBP, low back pain; OA, osteoarthritis; OB, obesity; RA, rheumatoid arthritis; SF-36 PCS, physical component summary; SF-36 MCS, mental component summar.

Table II. The two different sets of ICF categories used in the analyses.

Body functions	Body structures	Activities and participation
Set 1: Names of the chapters and category codes from 129 categories of the extended ICF Core Set for persons post-stroke		
Mental functions: b110 (823); b114 (825); b117 (750); b126 (777); b130 (799); b134 (762); b140 (807); b144 (804); b147 (474); b152 (701); b156 (795); b160 (455); b164 (744); b167 (800); b172 (739); b176 (803); b180 (808)	*Structures of the nervous system:* s110 (716); s120 (379); s130 (402)	*Learning and applying knowledge*: d110 (480); d115 (815); d120 (462); d130 (441); d135 (424); d155 (728); d160 (788); d166 (781); d170 (788); 172 (737); d175 (750); d177 (403)
Sensory functions and pain: b210 (754); b215 (757); b230 (458); b235 (461); b240 (454); b260 (794); b265 (795); b270 (776); b280 (791)	*Structures of the cardiovascular, immunological and respiratory systems*: s410 (626); s430 (416)	*General tasks and demands*: d210 (805); d220 (767); d230 (808); d240 (727)
Voice and speech functions: b310 (814); b320 (819); b330 (808); b340 (407)	*Structures related to the digestive, metabolic and endocrine systems*: s530 (416)	*Communication*: 310 (807); d315 (766); d325 (768); d330 (827); d335 (786); d345 (765); d350 (816); d360 (778)
Functions of the cardiovascular, haematological, immunological and respiratory systems: b410 (666); b415 (655); b420 (652); b430 (386); b435 (397); b440 (436); b450 (454); b455 (752)	*Structures related to movement*: s710 (428); s720 (727); s730 (742); s750 (755)	*Mobility*: d410 (823); d415 (826); d420 (814); d430 (820); d440 (820); d445 (820); d450 (823); d455 (767); d460 (806); d465 (713); d470 (717); d475 (517)
Functions of the digestive, metabolic and endocrine systems: b510 (787); b515 (426); b525 (760); b530 (459); b535 (423); b540 (384); b545 (370); b550 (406)	*Skin and related structures*: s810 (428)	*Self-care*: d510 (824); d520 (813); d530 (814); d540 (822); d550 (821); d560 (493); d570 (761)
Genitourinary and reproductive functions: b620 (773); b630 (426); b640 (271)		*Domestic life*: d620 (718); d630 (681); d640 (658)
Neuromusculoskeletal and movement-related functions: b710 (792); b715 (787); b730 (814); b735 (809); b740 (783); b750 (750); b755 (785); b760 (810); b770 (813);		*Interpersonal interactions and relationships*: d710 (785); d750 (762); d760 (745); d770 (454)
Functions of the skin and related structures: b810 (430)		*Major life areas*: d845 (286); d850 (295); d855 (314); d860 (663); d870 (652)
		Community, social and civic life: d910 (627); d920 (684); d930 (343); d940 (377)
Set 2: category codes from 21 categories common to all five health conditions		
b130 (3156); b134 (3109); b152 (3147); b280 (3172); b710 (3165);		d410 (3198); d415 (3202); d430 (3184); d450 (3201); d455 (3013); d470 (3043); d475 (2213); d510 (3207); d530 (3196); d540 (3206); d620 (3066); d640 (2959); d770 (2238); d850 (1807); d910 (2788); d920 (2993)

The numbers in brackets denote the valid cases of the respective ICF category.

depends on the purpose. For example, in an epidemiological study it may be more interesting and informative to use component scales which will allow examination of the relationship and the relative contribution of each component to a predicted variable such as mortality. On the other hand when predicting resource utilisation from a health system perspective an overall score would probably be more feasible. We recommend that researchers determine whether to analyse data unidimensionally or multi-dimensionally based on the purpose and dimensionality that is most appropriate. Therefore, we provide no global recommendation for the most appropriate dimensions when applying the ICF.

As with most retrospective analyses of dimensionality, we found categories that did not fit the assumed underlying dimension. The fact that some categories did not match an underlying dimension does not mean they are not important. It may be that in our study misfitting categories could still be carefully integrated in instruments addressing respective dimensions. The misfitting categories are different across different models and across health condition. There is no category which shows a consistent misfit. It goes beyond the scope of this study to hypothesise why certain categories did not fit, but one possible reason for misfitting categories is a low category to dimension ratio. For some categories missing values might be a factor contributing to misfit. For example, d850 (remunerative employment) had the highest proportion of missing values. Since the age of our patients was high it is likely that many were already retired and therefore the category was coded as not applicable. Generally, the reliabilities for the S

Table III. Comparison of models used for different aims.

Model	Category set	Aim 1 Extended Stroke	Aim 2 Common	Aim 3 Common	Aim 4		
					Common	Common	Common
	Condition	Stroke	Stroke^{+}	OA^{+}	All^{+}	All^{+}	All^{+}
	Region	All	All	All	All	German SC	Montenegro
	Cases (*n*)	833	833	624	3227	686	583
	No. categories	129	21	21	21	21	21
B/S/A/P	Deviance	170637.61	35083.60	28882.27	146380.00	31233.70	21194.85
	No. parameters	491	85	85	85	85	85
	No. misfitting cat. (over discriminating)	18 (0)	6 (3)	2 (0)	4 (0)	7 (3)	8 (4)
	Misfitting cat.	b134, b410, b415, b420, b435, b530, b540, b545, b630, b640, s120, s130, s410, s730, s810, d870, d940	b134, b280, d410$^{\&}$, d510$^{\&}$, d540$^{\&}$, d850	b134, b280	b134, b152, b280, d770	b130, b134, b152, b280, d510$^{\&}$, d530$^{\&}$, d620$^{\&}$	b134, b152, d430$^{\&}$, d530, d620$^{\&}$, d640$^{\&}$, d770, d850$^{\&}$
	Reliability B/S/A/P (cat.)	0.97 (129)	0.94 (21)	0.93 (21)	0.92 (21)	0.94 (21)	0.92 (21)
B/S-A/P	Deviance	161863.65	34221.52	28673.32	143447.83	30038.49	20892.23
	No. parameters	493	87	87	87	87	87
	No. misfitting cat.	4 (2)	4 (1)	1	1 (0)	2 (1)	5 (2)
	Misfitting cat.	d510$^{\&}$, d520$^{\&}$, d850, d870	d540$^{\&}$, d770, d850, d910	d770	d770	d640$^{\&}$, d770	b152, d510, d620$^{\&}$, d770, d910$^{\&}$
	Reliabilitity B/S (cat.)	0.93 (70)	0.73 (5)	0.76 (5)	0.72 (5)	0.67 (5)	0.84 (5)
	Reliabilitity A/P (cat.)	0.94 (59)	0.95 (16)	0.90 (16)	0.92 (16)	0.87 (16)	0.93 (16)
	r B/S-A/P	0.88	0.68	0.74	0.72	0.47	0.92
B/S-A-P	Deviance	161153.76	33693.11	28635.67	142680.92	29871.40	20762.05
	No. parameters	496	90	90	90	90	90
	No. misfitting cat. (over discriminating)	4 (3)	4 (1)	0	1 (0)	3 (0)	7 (3)
	Misfitting cat.	d510$^{\&}$, d520$^{\&}$, d540$^{\&}$, d870	d470, d475, d540$^{\&}$, d620		d770	d415, d475, d770	b152, b280$^{\&}$, b710$^{\&}$, d510, d530, d620$^{\&}$, d770
	Reliabilitity B/S (cat.)	0.97 (70)	0.69 (5)	0.74 (5)	0.77 (5)	0.65 (5)	0.77 (5)
	Reliabilitity A (cat.)	0.96 (46)	0.94 (12)	0.90 (12)	0.92 (12)	0.89 (12)	0.85 (12)
	Reliabilitity P (cat.)	0.87 (13)	0.74 (4)	0.80 (4)	0.83 (4)	0.76 (4)	0.79 (4)
	r B/S-A	0.88	0.62	0.74	0.71	0.44	0.91
	r B/S-P	0.78	0.77	0.67	0.69	0.50	0.93
	r A-P	0.82	0.77	0.88	0.86	0.87	0.91
B-S-A-P	Deviance	160694.55	n.a.	n.a.	n.a.	n.a.	n.a.
	No. parameters	500					
	No. msifiting cat. (over discriminating)	5 (2)					
	Misfitting cat.	d240, d330, d520$^{\&}$, d540$^{\&}$, d870					
	Reliabilitity B (cat.)	0.92 (59)					
	Reliabilitity S (cat.)	0.77 (11)					
	Reliabilitity A (cat.)	0.93 (46)					
	Reliabilitity P (cat.)	0.87 (13)					
	r B-S	0.62					
	r S-A	0.60					
	r S-P	0.36					
	r B-A	0.87					
	r B-P	0.80					
	r A-P	0.81					

-,separation between components; /, components were analysed statistically as one dimension; &, over discriminating category; German SC, German speaking countries (D/A/CH); cat., categories; (cat.), no. of categories; r, correlation; n.a., not applicable; $^{+}$, no categories from S covered; Extended stroke, categories of the extended ICF Core Set for post-stroke; Common, categories common to all five health conditions.

tended to be lower than for A or other dimensions. This may be not only because of the limited numbers of categories included, but also because the S are more difficult to document or to rate.

The correlations among the dimensions seem plausible. There are high correlations between A and P. The weakest association tends to be between S and P. It is a daily experience in rehabilitation practice that the impairments of S are less important than impairments in B when assessing different A or P. There is currently no literature available to compare the results directly because the instruments used are predominately not based on the ICF. Moreover, when using different questionnaires to measure an identical content the amount of the association may differ due to different operationalisations. Nevertheless, there is some support for the correlations obtained from this study in the literature [32,33].

From this study we cannot explore all of the reasons for the more homogeneous correlations among the dimensions for Serbia-Montenegro in comparison to the sample from the German-speaking countries. The difference between regions is likely the result of the different distribution of health conditions in the two samples. Additionally, the case record forms in which the ICF categories of the ICF Core Sets contained were in English. Therefore a second reason may be the differences in the data collection. The information was coded and therefore processed by health professionals who were trained in the ICF, but who could have had differences in scoring, or differences in level of English ability for interpreting the ICF Core Sets. In addition, the German health professionals received a one-day workshop for training; all other health professionals were trained by a video and complementary material. This may also have had an impact on the data. And third, we do not know whether the persons of different cultures provided information differently about patterns of disability. More than half of the data were collected by unknown type of health professionals. The distribution of different health professionals collecting the data might have also been contributed to different results, assuming that each health profession has its own professional view.

Our study has a number of limitations. The sampling process was based on convenience samples from multiple countries. We lacked a systematic approach to get representation across world regions, additional health conditions and health care settings besides rehabilitation centres. Additionally, we did not control for many factors influencing human functioning [34]. We did not cover the whole spectrum of chapters or all categories of the ICF, thus limiting the generalisability of the results.

Because we used categories from the ICF that had been rated by health professionals, we would not expect identical results from patient-reported questionnaires. Nevertheless despite different views of health professionals and patients on patients' functioning [35], it would be interesting to compare both judgements on the components of the ICF.

We did not exclude patients with comorbidities. As a consequence, we assume to have more variance in the data which is not covered by our models. Thus would make it more difficult to test the structure of the data. Despite this disadvantage we were able to confirm the structure of the ICF framework. Since patients with comorbidities reflecting the 'reality' of rehabilitation we interpret this as part of ecological validity of our study.

We did not fully explore the psychometric properties of the generic ICF qualifier scale which was used in the data collection process. With respect to the distinction between A and P we do not want to overinterpret the results. The different contents of the chapters may have contributed to the distinction between the responses to A and P categories; other settings (like acute rehabilitation) or health conditions may show different results. Despite these limitations, the presented result is plausible. For other purposes it is arguable that a common list of activities and participation may be preferable since the two dimensions were highly related in our sample. A recent empirical study found in contrast to an earlier study [15] no clear distinction between A and P [14]. Since we directly assessed the ICF domains, did not use a questionnaire, and performed different statistical analyses, the results are not comparable. However, the meaning of A and P may differ for persons with different health status, health conditions and culture.

In conclusion the components and overall functioning as described in the ICF represent distinct underlying statistically called dimensions. These distinct dimensions may facilitate the alignment with ICF components of other measures used clinically and in research, thus improving the communication of findings across world regions and between health care professionals familiar with the ICF. Relatively high correlations between dimensions and small numbers of misfitting categories in the unidimensional models also support the integration of data across components to provide an overview of the functioning of a patient.

Acknowledgement

The authors wish to thank Dr. A. Cieza, Dr. A. Weidmann, G. Immich, A. Garza, M. Kirschneck and H. Gall for their thoughtful comments, data collection conduction and planning as well as support preparing the manuscript.

References

1. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva: World Health Organization; 2001.
2. Stucki G, Grimby G. Foreword: applying the ICF in medicine. J Rehabil Med 2004;44:5–6.
3. Grimby G, Melvin J, Stucki G. The ICF: a unifying model for the conceptualization, organization and development of human functioning and rehabilitation research. Foreword. J Rehabil Med 2007;39:277–278.
4. Jelsma J. Use of the International Classification of Functioning, Disability and Health: a literature survey. J Rehabil Med 2009;41:1–12.
5. Allan CM, Campbell WN, Guptill CA, Stephenson F, Campbell KE. A conceptual model for interprofessional education: the International Classification of Functioning, Disability and Health (ICF). J Interprof Care 2006;20:235–245.
6. Stucki G, Cieza A, Ewert T, Konstanjsek N, Chatterji S, Ustun TB. Application on the international classification of functioning, disability and health (ICF) in clinical practice. Disabill Rehabil 2002;24:281–282.
7. Stucki G. International Classification of Functioning, Disability and Health (ICF): a promising framework and classification for rehabilitation medicine. Am J Phys Med Rehabil 2005;84:733–740.
8. Frontera WR, Grimby G, Basford J, Muller D, Ring H. Publishing in physical and rehabilitation medicine. J Rehabil Med 2008;40:161–165.
9. Stucki G, Boonen A, Tugwell P, Cieza A, Boers M. The World Health Organisation International Classification of Functioning, Disability and Health: a Conceptual Model and Interface for the OMERACT Process. J Rheumatol 2007;34: 600–606.
10. Soberg HL, Sandvik L, Ostensjo S. Reliability and applicability of the ICF in coding problems, resources and goals of persons with multiple injuries. Disabil Rehabil 2008;30:98–106.
11. Anttila H, Autti-Ramo I, Suoranta J, Makela M, Malmivaara A. Effectiveness of physical therapy interventions for children with cerebral palsy: a systematic review. BMC Pediatr 2008; 8:14.
12. Jerosch-Herold C, De Carvalho LJC, Song F. A systematic review of outcomes assessed in randomized controlled trials of surgical interventions for carpal tunnel syndrome using the International Classification of Functioning, Disability and Health (ICF) as a reference tool. BMC Musculoskelet Disord 2006, 7, Article Number: 96.
13. World Health Organization. International Statistical Classification of Diseases and Health Related Problems ICD-10. 2nd ed. Geneva: World Health Organization; 2007.
14. Jette AM, Tao W, Haley SM. Blending activity and participation sub-domains of the ICF. Disabil Rehabil 2007; 29:1742–1750.
15. Jette AM, Haley SM, Kooyoomjian JT. Are the ICF activity and participation dimensions distinct? J Rehabil Med 2003; 35:145–149.
16. Haley SM, Coster WJ, Andres PL, Ludlow LH, Ni P, Bond TLY, Sinclair SJ, Jette AM. Activity outcome measurement for postacute care. Med Care 2004;42:I49–161.
17. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). A. Conceptual framework and item selection. Med Care 1992;30:473–483.
18. Üstün B, Chatterji S, Kostanjsek N. Comments from WHO for the journal of rehabilitation medicine special supplement on ICF core sets. J Rehabil Med 2004;44:7–8.
19. Cieza A, Ewert T, Üstün TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. J Rehabil Med 2004;44:9–11.
20. Stucki G, Üstün TB, Melvin J. Applying the ICF for the acute hospital and early post-acute rehabilitation facilities. Disabill Rehabil 2005;27:349–352.
21. Grill E, Ewert T, Chatterji S, Kostanjsek N, Stucki G. ICF Core Set development for the acute hospital and early post-acute rehabilitation facilities. Disabil Rehabil 2005;27:361–366.
22. Stucki G, Cieza A, Geyh S, Battistella L, Lloyd J, Symmons S, Kostanjsek N, Schouten J. ICF Core Sets for rheumatoid arthritis. J Rehabil Med 2004;44:87–93.
23. Dreinhöfer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner Ch, Kostanjsek N, Cieza A. ICF core sets for osteoarthritis. J Rehabil Med 2004;44:75–80.
24. Cieza A, Stucki G, Weigl M, Disler P, Jäckel W, van der Linden S, Kostanjsek N, de Bie R. ICF Core Sets for low back pain. J Rehabil Med 2004;44:69–74.
25. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, Kostanjsek N, Ring H, Stucki G. ICF Core Sets for stroke. J Rehabil Med 2004;44:135–141.
26. Stucki A, Daansen P, Fuessl M, Cieza A, Huber E, Atkinson R, Kostanjsek N, Stucki G, Ruof J. ICF Core Sets for obesity. J Rehabil Med 2004;44, 107–113.
27. Ewert T, Grill E, Bartholomeyczik S, Finger M, Mokrusch T, Kostanjsek N, Stucki G. ICF Core Set for patients with neurological conditions in the acute hospital. Disabil Rehabil 2005;27:367–374.
28. Stier-Jarmer M, Grill E, Ewert T, Bartholomeyczik S, Finger M, Mokrusch T, Kostanjsek N, Stucki G. ICF Core Set for patients with neurological conditions in early post-acute rehabilitation facilities. Disabill Rehabil 2005;27:389–396.
29. Masse LC, Heesch KC, Eason KE, Wilson M. Evaluating the properties of a stage-specific self-efficacy scale for physical activity using classical test theory, confirmatory factor analysis and item response modeling. Health Educ Res 2006;21:i33–i46.
30. Wilson M, Allen DD, Li JC. Improving measurement in health education and health behavior research using item response modeling: introducing item response modeling. Health Educ Res 2006;21:i4–i18.
31. Wu ML, Adams RJ, Wilson MR. ACER ConQuest: Generalised Item Response Modelling Software. Melbourne: ACER Press; 1998.
32. Carod-Artal FJ, Gonzalez-Gutierrez JL, Herrero JA, Horan T, De-Seijas EV. Functional recovery and instrumental activities of daily living: follow-up 1-year after treatment in a stroke unit. Brain Inj 2002;16:207–216.
33. Edwards B, O'Connell B. Internal consistency and validity of the Stroke Impact Scale 2.0 (SIS 2.0) and SIS-16 in an Australian sample. Qual Life Res 2003;12:1127–1135.
34. Orfila F, Ferrer M, Lamarca R, Tebe C, Domingo-Salvany A, Alonso J. Gender differences in health-related quality of life among the elderly: the role of objective functional capacity and chronic conditions. Soc Sci Med 2006;63:2367–2380.
35. Puhana MA, Behnkeb M, Devereauxc PJ, Montorid VM, Braendlie O, Frey M, Schunemann HJ. Measurement of agreement on health-related quality of life changes in response to respiratory rehabilitation by patients and physicians – a prospective study. Respir Med 2004;98:1195–1202.