ORIGINAL REPORT

AGREEMENT BETWEEN SINGLE RATERS AND TEAM RATING WHEN APPLYING THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH'S REHABILITATION SET

Malan ZHANG, MD, PhD[1], Yun ZHANG, BMS[2], Minghong SUI, MD, PhD[3], Liyin WANG, BSc[4], Ziling LIN, MMSc[5] Wei SHEN, MMSc[6], Jiani YU, MD, PhD[7] and Tiebin YAN, MD, PhD[8,9]

From the [1]Department of Exercise Rehabilitation, College of Exercise and Health, Guangzhou Sport University, Guangzhou; Departments of Rehabilitation: [2]The Fifth Hospital of Xiamen, Xiamen, [3]Shenzhen Nanshan People's Hospital, Shenzhen, [4]Clifford Hospital, Guangzhou, [5]The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, [6]Guangdong 999Brain Hospital, Guangzhou, [7]Guangdong Province Hospital of Chinese Medicine, Guangzhou, [8]Department of Rehabilitation Medicine, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou and [9]Guangdong Engineering Technology Research Center for Rehabilitation and Elderly Care, Guangzhou, China

Objective: To quantify the agreement between functional assessments by a single rater and a team using the Chinese version of the International Classification of Functioning, Disability and Health Rehabilitation Set in a clinical situation.
Design: Inter-rater, multi-centre agreement study.
Subjects: A total of 193 adult inpatients admitted to 5 rehabilitation centres at 5 hospitals in China
Methods: The Chinese version of the International Classification of Functioning, Disability and Health Rehabilitation Set was used by either a single rater or a team to assess 193 patients at 5 Chinese hospitals. Percentage of agreement and quadratic-weighted kappa coefficients were computed. Evaluation times were compared with paired *t*-tests.
Results: The mean team and individual evaluation times were not significantly different. The percentage of agreement ranged from 46.1% to 94.2% depending on the item, and the quadratic-weighted kappas ranged from 0.43 to 0.92. Eight categories (26.6%) showed a weighted kappa exceeding 0.4, 11 others (36.7%) exceeded 0.6, and another 11 (36.7%) produced kappas of more than 0.8.
Conclusion: Either a single rater or a team of raters can produce valid and consistent ratings when using the Chinese version of the International Classification of Functioning, Disability and Health Rehabilitation Set to assess patients in a rehabilitation department. The team rating approach is suitable for clinical application.

LAY ABSTRACT

A new team evaluation approach to implementing the rehabilitation measures of the World Health Organization's International Classification of Functioning, Disability and Health was tested by asking teams including a physician, a nurse, and a physiotherapist or an occupational therapist to evaluate 193 adult inpatients admitted to the rehabilitation departments of 5 hospitals in China. The teams' ratings were compared with those of single physicians and therapists. The agreement of the assessment results and the time taken by a single rater and a team were compared. There was moderate to high consistency in the ratings, and the mean times taken by the teams and the individual raters were not significantly different. In conclusion, team and single rating can both produce consistent assessments.

The International Classification of Functioning, Disability and Health (ICF) is officially endorsed by the World Health Organization (WHO) as the international standard for describing and measuring functioning and disability (1). It conceives of functioning as a dynamic interaction between a person's health, environmental factors and other personal factors (2). In its Global Disability Action Plan 2014–2021 (3) the WHO recommended the ICF as a framework for collecting comprehensive information on functioning and disability. The ICF includes nearly 1,500 categories covering diverse domains of functioning and a wide range of content and related concepts. This makes the ICF difficult to apply in clinical practice (4). To address this problem, ICF Core Sets, condensed from the whole set of ICF categories, have been developed to provide application-tailored shorter lists better related to specific health conditions and healthcare contexts (5–8).

Among the ICF Core Sets, one is a minimal, generic rehabilitation Core Set designed to address one of the most important challenges in health measurement: the comparability of data across studies and countries (9, 10). Although the ICF generic set has demonstrated application feasibility and good properties (11–13), it has only 7 categories, which limits its clinical application. An ICF Rehabilitation Set (ICF-RS) was therefore developed from the ICF generic set to reflect more key functional information universal among different patient populations (10). The ICF-RS includes 9

categories specifically for physical functioning and 21 categories for activities and participation. It can serve as a starting point for developing practical tools that compare a minimum set of data on disability across studies and countries (14).

The original ICF-RS had only a list of categories with some rather unclear definitions. Chinese rehabilitation professionals have been working with the ICF Research Branch to generate simple, intuitive descriptions of the categories in Chinese to promote their nationwide implementation (15). However, the detailed information professionals need to guide the application of the categories in clinical settings is still lacking. To alleviate this problem, an assessment standard has been developed for each category in the Chinese version of the ICF-RS. This provides detailed items easily applied in rehabilitation practice. The standards have demonstrated good validity and reliability (16, 17).

The clinical application of the Chinese assessment standards has, however, raised some problems. An evaluation using the standards involves interviews and clinical examination. It was difficult for a rater to complete the entire evaluation in a single setting, especially with a patient with complex complaints or poor language expression. In addition, the categories refer to 3 dimensions: body functioning, activity and participation. Some of the categories may be more relevant to and better rated by certain professionals.

To address these difficulties, a Delphi expert survey was conducted aiming to develop a new team evaluation approach rather than the default single rater approach to implementing the ICF-RS. It groups the 30 categories into 4 groups to be rated by a physician, a nurse, a physiotherapist, or an occupational therapist according to the content, with 6 categories assigned to the physician, 7 to the nurse, 9 to the physiotherapist, and 8 to the occupational therapist (18, 19). Using this team rating approach, each professional is responsible for evaluating the categories closest to their routine practice. Thus, assessments can be completed more easily without investing too much time. The assessors can easily generate the necessary information in their routine work.

The aim of this study was to quantify the agreement between functional assessments by a single rater and a team using the Chinese version of the ICF-RS in a clinical situation. The study compared the agreement between a single rater and a team of raters and also the time taken to complete the evaluation.

METHODS

Participants

This study applied a design in which each patient was evaluated separately by a single rater and a team of raters who were blinded to each other's collection of the data. Five rehabilitation departments from general or specialized hospitals participated. Four were from Guangdong Province, including 2 from Guangzhou and 1 each from Shenzhen and Zhuhai. The other participating hospital was in Fujian Province. The Chinese qualitative standards of the ICF-RS have been applied for years in those rehabilitation departments to assess patients' functioning. Many staff there have been formally trained to use the ICF-RS, so they are familiar with the assessment process.

The participants were recruited from among the inpatients admitted to the 5 rehabilitation departments between July and December 2019. The following inclusion criteria were applied: older than 18 years; at least 2 weeks since onset; conscious with a score ≥6 on the Chinese version of Hadkinson's abbreviated mental test (good cognitive ability); and continuously able to communicate verbally. Patients scheduled for discharge within 3 days, or those who were critically ill with unstable vital signs and any who were unwilling to cooperate with the whole evaluation process were excluded.

Participants meeting the inclusion criteria were recruited by quota sampling. Candidates were first classified in terms of neurological, musculoskeletal, cardiopulmonary or another condition. The proportions of the candidates selected at each rehabilitation department were then specified as nervous system dysfunction 50%, musculoskeletal system 25%, and 25% cardiopulmonary and others (e.g. tumour, geriatric) (17). The only exception was the Guangdong 999 Brain Hospital, which is a specialized hospital for neurological diseases.

Sample size

A sample size of at least 50 is considered acceptable for reliability studies (20). Allowing for 20% wastage, the target minimum sample size was therefore set as 63 in this study. The purpose, benefits, risks and confidentiality of the study were explained to each candidate. Any patient could withdraw from the study at will and their treatment would not be affected. The study protocols were approved by the ethics committees of the collaborating hospitals.

Raters

Five professionals were recruited at each collaborating rehabilitation department. In the single-rater approach, either a physician or therapist served as the single rater. The others formed a team of 4 raters with 1 physician, 1 nurse, 1 physiotherapist and 1 occupational therapist as suggested in the Delphi survey (19). All of the raters

had passed a 2-day unified and rigorous training course, which included theoretical study and clinical practice. After the training, each had independently passed the test for ICF-RS raters with an inpatient under the supervision of a trainer to make sure that they had mastered the basic concepts, the evaluation rules and matters needing attention with the Chinese assessment standards. A special group was set up to provide further assessment guidance and to answer any questions in the process of independent evaluation. All of the raters were registered members of their profession and had worked in a rehabilitation department for at least 3 years; hence they had the necessary knowledge and experience related to rehabilitation assessment.

Questionnaires

At the beginning of the rating process, the single raters completed a personal and disease information questionnaire describing each person rated, including their age, sex, marital status, education level, occupation, diagnosis and other information.

Hadkinson's abbreviated mental test (AMT) assesses basic cognitive functioning (21). It has 10 items covering directivity, memory, attention, computation ability, and recall. The questions are scored with 1 point for each correct answer and a total possible score of 10 points (22). The test was administered to each candidate and patients with an AMT score of 6 or more were included in the subsequent formal evaluation.

The Chinese assessment standard of the original ICF-RS had 9 categories for body function, 14 for activities, and 7 for participation (10). It is used across China to assess the key functions of patients from the acute to the chronic stage (16, 17). In each category, the severity of dysfunction receives 1 of 5 grades. No dysfunction is graded 0; mild dysfunction earns a 1; moderate dysfunction means grade 2; severe dysfunction means grade 3 and complete dysfunction is graded 4. There is also a grade 8 for failure to provide relevant information and a grade 9 used when a category is not applicable to a patient (23).

In contrast, the team evaluation version of the ICF-RS consists of 4 parts (19). In this study 6 categories were assigned to the physician, 7 to the nurse, 9 to the physical therapist (PT), and 8 to the occupational therapist (OT) (see Table I).

Data collection

The single raters evaluated all 30 items independently. The team raters arranged themselves to complete their parts separately whenever they had free time during working hours but within 3 days of patient's admission to the hospital. There were team meetings but the assessment results were not shared among the raters. To further demonstrate consistency of the 2 rating approaches, the evaluation time taken by each rater was also recorded (except at the specialized Guangdong 999 Brain Hospital). The reasons for failure to assess were recorded by the rater if any part of the whole rating was not completed within a patient's 3-day window. The case was excluded if more than 10% of the data on the 30 categories were missing (24).

Data analysis

The data were analysed with the help of version 25 of the SPSS software (IBM, Armonk, NY, USA) suite and version 12.0 of the Stata software package. Descriptive statistics were compiled summarizing the patients' demographic and disease-related information. Measurement data were expressed as mean ± standard deviation (SD). Paired t-tests were used for intra-group comparison of patients with normally-distributed data and paired Wilcoxon tests were used when the data were not normally distributed. A confidence level of $p \leq 0.05$ was considered to indicate statistical significance.

Paired t-tests were also applied to relate the evaluation times reported by the single raters and the teams. The agreement of each category's rating between a single rater and a team was another important result along with a weighted κ and a bias-corrected, bootstrapped 95% confidence interval (95% CI). Weighted kappa coefficients are commonly used to quantify the agreement between 2 raters on K-ordinal scales. A linear-weighted kappa coefficient relates the mean distance between 2 raters' classifications with respect to what would be expected by chance. That makes

Table I. Category assignments in team rating

Physician	Nurse	PT	OT
1. b134 Sleep functions	7. d770 Intimate relationships	14. b455 Exercise tolerance functions	23. d230 Carrying out daily routine
2. b152 Emotional functions	8. b620 Urination functions	15. b710 Mobility of joint functions	24. d640 Doing housework
3. b280 Sensation of pain	9. d570 Looking after one's health	16. b730 Muscle power functions	25. d660 Assisting others
4. b640 Sexual functions	10. d510 Washing oneself	17. d410 Changing basic body functions	26. d470 Using transportation
5. b130 Energy and drive functions	11. d520 Caring for body parts	18. d415 Maintaining a body position	27. d710 Basic interpersonal interactions
6. d240 Handling stress and other psychological demands	12. d530 Toileting	19. d420 Transferring oneself	28. d920 Recreation and leisure
	13. d550 Eating	20. d450 Walking	29. d540 Dressing
		21. d465 Moving around using equipment	30. d850 Remunerative employment
		22. d455 Moving around	

it suitable here, since statistical distributions are usually primarily described in terms of location and variability. A quadratic-weighted kappa coefficient provides changes in the centre of inertia about the agreement cells. Both coefficients were computed because they provide complementary information about the distribution of any disagreements (25, 26). Weighted kappas range from −1 to 1, where 1 indicates perfect agreement, 0 indicates no additional agreement beyond what is expected by chance alone, and a negative value indicates disagreement. A kappa value of 0.81–1.00 is viewed as almost perfect agreement, 0.61–0.80 as substantial, 0.41–0.60 as moderate, 0.21–0.40 as fair, and 0.00–0.20 as slight agreement (27).

RESULTS

Characteristics of the participants

A total of 217 patients were initially contacted. Six produced AMT scores <6, and 8 declined to participate, hence 203 patients were eventually recruited. Of those, 10 could not be included in the final statistical analysis because of incomplete data. Among them, 6 were excluded because a rater did not complete the assessment within 3 days of admission. Another 2 were discharged early, and the other 2 subjects dropped out for personal reasons. Hence, 193 patients were included in the final data analyses.

The participants had a mean age of 52.6 ± 16.7 years, with 69.4% younger than 60 years. Sixty percent ($n = 116$) were men. 73.6% said they had not attended university. Most of the patients were unemployed after their injury (112, 58%). 139 (72%) had a nervous system dysfunction, 43 (22.3%) had musculoskeletal problems, 11(5.7%) had cardiopulmonary system diseases. The patients' general characteristics are shown in Table II.

Characteristics of the raters

There were 5 raters at each of the 5 rehabilitation departments. Five of them worked as a single rater assessing all 30 categories. The other 20 participated as team raters. They had a mean age of 40.5 ± 6.76 years, with 60% older than 30 years. Sixteen (64%) were men. 48% had an intermediate title or better. They had a mean of 6.6 ± 4.9 years of experience working in a rehabilitation centre and most of them (80%) had 3–9 years of work experience. Almost all of the raters (23, 92%) had started learning about the ICF-RS within the previous year. The general characteristics of the raters are shown in Table III.

Table II. Characteristics of the study's population

|	Frequency	Percentage (%)	Mean±SD
Sex			
Male	116	60.1	
Female	77	39.9	
Age			52.6±16.7
20–40 years	49	25.4	
41–60 years	85	44.0	
≥61 years	59	30.6	
Education			
Primary school	33	17.1	
Junior middle school	45	23.3	
Senior middle school	64	33.2	
College and above	51	26.4	
Marital status			
Single	21	10.9	
Married	155	80.7	
Divorced or widowed	16	8.4	
Occupation			
Employed	81	42.0	
Unemployed	112	58.0	
Mean income monthly			
<¥3000	36	19.6	
¥3000–5000	46	23.8	
¥5000–10,000	64	33.2	
>¥10,000	47	24.4	
Rehabilitation group			
Nervous	139	72.0	
Musculoskeletal	43	22.3	
Cardiopulmonary & other	11	5.7	

SD: standard deviation.

Evaluation time

Of the 193 cases collected, 29 were from the Guangdong 999 Brain Hospital without time data. 54 patients' time data at the remaining 4 rehabilitation departments were invalid because a rater forgot to record the time, so a final total of 110 assessments with full evaluation time were analysed. The mean time taken to

Table III. Demographic characteristics and professional experience of the raters

Items	Frequency, n	Percentage (%)	Mean±SD
Sex			
Male	9	36	
Female	16	64	
Age			40.5±6.94
≤ 29 years	10	40	
30–50 years	15	60	
Profession			
Physician	7	28	
Nurse	5	20	
PT	6	24	
OT	6	24	
ST	1	4	
Professional title			
Primary	13	52	
Intermediate or above	12	48	
Years working in rehabilitation			6.6±4.9
3–9	20	80	
10–20	5	20	
ICF-RS experience			
< 1 year	23	92	
1–3 years	2	8	

PT: physiotherapist; OT: occupation therapist; ST: speech therapist; SD: standard deviation.
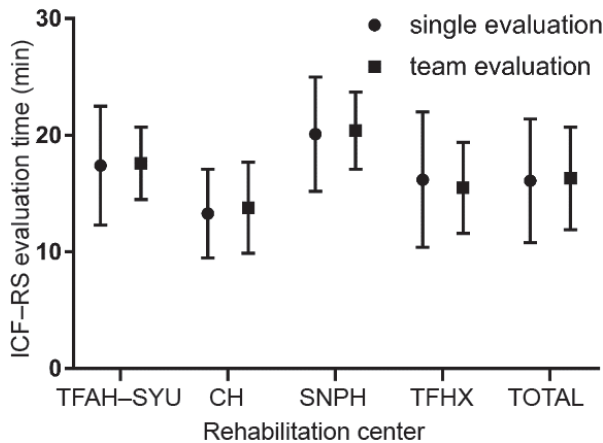
Fig. 1. Evaluation times at each department by a single rater and a team. ICF-RS: International Classification of Functioning, Disability and Health Rehabilitation Set; TFAH-SYN: The Fifth Affiliated Hospital, Sun Yat-sen University; CH: Clifford Hospital; SNPH: Shenzhen Nanshan People's Hospital; TFHX: The Fifth Hospital of Xiamen; TOTAL: mean evaluation time of a single rater or a team.

complete an evaluation was 16.1 ± 5.3 min for a single rater and almost the same (16.3 ± 4.4 min) for a team. A paired *t*-test confirmed that there was no significant difference between the 2 groups (t = –0.429, *p* = 0.67). Paired *t*-tests also showed that there was no significant difference between a single rater or a team at any of the individual rehabilitation centres (Fig. 1).

Weighted kappa coefficients

The observed agreement and the weighted kappas with bootstrapped 95% CIs are shown in Table IV. The percentage of agreement ranged from 46.1 to 94.2% depending on the category. A category's weighted kappa ranged from 0.43 to 0.92, with 8 categories (26.6%) showing a weighted kappa exceeding 0.4. Eleven (36.7%) had weighted kappas of more than 0.6, and for another 11 (36.7%) it was more than 0.8. The categories are ranked by kappa value from highest to lowest in Table IV. The category "d450 Walking" had the highest weighted kappa, while the category "d710 Basic interpersonal interactions" had the lowest.

DISCUSSION

The ICF is used as a reference model in the assessment of functioning, mostly in assessing specific health conditions in a rehabilitation context (28). Many researchers have sought to reduce the size or perceived complexity of the ICF by creating short lists of ICF domains for specific recording or measurement purposes (29). The team evaluation approach was developed through a Delphi study to facilitate the evaluation of the Chinese version of the ICF-RS (19). In that study the 30 categories were grouped into 4 parts to suit the

Table IV. Observed agreement and weighted kappas of ratings between single raters and teams

Team member	Category	Agreement		95% CI	
		Percentage	Kappa	Lower	Upper
PT	d450 Walking	94.2	0.92	0.88	0.96
PT	d420 Transferring oneself	80.8	0.91	0.88	0.95
PT	d465 Moving around using equipment	89.5	0.89	0.73	1.00
Nurse	d510 Washing oneself	65.8	0.85	0.80	0.90
OT	d540 Dressing	70.5	0.84	0.78	0.90
OT	d470 Using transportation	58.0	0.83	0.79	0.88
PT	D455 Moving around	64.1	0.83	0.72	0.89
OT	d640 Doing housework	70.7	0.82	0.76	0.89
PT	d410 Changing basic body functions	66.3	0.82	0.76	0.88
Nurse	d530 Toileting	63.2	0.81	0.75	0.87
PT	d415 Maintaining a body position	66.8	0.81	0.73	0.88
OT	d850 Remunerative employment	66.0	0.78	0.72	0.85
PT	b730 Muscle power functions	62.0	0.77	0.71	0.83
	d240 Handling stress and other psychological demands	56.5	0.74	0.66	0.82
PT	b710 Mobility of joint functions	63.8	0.74	0.66	0.82
Physician	b280 Sensation of pain	69.1	0.72	0.64	0.81
Physician	b640 Sexual functions	69.5	0.70	0.60	0.81
PT	b455 Exercise tolerance functions	60.6	0.69	0.60	0.78
Nurse	b620 Urination functions	72.0	0.69	0.56	0.81
OT	d230 Carrying out daily routine	48.2	0.66	0.57	0.75
Nurse	d550 Eating	66.3	0.64	0.50	0.78
Nurse	d520 Caring for body parts	61.6	0.62	0.51	0.72
Nurse	d570 Looking after one's health	49.0	0.60	0.49	0.70
OT	d920 Recreation and leisure	49.0	0.59	0.49	0.69
Physician	b130 Energy and drive functions	46.1	0.56	0.46	0.67
Physician	b134 Sleep functions	50.8	0.56	0.45	0.67
Nurse	d770 Intimate relationships	83.5	0.54	0.27	0.81
Physician	b152 Emotional functions	52.8	0.52	0.41	0.64
OT	d660 Assisting others	46.6	0.44	0.34	0.58
OT	d710 Basic interpersonal interactions	62.7	0.43	0.27	0.59

A kappa value of 0.81–1.00 is viewed as almost perfect agreement, 0.61–0.80 as substantial, 0.41–0.60 as moderate, 0.21–0.40 as fair, and 0.00–0.20 as slight agreement.
PT: physiotherapist; OT: occupation therapist; 95% CI: 95% confidence interval.

4 types of professionals and best bring to bear their diverse skills and experience. This study then aimed to evaluate the agreement between the results of a single rater and those of a team with specialist expertise. The results demonstrate that there were no significant differences in terms of evaluation time and that the ratings of a single rater and a team demonstrated moderate to high agreement.

Much has been published about the reliability of scales used among 2 or several single professional raters (30–33), but the reliability of team evaluation has been reported relatively rarely. Alvsåker reported observing good inter-rater reliability when the Early Functional Abilities scale was used by experts from 4 different professions independently. However, there was no team division of labour (34). Functional Independence Measure (FIM) is a team rating scale especially designed for use by a multidisciplinary team (35–36). A group led by Young has reported (37) that the mean total FIM rating was similar regardless of whether a team of healthcare professionals (generally consisting of a nurse, a physical therapist, an occupational therapist and a social worker) or a single non-clinician was the interviewer. The Catz-Itzkovich Spinal Cord Independence Measure (SCIM) can also be scored by a team of professionals (38), but in that case research has shown that assessment by a single nurse is not as accurate as by a multidisciplinary team (39).

In this study the single raters and the team both produced valid and consistent ratings similar to those reported in previous studies. Those studies with the FIM and SCIM tested the feasibility of single raters because they thought single scoring might be less burdensome and expensive, but the single raters selected were only nurses or non-clinicians. That may not suitable for the Chinese ICF-RS, as it includes categories that require more specialized skills such as "b710 Mobility of joint functions". Physicians and therapists in the clinic are the preferred raters. Team rating did not, however, increase the cost of the rating and promises better accuracy in functional assessment. Also, reduced personal assessment time in a busy workday and more targeted assessment may increase willingness to use the instrument, increase the raters' attention and promote more active intervention in clinical practice. In addition, consistent category ratings can be shared via a mobile application (18).

The weighted κ results of all of the categories were greater than 0.4, indicating moderate to high consistency between the single raters and the teams. ICF-RS assessments involve interviews and clinical examinations. The category "d450 Walking" had the highest weighted kappa. It is scored by looking at the patient's ability to walk 10 m on flat ground, wearing a brace or prosthetic limb or using a walking aid if necessary. It is graded according to the need for "supervision, prompting, or assistance". High consistency can be achieved because most professionals are familiar with the 10-m walking evaluation, and can give objective ratings through simple observation. The assessment of "d710 Basic interpersonal interactions" calls for the rater to make a judgment based on the subject's enthusiasm, appropriateness, language organization ability, expression ability, etc. in interpersonal communication. The patient's self-assessment and the opinions of family members may also be considered. The ratings range from excellent (0) to very poor (4) using Likert 5-level scoring. The ratings in that category demonstrated low consistency because the professionals, the patients and their families made different evaluations of the interviewees' interpersonal communication. To improve the situation the evaluation could be based entirely on the professional's rating after communicating with the interviewee.

The relatively low agreement in some categories is where subjective judgment is more important. And of course, the results of a particular evaluation depend to some extent on the degree of cooperation from the patient at that time as well as the rater's skill. The ICF is helpful in establishing a common language between different professionals and with patients, caregivers, administrators and health policy-makers (40). This study has shown that the team approach to ICF-RS assessment is feasible and gives results very consistent with those of a single rater.

Limitations

An obvious limitation is that all 5 rehabilitation departments involved in this study were in China. The findings need to be extended to other contexts. Also, this study was only conducted in the rehabilitation departments of third-class general or specialized hospitals in China. Further research will be needed to verify the suitability of team rating in community and rural rehabilitation centres. Furthermore, the quota sampling did not cover all of the patient population available during the study period. To do so would have disallowed random sampling. There was no formal debriefing. Ideally, semi-structured interviews should have been conducted with the raters involved in the team evaluations to better understand the acceptability of the team assessment approach. And it was also a limitation that the patients were not interviewed to collect their perspectives on single or team rating.

Team rating using the ICF-RS produces ratings the same as those of a single rater. Team assessment is thus potentially useful in the clinic. It can be an effective technique for producing consistent ICF-RS assessments.

ACKNOWLEDGEMENTS

REFERENCES

1. World Health Organization (WHO). The International Classification of Functioning, Disability and Health. Geneva: World Health Organization; 2001.
2. Vreeman DJ, Richoz C. Possibilities and implications of using the ICF and other vocabulary standards in electronic health records. Physiother Res Int 2013; 20: 210–219.
3. Gutenbrunner C, Negrini S, Kiekens C, Zampolini M, Nugraha B. The Global Disability Action Plan 2014–2021 of the World Health Organisation (WHO): a major step towards better health for all people with disabilities. Euro Phys and Rehab Med 2015; 51: 1–4.
4. Ustun 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 Suppl: 7–8.
5. Cieza A, Ewert T, Ustun TB, Stucki G. Development of ICF core sets for patients with chronic conditions. J Rehabil Med 2004; 44 Suppl: 9–11.
6. Kraus de Camargo O. International Classification of Functioning, Disability and Health core sets: moving forward. Dev Med Child Neurol 2018; 60: 857–858.
7. Tofani M, Mustari M, Tiozzo E, Dall'Oglio I, Morelli D, Gawronski O, et al. The development of the International Classification of Functioning, Disability and Health for child and youth (ICF-CY) core sets: a systematic review. Disabil Rehabil 2022 Oct 22: 1–10 [Online ahead of print].
8. Karlsson E, Gustafsson J. Validation of the International Classification of Functioning, Disability and Health (ICF) core sets from 2001 to 2019: a scoping review. Disabil Rehabil 2022; 44: 3736–3748.
9. Cieza A, Oberhauser C, Bickenbach J, Chatterji S, Stucki G. Towards a minimal generic set of domains of functioning and health. BMC Public Health 2014; 14: 218.
10. Prodinger B, Cieza A, Oberhauser C, Bickenbach J, Üstün TB, Chatterji S, et al. Toward the International Classification of Functioning, Disability and Health (ICF) rehabilitation set: a minimal generic set of domains for rehabilitation as a health strategy. Arch Phys Med Rehab 2016; 97: 875–884.
11. Li J, Prodinger B, Reinhardt JD, Stucki G. Towards the system-wide implementation of the International Classification of Functioning, Disability and Health in routine practice: lessons from a pilot study in China. J Rehabil Med 2016; 48: 502–507.
12. Reinhardt JD, Zhang X, Prodinger B, Ehrmann-Bostan C, Selb M, Stucki G, Li J. Towards the system-wide implementation of the International Classification of Functioning, Disability, and Health in routine clinical practice: Empirical findings of a pilot study from mainland China. J Rehabil Med 2016; 48: 515–521.
13. Ehrmann C, Prodinger B, Stucki G, Cai W, Zhang X, Liu S, et al. ICF generic set as new standard for the system wide assessment of functioning in China: a multicentre prospective study on metric properties and responsiveness applying item response theory. BMJ Open 2018; 8: e021696.
14. Senju Y, Mukaino M, Prodinger B, Stucki G. Development of a clinical tool for rating the body function categories of the ICF generic-30/rehabilitation set in Japanese rehabilitation practice and examination of its interrater reliability. BMC Med Res Methodol 2021; 21: 121.
15. Prodinger B, Reinhardt JD, Selb M, Stucki G, Yan T, Zhang X, et al. Towards system-wide implementation of the International Classification of Functioning, Disability and Health (ICF) in routine practice: developing simple, intuitive descriptions of ICF categories in the ICF generic and rehabilitation set. J Rehabil Med 2016; 48: 508–514.
16. Gao Y, Yan T, You L, Li K. Developing operational items for the International Classification of Functioning, Disability and Health rehabilitation set: the experience from China. Int J Rehabil Res 2018; 41: 20–27.
17. Gao Y, Yan T, You L, Li K, Zhang L, Zhang, M. Psychometric properties of the International Classification of Functioning, Disability and Health Rehabilitation Set: a Rasch analysis. Int J Rehabil Res 2021; 44: 144–151.
18. Zhang M, Yu J, Shen W, Zhang Y, Xiang Y, Zhang X, et al. A mobile app implementing the International Classification of Functioning, Disability and Health Rehabilitation Set. BMC Med Inform Decis Mak 2020; 20: 12–22.
19. Zhang M, Zhang Y, Xiang Y, Lin Z, Shen W, Wang Y, et al. A team approach to applying the International Classification of Functioning, Disability and Health Rehabilitation set in clinical evaluation. J Rehabil Med 2021; 53: jrm00147.
20. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: a practical guide. Cambridge, UK: Cambridge University Press; 2011.
21. Hadkinson H. Evaluation of a mental test score for assessment of mental impairment in the elderly. Age Ageing 1972; 2: 275–279.
22. Tanglakmankhong K, Hampstead BM, Ploutz-Snyder RJ, Potempa K. Does the Abbreviated Mental Test accurately predict cognitive impairment in Thai older adults? A retrospective study. Pac Rim Int J Nurs Res Thail 2021; 25: 23–33.
23. World Health Organization (WHO). How to use the ICF: a practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment. Geneva: World Health Organization; 2013.
24. Luo WY, Ni P, Chen L, Pan Q, Zhang H, Zhang Y. Development of the ICF-CY set for cardiac rehabilitation after pediatric congenital heart surgery. Frontiers in Pediatrics 2022; 10: 790431.
25. Vanbelle S. A new interpretation of the weighted kappa coefficients. Psychometrika 2016; 81: 399–410.
26. Johansson C, Åström S, Kauffeldt A, Carlström E. Daily life dialogue assessment in psychiatric care: Face validity and inter-rater reliability of a tool based on the International Classification of Functioning, Disability and Health. Arch Psychiatr Nurs 2013; 27: 306–311.
27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174.
28. Leonardi M, Lee H, Kostanjsek N, Fornari A, Raggi A,

Martinuzzi A, et al. 20 years of ICF International Classification of Functioning, Disability and Health: uses and applications around the world. Int J Environ Res Public Health 2022; 19: 11321.

29. Madden RH, Bundy A. The ICF has made a difference to functioning and disability measurement and statistics. Disabil Rehabil 2019; 41: 1450–1462.

30. Mukaino M, Prodinger B, Yamada S, Senju Y, Izumi SI, Sonoda S, et al. Supporting the clinical use of the ICF in Japan: development of the Japanese version of the simple, intuitive descriptions for the ICF generic-30 set, its operationalization through a rating reference guide, and inter-rater reliability study. BMC Health Serv Res 2020; 20: 66.

31. Liu S, Reinhardt JD, Zhang X, Ehrmann C, Cai W, Prodinger B. System-wide clinical assessment of functioning based on the International Classification of Functioning, Disability and Health in China: interrater reliability, convergent, known group, and predictive validity of the ICF Generic-6. Arch Phys Med Rehab 2019; 100: 1450–1457.

32. De Vrieze T, Frippiat J, Deltombe T, Gebruers N, Tjalm WAA, Nevelsteen I, et al. Cross-cultural validation of the French version of the Lymphedema Functioning, Disability and Health Questionnaire for Upper Limb Lymphedema (Lymph-ICF-UL). Disabil Rehab 2020; 28: 1–8.

33. Li K, Yan T, You L, Xie S, Li Y, Tang J, et al. The inter-rater reliability of the International Classification of Functioning, Disability and Health set for spinal cord injury nursing. Int J Rehab Res 2016; 39: 240–248.

34. Alvsåker K, Walther SM, Kleffelgård I, Mongs M, Drægebø RA, Keller A. Inter-rater reliability of the Early Functional Abilities scale. J Rehab Med 2011; 43: 892–899.

35. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the Functional Independence Measurement and its performance among rehabilitation inpatients. Arch Phys Med Rehab 1993; 74: 531–536.

36. Grey N, Kennedy P. The Functional Independence Measure: a comparative study of clinician and self-ratings. Paraplegia 1993; 31: 457–461.

37. Young Y, Fan MY, Hebel JR, Boult C. Concurrent validity of administering the functional independence measure (FIM) instrument by interview. Am J Phys Med Rehab 2009; 88: 766–770.

38. Catz A, Itzkovich M, Steinberg F, Philo O, Ring H, Ronen J, Spasser R, Gepstein R, Tamir A. The Catz–Itzkovich SCIM: a revised version of the Spinal Cord Independence measure. Disability Rehab 2001; 23: 263–268.

39. Catz A, Itzkovich M, Steinberg F, Philo O, Ring H, Ronen J, et al. Disability assessment by a single rater or a team: a comparative study with the Catz-Itzkovich Spinal Cord Independence measure. J Rehab Med 2002; 34: 226–230.

40. Leonardi M, Fheodoroff K. Goal Setting with ICF (International Classification of Functioning, Disability and Health) and Multidisciplinary Team Approach in Stroke Rehabilitation. In: Platz T,editor. Clinical Pathways in Stroke Rehabilitation: Evidence-based Clinical Practice Recommendations. Cham (CH): Springer; p.35–56.