Annals of Physical and Rehabilitation Medicine 64 (2021) 101533

Available online at ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com

Original article

Development and validation of a short version of the French Hand Function Sort questionnaire in vocational rehabilitation

Zineb Benhissen [a,*], Michel Konzelmann [a,b], Philippe Vuistiner [b], Bertrand Leger [b], François Luthi [a,b,c], Hervé Devilliers [d], Roger Hilfiker [e], Charles Benaim [b,c]

[a] Department of Musculoskeletal Rehabilitation, Clinique Romande de Réadaptation Suva, Sion, Switzerland
[b] Department of Medical Research, Clinique Romande de Réadaptation Suva, Sion, Switzerland
[c] Division of Physical Medicine and Rehabilitation, Lausanne University Hospital, Lausanne, Switzerland
[d] CIC INSERM 1432, Dijon University Hospital, Dijon, France
[e] School of Health Sciences, HES-SO Valais-Wallis, University of Applied Sciences and Arts Western Switzerland Valais, Sion, Switzerland

ARTICLE INFO

Article history:
Received 27 July 2020
Accepted 28 March 2021

ABSTRACT

Background: The Hand Function Sort (HFS) is a pictorial self-administered questionnaire with 62 items. It is a valid and reliable scale focused on the physical function of the upper limbs. It is used to predict the return to work.
Objectives: We aimed to develop and validate a short version of the French version of the HFS (HFS-F) to simplify its use in clinical practice.
Methods: We included patients with upper-limb chronic pain hospitalised for vocational rehabilitation from 2012 to 2019. Vocational rehabilitation aims to improve the autonomy of patients to regain their previous working capacity. The 62 items of the HFS-F were analysed in terms of patient and expert assessments, floor/ceiling effect, item-to-total correlation, principal component analysis, and Rasch analysis. A short HFS-F was developed. Thereafter, we assessed its internal consistency, test–retest reliability, criterion validity with the full-length HFS-F, construct validity with different scales (Disabilities of the Arm, Shoulder, and Hand [DASH]; Brief Pain Inventory [BPI]; Hospital Anxiety and Depression [HAD]), standard error of measurement (SEM), and minimal detectable change (MDC).
Results: Six experts were consulted, 34 patients were interviewed, and 629 questionnaires were analysed. Among the items, 25 were selected after the final round with the six experts. The internal consistency and test–retest reliability were excellent (Cronbach $\alpha = 0.95$, intraclass correlation coefficient = 0.92, 95% confidence interval [95% CI] 0.87 to 0.95). The correlation coefficient between scores of the short and full-length HFS-F was 0.841 (95% CI: 0.752 to 0.897, $P < 10–4$), and those between the short HFS-F score and the DASH, BPI, HAD-Anxiety, and HAD-Depression scores were −0.816 (95% CI: −0.714 to −0.881, $P < 10–4$), −0.529 (95% CI: −0.338 to −0.674, $P < 10–4$), −0.451 (95% CI: −0.244 to 0.614, $P = 0.0001$), and −0.360 (95% CI: −0.140 to −0.542, $P = 0.0018$), respectively. The SEM and MDC values were estimated at 6/100 and 17/100, respectively.
Conclusions: A short version of the HFS-F was developed and validated. We named this questionnaire the 25 HFS-F.

1. Introduction

The International Classification of Functioning (ICF) [1] is a conceptual framework that regroups several hierarchised classifications and assesses 3 key dimensions: body structure and function, activities, and participation. Patient-reported outcomes (PROs) evaluating the ICF dimensions are numerous and have been used for a long time. Examples include the assessment of the upper limb with the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire, which has been used for > 20 years [2].

Most of the newly validated clinical outcome (PRO) measures have focused on evaluating activities and participation [2]. Among them, the Hand Function Sort (HFS) is a pictorial self-administered questionnaire with 62 items that evaluate activities, participation, and self-efficacy in patients with musculoskeletal pathologies of

* Corresponding author.
E-mail address: zineb.benhissen@hcuge.ch (Z. Benhissen).

https://doi.org/10.1016/j.rehab.2021.101533

the upper limb. The responses on the HFS were developed from a review of the Dictionary of Occupational Titles physical demand taxonomy. Many items assess work tasks that place particular demand on the upper extremities [3].

A French version of the HFS (HFS-F) is available and has excellent psychometric properties similar to the original English version (Cronbach $\alpha = 0.98$; test–retest reliability = 0.921, 95% confidence interval [95% CI] 0.871 to 0.971) [4,5]. Recently, a Dutch version was developed with the same good psychometric properties [6]. The pictorial aspect makes it particularly easy to understand. The instrument is a useful and complementary tool for objective clinical evaluations [7]. It is also complementary to other PRO measures that more specifically assess disability, such as the DASH questionnaire [2]. In vocational rehabilitation, the HFS can be used to predict the return to work by comparing the self-efficacy assessed by the HFS to job demands and functional capacity evaluation (FCE) [7]. An example of an activity and participation assessed by the HFS is shown in Fig. 1, which illustrates handy-work participation and a position-holding activity.

Assessment in physical medicine and rehabilitation is a multidimensional, long, and complex procedure that requires several evaluation tools [8], including PROs, which can be time-consuming to complete. Because the development of new, shorter questionnaires is long and costly, several authors have proposed short versions of already validated tools, such as the QuickDASH [9], Brief Michigan Hand Questionnaire [10], and Modified Spinal Function Sort [11]. Some authors argue that the use of short assessments can promote the observance of handover delivery guidelines, increase the reliability of the results, and decrease the proportion of missing data. Most important, it could increase patient and caregiver satisfaction [10,12,13]. In a meta-analysis of response burden and questionnaire length, Rolstad [14] found a general association between response rate and questionnaire length (i.e., response rates were somewhat lower for longer questionnaires); however, the *P* value for testing homogeneity indicated heterogeneity in odds ratios among studies. This result suggests that the content of the questionnaires, rather than their length, is implicated in the response burden.

Fig. 1. An example of a Hand Function Sort item assessing handy-work participation and a position-holding activity. Item 31: Changing a light bulb overhead. https://www.mathesondevelopment.com.

In our clinic, patients complete 6 to 7 assessment questionnaires at admission, which is time-consuming for both patients and nurses. The HFS is the longest questionnaire (62 items), requiring 10 to 15 min to complete. To optimise our practice and motivate patients to correctly answer the HFS, we decided to shorten the questionnaire. We used the same approach as that previously used for the Modified Spinal Function Sort [11].

The aim of this study was to develop a short version of the HFS-F and to investigate its psychometric properties, especially validity and reliability. The study is reported according to COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) guidelines [15–17].

2. Methods

2.1. Study design, settings, and participants

The study was performed in a musculoskeletal and vocational rehabilitation centre. It was conducted in the same clinic and with the same patient profile as in the HFS-F validation study [5]. Data were collected from patients aged 18 to 65 years who were French speakers and hospitalised from January 1, 2012 to January 31, 2019. The data collection has been prospective since 2012.

Patients had various orthopaedic pathologies of the upper limb (shoulder, elbow, and hand) resulting from work, road, or domestic accidents. They had chronic pain for > 6 months. The exclusion criteria were:

- inability to read and understand the questionnaire in French;
- serious psychiatric disorders (psychosis, mood disorders, and post-traumatic stress disorder) and;
- brachial plexus- (or central nervous system)-associated pathologies.

At admission to our centre, all participants provided formal consent for the anonymous use of their data. The interviewed patients provided oral informed consent. The study was conducted according to the principles expressed in the Declaration of Helsinki 2008 (ethics committee No. CCVEM 001/10, Wallis, Switzerland).

2.2. Variables and measurement

From each included patient, we collected the following data: age, sex, educational level (≤ 9 or > 9 years), profession, injury localisation, HFS-F score, DASH score, Brief Pain Inventory (BPI) severity and interference item scores [18], and Hospital Anxiety and Depression (HAD) scale scores [19]. We chose 3 scales (DASH, BPI, and HAD) evaluating different dimensions based on the hypothesis that the score for the short version of the HFS will be correlated in a divergent way with scores for scales evaluating pain, depression, and anxiety and in a convergent way with the score for the scale evaluating disability.

The HFS items cover a large part of daily-living and professional activities involving the upper limbs (Fig. 1). They assess 7 areas: brief fingering, full body heavy, full body medium, full body light, hand–wrist–forearm stability, sustained fine coordination, and sustained force [3]. Questions 1–16 correspond to sedentary

activities, questions 17–34 light activities, questions 35–52 medium activities, and questions 53–62 heavy activities. Each item is rated on a 5-point Likert scale from 0 (“unable to do”) to 4 (“able to do”). An additional response possibility is defined by a question mark (“?”), which means “I don’t know” (score 0). The authors of the HFS identified 4 particular ranges over the entire range of the score (0–248): 100–136, 154–190, 200–228, and 238–248, corresponding to “sedentary”, “light”, “medium” and “heavy” activity levels, respectively. Three similar tasks are represented twice in the questionnaire for controlling the coherence of the answers. The questionnaire is considered “reliable” with fewer than 3 inconsistencies, fewer than 6 missing responses, and fewer than 6 “I don’t know” responses. For interpretation, percentile distributions for sex and working status (employed or unemployed) are available [3].

The DASH questionnaire is the most often used questionnaire for assessing function in upper-limb pathologies. It includes 30 items. After transformation, the scores range from 0 (“no disability”) to 100 (“most severe disability”) [2]. The DASH assesses disability and symptom severity.

For BPI, we used the following 2 dimensions: pain severity (4 items) and pain interference (7 items). The patients rate each of these dimensions on a scale from 0 to 10, with 10 indicating excruciating pain intensity and complete interference with life [18].

The HAD scale measures anxiety and depressive disorders. It has 14 items rated from 0 to 3. Seven questions relate to anxiety, and 7 others relate to the depressive dimension. For each dimension, the scores range from 0 (asymptomatic) to 21 (most depressed or anxious) [20].

2.3. Shortening the questionnaire

Various approaches have been used to reduce the number of items in questionnaires [21,22]. These approaches were based on clinical relevance, expert judgement of items, and statistical methods. We decided to use all of them, following a procedure similar to that used to develop the modified version of the Spinal Function Sort [11]. The choice of items to keep was based on rating each item according to its relevance to experts/patients and statistical performance according to the Stanton criteria [21]. This involves weighting the various qualitative and quantitative factors by adding a score assigned to each of the methods (Table 1).

2.4. Qualitative analyses

Patient interviews: We interviewed eligible inpatients between November 2016 and April 2017. We asked 34 patients to evaluate the overall frequency of the situation described by each item. A sample size > 30 was chosen to ensure that our estimations were unbiased and to allow for computation of normal-based Cis, even if precise guidelines are not available [23]. The question asked was, “In general, do you perform this task (never, rarely, sometimes, often)?” The Stanton score depended on how many times the answers “sometimes” and “often” were given by the participant: score “2” if > 24 times, “1” if 13–24 times, and “0” if < 13 times (Table 1).

Expert interviews: We consulted 6 clinical experts in the orthopaedic department who had at least 10 years of experience in upper-limb rehabilitation (2 physiatrists, 2 physiotherapists, and 2 occupational therapists), and asked them the following: “In order to reduce the number of items from 62 to 20–30, please indicate for each item if it should be kept (answer ‘yes’) or eliminated (answer ‘no’)”. The Stanton score for an item was the percentage of “yes” answers among the 6 experts, multiplied by 2 to obtain a score ranging from 0 to 2 (Table 1).

Table 1
Preselection of items of the short version of the French Hand Function Sort (HFS-F) questionnaire and scoring according to Stanton criteria.

Criteria	Cut-off	Scores
Patient interviews ($n = 34$)	> 24 times	2
Item frequency: sometimes and often	13–24 times	1
	< 13 times	0
Expert interviews ($n = 6$)	All experts	2
Expert opinion of item relevance, no. of “yes” answers	5/6	1.67
	4/6	1.33
	3/6	1
	2/6	0.67
	1/6	0.33
	None	0
Item-to-total correlation	0.6–0.9	2
	0.4–0.59	1
	< 0.4 or > 0.9	0
Principal component analysis–loading > 0.5	On one factor	2
	On two or no factors	0
Rasch analysis–infit and outfit	Both 0.5–1.5	2
	Otherwise	1
	Both < 0.5 or > 2	0
Floor/ceiling effect–proportion of max/min levels	$< 0.85\%$	2
	$\geq 0.85\%$	0
Total score		0–12

Quantitative analyses: We collected HFS-F questionnaires from eligible inpatients hospitalised from January 1, 2012 to April 30, 2017, and performed several calculations.

Item-to-total correlation: To estimate the representativeness of each item, we calculated the Pearson correlation coefficient between the item score and the overall assessment score [24]. The Stanton score of an item depended on its correlation (r) with the global HFS score: “0” for $r < 0.4$ and $r > 0.9$ (low representativity or redundancy with other items), “1” for $r = 0.4$ to 0.59, and “2” for $r = 0.6$ to 0.9 (good representativity) (Table 1).

Principal component analysis (PCA): This data analysis method consists of transforming variables linked to each other into new decorrelated variables (factors), allowing for reducing the number of variables. To determine the number of factors to be retained in the analysis, we used the Kaiser criterion: eigenvalues (proportion of the explained variance) > 1[25]. PCA was performed with Varimax rotation to force the loadings to be large or small for easier interpretation. Loadings > 0.5 on one factor (good and specific representation of one clinical dimension measured by the questionnaire) were considered good (Stanton score = 2), whereas loadings > 0.5 on 2 factors (representation of at least 2 dimensions) or on any factor (poor representation) were considered insufficient (Stanton score = 0) (Table 1).

Rasch analysis: The fit of the items to the Rasch model was examined with mean square infit and outfit statistical analysis [26]. We interpreted infit and outfit values as follows:

- both values between 0.5 and 1.5 indicate good fit (Stanton score = 2);
- both values < 0.5 or > 2 indicate insufficient fit (Stanton score = 0) and;
- other cases indicate low but still sufficient fit (Stanton score = 1) (Table 1).

Floor or ceiling effect: This effect was determined present if an item was scored at the lowest or highest global score of the HFS-F by $> 85\%$ of the patients [27]. A Stanton score of 0 was given to the item with a floor or ceiling effect, and 2 otherwise.

The 6 Stanton scores described above were summed to form a global Stanton score ranging from 0 to 12. Although this score in itself has no particular significance (no cut-off), it allowed us to rank the items according to their performance.

Finally, a last round with 6 experts was organised to finalise the short version of the HFS. The experts had access to the global Stanton score data and were asked to produce a version with 20 to 30 items, while keeping the same structure as the original HFS (levels of activity and domains) and with at least 2 pairs of redundant items (for classifying the answers as "reliable" or "unreliable"). Three experts had already participated in the first round, and 3 were new experts with exactly the same profile as the others.

Validation of the short version: Validation was performed from March 2018 to January 2019.

We studied the following psychometric properties of the short HFS-F with 57 eligible patients, as recommended by Terwee et al. [28]:

- Internal consistency: This property of the short HFS-F was determined with the Cronbach α, a general coefficient of homogeneity between items. Cronbach α values range from 0 (no internal consistency) to 1 (perfect internal consistency). A value of > 0.8 was expected [29];
- Test–retest reliability: The short HFS-F was administered to patients on days 4 and 7 of hospital stay. In the context of self-assessment, the intraclass correlation coefficient (ICC) (2.1) was calculated [30]. An ICC > 0.8 was expected [31]. We used Bland–Altman plots for graphic representations of reliability;
- Criterion validity: This assesses how well the PRO is correlated with another measure of the same nature, considered a "gold standard". As COSMIN recommends, we chose the original HFS-F as the gold standard: day 0 (HFS-F) and day 4 (short HFS-F). A Pearson correlation coefficient of > 0.8 was expected [29];
- Construct validity: We compared the short HFS-F to the DASH, BPI, and HAD-Anxiety and HAD-Depression questionnaires. DASH, BPI, and HAD questionnaires were administered to the patients on the first day of the hospital stay. We calculated the Pearson correlation coefficient. We expected to find any correlation or at least a weaker correlation with the HAD and BPI scores because they had different "constructs";
- Minimal detectable change (MDC) and standard error of measurement (SEM): The MDC is a distribution-change index that represents a "real change" that should not be due to measurement error or chance. It has been used to assess changes in many clinical scores in various patient populations. Its formula is as follows: $\text{MDC} = \text{SEM} \times 1.96 \times \sqrt{2}$, where the SEM agreement is a measure of how much scores are spread around a "true" score, based on the variance components [32].

Determination of thresholds: In the original HFS, the thresholds of "sedentary", "light", "medium", and "heavy" activity for the short HFS-F were determined by different percentiles. For example, the "sedentary" level of the original HFS corresponds to the 25th to 40th percentiles [3]. We calculated the scores of the short version corresponding to these percentiles in patients who participated in the validation step. We did the same for the other thresholds: 50th to 80th (light), 85th to 95th (medium), and 95th to 100th percentile (heavy).

The Number Cruncher Statistical System 12 (NCSS Atlanta 2018) [33] and Winsteps 3.8 were used for statistical analyses [34].

3. Results

3.1. Descriptive statistics

In total, 693 eligible patients completed the HFS-F questionnaire; 64 questionnaires were excluded because they were classified as "unreliable" (3 inconsistencies, 6 missing responses, or 6 "I don't know" responses). Therefore, 629 questionnaires were analysed (Fig. 2). The characteristics of patients are summarised in Table 2. Most of the patients were working-age men with musculoskeletal pathologies of the proximal upper limb (fractures, dislocation, or rotator cuff injury). Of the 629 patients, 34 were interviewed and 57 additional patients participated in the validation step (Fig. 2).

3.1.1. Shortening of the questionnaire

3.1.1.1.1. Qualitative analyses. According to the 34 interviewed patients, 12 items had a score of 2, 20 items a score of 1, and 30 items a score of 0. In expert interviews, 9 items had a score of 2, 6 items a score of 0, and 47 items a score of 0 to 2 (Table 3).

3.1.1.1.2. Quantitative analysis. In item-to-total correlation, 58 items had a rating of 2 and 4 items had a rating of 1. In PCA, 5 factors had an eigenvalue of > 1 and accounted for 72% of the total variation; 49 items had a factor loading of > 0.5 with only 1 factor (score 2 points), 4 items with 2 factors (score 0 points), and 9 items with 0 factors (score 0 points). In Rasch analysis, 51 items had a score of 2, 7 items a score of 1, and 4 items a score of 0. We did not find a floor or ceiling effect for any of the 62 items (all items scored 2).

3.1.1.1.3. Synthesis. The global Stanton score ranged from 5.67 to 11.67/12. We ranked the items according to this score and presented them to the 6 experts who made the following final selection: items 1, 5–8, 10–14, 23–25, 27, 29, 37, 41, 43, 45, 50–51, 55, 57, and 60–61 (Table 3). These 25 items were in the top 36 according to the global Stanton score. The 2 pairs of redundant items were items 1 and 14 and items 55 and 60. The short version of the HFS contains 25 items rated from 0 to 4 and "?" (total score 0–100).

4. Validation of the short version of the HFS

A total of 70 patients completed the short HFS-F, and 13 questionnaires were excluded because they were classified as "unreliable". On day 4, the mean (SD) score of the short HFS-F was 55 (21)/100. The mean scores of the HFS-F and DASH on day 0 were 145 (57)/248 and 51 (15)/100, respectively:

- the internal consistency of the short HFS-F was excellent (Cronbach α = 0.95);
- the test–retest reliability was excellent, with ICC = 0.92 (95% CI 0.87 to 0.95). The scores for reliability for most patients were well within the usual interval (Fig. 3);
- for criterion validity, the correlation between the short HFS-F and long HFS-F scores was good (0.841, 95% CI: 0.752 to 0.897, $P < 10^{-4}$);
- for construct validity, high scores on the DASH, BPI, HAD-Anxiety, and HAD-Depression corresponded to high disability, pain, anxiety, and depression, respectively. The correlation coefficients between the short HFS-F and DASH, BPI, HAD-Anxiety, and HAD-Depression scores were −0.816 (95% CI: −0.714 to −0.881, $P < 10–4$), −0.529 (95% CI: −0.338 to −0.674, $P < 10–4$), −0.451 (95% CI: −0.244 to 0.614, $P = 0.0001$), and −0.360 (95% CI: −0.140 to −0.542, $P = 0.0018$), respectively;
- the SEM and MDC values were estimated at 6/100 and 17/100, respectively.

Determination of thresholds: The preliminary results in the 57 patients who used the short HFS-F suggested the following score thresholds: sedentary, 44–52/100; light, 58–79/100; medium, 84–94/100; and heavy, 95–100/100.

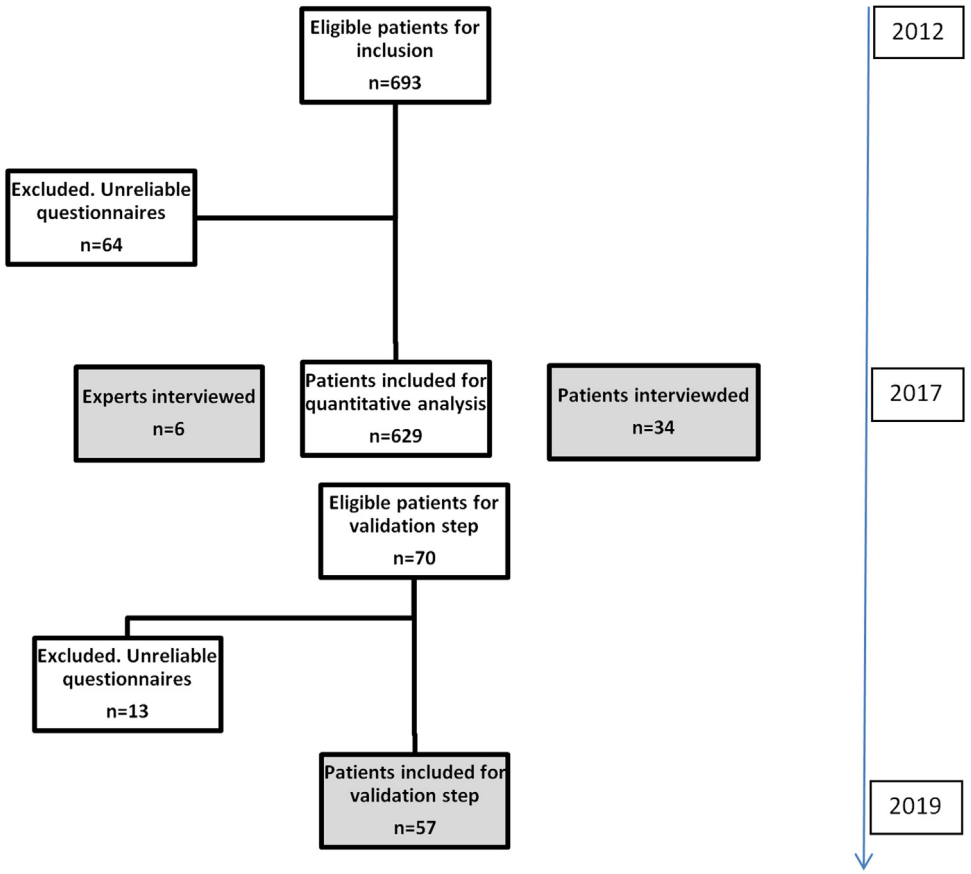

Fig. 2. Flowchart development and validation steps of the short version of the French version of the Hand Function Sort questionnaire.

Table 2
Characteristics of included patients.

Patients	Quantitative analyses ($n = 629$)	Validation step ($n = 57$)
Age, years, mean (SD)	43 (12)	45 (12)
Sex, male/female	540 (86%)/89 (14%)	51 (90%)/6 (10%)
Injury localization		
Proximal (shoulder, elbow)	402 (64%)	38 (68%)
Distal (wrist, hand)	227 (36%)	19 (32%)
Educational level (years)		
≥ 9	320 (51%)	27 (48%)
< 9	309 (49%)	30 (52%)
Profession		
Skilled or unskilled worker	496 (79%)	39 (69%)
Employee, business executive, other	133 (21%)	18 (31%)
Brief Pain Inventory score (/10), mean (SD)	4 (2)	5 (2)

5. Discussion

A short 25-item version of the HFS-F was developed and validated. The 25 items were chosen among the top 36 items according to the global Stanton score.

This shortened PRO instrument showed good psychometric properties, similar to the long version: 0.95 versus 0.98 for internal consistency, the same test–retest reliability, and good correlation (0.83 vs 0.77) of the score with the DASH score. As expected, the correlation between HAD and BPI scores was moderate but statistically significant, which agrees with the finding of previous studies that disability is a risk factor for pain, depression, and anxiety, and vice versa [35–38].

The reliability check is one of the components of the long-version HFS for detecting inconsistencies. Thus, we maintained this check by keeping 4 items representing 2 similar tasks (items 1 and 14, items 55 and 60).

The total score ranges from 0 to 100, and the instrument covers the same effort levels as the full version. The short HFS-F is a multidimensional PRO instrument that measures global activity, participation, and self-efficacy. The conceptual framework is identical to the original version, whose development was based on a review of the literature (US Dictionary of Occupational Titles) and on patient opinion [3]. The items encompass patient tasks involving the upper limbs, at home, at work, and in the community.

Table 3
Stanton criteria for statistical performance for 62 items of the HFS-F questionnaire.

Domain	Item	Patients	Expert	Item-to-total correlation	PCA	Rasch	Floor/ceiling effect	Total
Brief fingering	hfs_6*	2*	1.67*	2*	2*	2*	2*	11.67*
	hfs_8*	2*	1.67*	2*	2*	2*	2*	11.67*
	hfs_12*	2*	1.67*	2*	2*	2*	2*	11.67*
	hfs_10*	2*	1.33*	2*	2*	2*	2*	11.33*
	hfs_30	2	1	2	2	2	2	11
	hfs_13*	2*	0.33*	2*	2*	2*	2*	10.33*
	hfs_9	1	0.67	2	2	2	2	9.67
	hfs_2	1	1.33	2	2	1	2	9.33
	hfs_21	0	0.67	2	2	2	2	8.67
	hfs_19	1	0	2	2	1	2	8
	hfs_3	0	0	2	2	1	2	7
	hfs_4	0	0.67	1	2	0	2	5.67
Full body heavy	hfs_53	0	2	2	2	2	2	10
	hfs_61*	0*	2*	2*	2*	2*	2*	10*
	hfs_57*	0*	1.67*	2*	2*	2*	2*	9.67*
	hfs_55*	0*	1.33*	2*	2*	2*	2*	9.33*
	hfs_60*	0*	1.33*	2*	2*	2*	2*	9.33*
	hfs_56	0	0.67	2	2	2	2	8.67
	hfs_58	0	0.67	2	2	2	2	8.67
	hfs_59	0	0.33	2	2	2	2	8.33
	hfs_54	0	0	2	2	2	2	8
	hfs_62	0	0.67	1	2	2	2	7.67
Full body light	hfs_11*	2*	2*	2*	2*	2*	2*	12*
	hfs_7*	2*	1.67*	2*	2*	2*	2*	11.67*
	hfs_25*	2*	1*	2*	2*	2*	2*	11*
	hfs_34	1	1	2	2	2	2	10
	hfs_20	2	1.33	2	0	2	2	9.33
	hfs_23*	1*	2*	2*	0*	2*	2*	9*
	hfs_17	1	0.33	2	0	1	2	6.33
Full body medium	hfs_45*	1*	1*	2*	2*	2*	2*	10*
	hfs_50*	0*	2*	2*	2*	2*	2*	10*
	hfs_27*	0*	1.67*	2*	2*	2*	2*	9.67*
	hfs_46	1	0.67	2	2	2	2	9.67
	hfs_5*	1*	1.33*	2*	2*	1*	2*	9.33*
	hfs_35	1	1.33	2	2	1	2	9.33
	hfs_40	0	1.33	2	2	2	2	9.33
	hfs_48	0	1.33	2	2	2	2	9.33
	hfs_51*	0*	1.33*	2*	2*	2*	2*	9.33*
	hfs_47	0	1	2	2	2	2	9
	hfs_52	0	0.67	2	2	2	2	8.67
	hfs_31	1	1	2	2	0	2	8
	hfs_44	0	0	2	2	2	2	8
	hfs_49	0	0	2	2	2	2	8
	hfs_15	0	1.33	2	0	2	2	7.33
Hand, wrist, forearm stability	hfs_43*	1*	2*	2*	2*	2*	2*	11*
	hfs_37*	0*	1.33*	2*	2*	2*	2*	9.33*
	hfs_28	0	0.33	2	2	2	2	8.33
	hfs_38	0	0.33	2	2	2	2	8.33
	hfs_18	1	1	2	0	2	2	8
	hfs_33	1	1	2	0	2	2	8
	hfs_32	1	0.33	2	0	2	2	7.33
Sustained fine coordination	hfs_1*	2*	1.33*	1*	2*	1*	2*	9.33*
	hfs_14*	2*	0.33*	2*	2*	0*	2*	8.33*
	hfs_16	1	2	1	2	0	2	8
Sustained force	hfs_24*	1*	2*	2*	2*	2*	2*	11*
	hfs_29*	0*	1.67*	2*	2*	2*	2*	9.67*
	hfs_41*	1*	2*	2*	0*	2*	2*	9*
	hfs_22	1	1.33	2	0	2	2	8.33
	hfs_39	1	1	2	0	2	2	8
	hfs_36	0	0.67	2	0	2	2	6.67
	hfs_26	0	0.33	2	0	2	2	6.33
	hfs_42	0	0	2	0	2	2	6

For each of the 7 groups of items, items are ranked from highest to lowest rated. PCA: principal component analysis.
* Items selected by experts (final round).

Our study has several strengths, including the high number of questionnaires analysed and the combination of several methods for item selection. We included patient opinions, according to the 2009 US Food and Drug Administration recommendations [39]. We gave preference to clinically relevant items, selected by experts from 3 different professions, reflecting good representativeness of the expert panel. However, all experts came from the same rehabilitation centre, which could imply biased representativeness of the jury.

The calculated (purely statistically) MDC is not based on patient opinion and represents approximately a 17% derivation of the score. The minimal clinically important difference calculated for

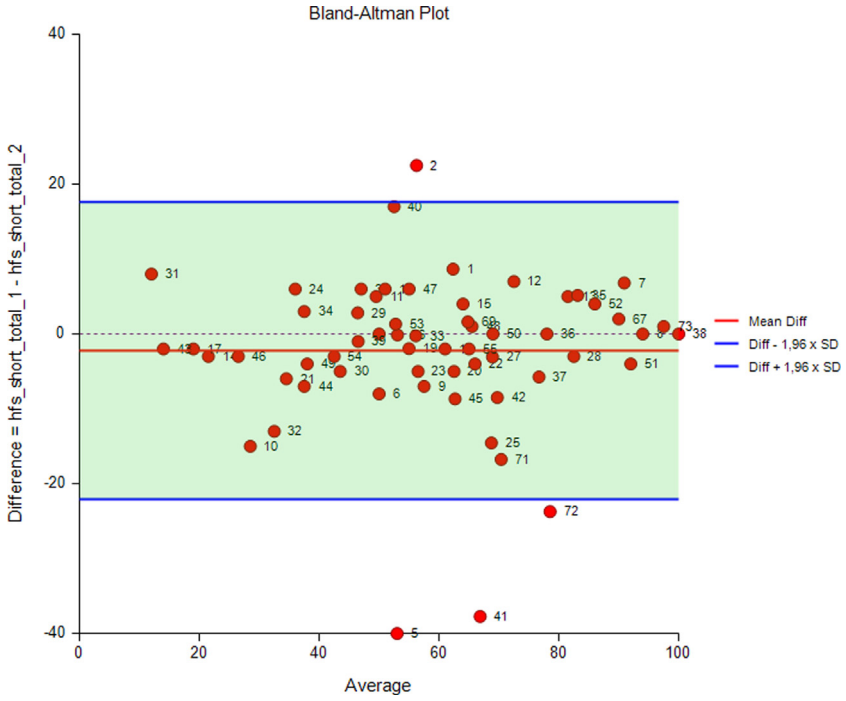

Fig. 3. Bland–Altman graphic representation of test–retest reliability of the short version of the French version of the Hand Function Sort questionnaire.

the long version represented about 10% derivation of the score [4]. Further studies are needed to assess the responsiveness of the short HFS-F [40].

6. Limitations

In this study, we were unable to demonstrate a lower proportion of unreliable results with the short versus long version of the HFS-F, possibly because the patients completed several other questionnaires at admission and exit from the clinic, which could have caused a lack of concentration and fatigue [41]. We did not measure the time taken to run the long versus short version of the HFS; however, 25 questions can be answered faster than 62 questions.

We assessed the correlation of the score for the short version of the HFS-F with those for 3 other scales (DASH, BPI, and HAD). Considering the complexity of the measured phenomenon of worker activity and participation, further studies are needed to reinforce the validity of the short HFS by assessing the correlation of its scores with those of other clinical scales and other categories of patients.

7. Conclusion

The short version of the HFS-F contains 25 items instead of the 62 items in its long version. We named this questionnaire the 25 HFS-F. This scale may be an interesting alternative to the long version, especially in the presence of a time constraint (e.g., in the context of a medical consultation) or when the patient needs to complete several questionnaires. Further studies are needed to compare the short and long versions in terms of patient satisfaction and missing data. Other studies including larger samples of patients are needed to establish the thresholds of effort according to the short HFS scores. In addition, this short version should be further tested on various populations.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Granlund M, Eriksson L, Ylvén R. Utility of international classification of functioning, disability and health's participation dimension in assigning ICF codes to items from extant rating instruments. J Rehabil Med 2004;36:130–7.
[2] Schoneveld K, Wittink H, Takken T. Clinimetric evaluation of measurement tools used in hand therapy to assess activity and participation. J Hand Ther 2009;22:221–36.
[3] Matheson LN, Kaskutas VK, Mada D. Development and construct validation of the Hand Function Sort. Journal of occupational rehabilitation 2001;11:75–86.
[4] Benhissen Z, Konzelmann M, Vuistiner P, Léger B, Luthi F, Benaim C. Determining the minimal clinically important difference of the hand function sort questionnaire in vocational rehabilitation. Ann Phys Rehabil Med 2019;62:155–60.
[5] Konzelmann M, Burrus C, Hilfiker R, Rivier G, Deriaz O, Luthi F. Cross-cultural adaptation, reliability, internal consistency and validation of the Hand Function Sort (HFS©) for French speaking patients with upper limb complaints. J Occup Rehabil 2015;25:18–24.
[6] Muskee A, Berduszek RJ, Dekker R, Reneman MF, van der Sluis CK. Cross-cultural adaptation and psychometric properties of the Dutch version of the Hand Function Sort in patients with complaints of hand and/or wrist. BMC Musculoskelet Disord 2019;20:1–11.
[7] Luthi F, Deriaz O, Vuistiner P, Burrus C, Hilfiker R. Predicting non return to work after orthopaedic trauma: the Wallis Occupational Rehabilitation RisK (WORRK) model. PloS one 2014;9:e94268.
[8] Béthoux F, Calmels P. Guide des outils de mesure et d'évaluation en médecine physique et de réadaptation. Frison-Roche; 2003.
[9] Budtz CR, Andersen JH, de Vos Andersen N-B, Christiansen DH. Responsiveness and minimal important change for the quick-DASH in patients with shoulder disorders. Health Qual Life Outcomes 2018;16:1–6.
[10] Waljee JF, Kim HM, Burns PB, Chung KC. Development of a brief, 12-item version of the Michigan Hand Questionnaire. Plast Reconstruct Surg 2011;128:208.
[11] Janssen S, Trippolini MA, Hilfiker R, Oesch P. Development of a modified version of the spinal function sort (M-SFS): a mixed method approach. J Occup Rehabil 2016;26:253–63.
[12] Ware JE. Improvements in short-form measures of health status: introduction to a series. J Clin Epidemiol 2008;61(1):1–5.
[13] Hawthorne G. Assessing utility where short measures are required: development of the short Assessment of Quality of Life-8 (AQoL-8) instrument. Value Health 2009;12:948–57.
[14] Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is shorter better?. A review and meta-analysis. Value Health 2011;14:1101–8.
[15] Mokkink LB, Prinsen CA, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 2016;20:105–13.
[16] Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:1–8.

[17] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[18] Cleeland C, Ryan K. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap 1994;23:129–38.
[19] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70.
[20] Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. J Psychosom Res 2002;52:69–77.
[21] Stanton JM, Sinar EF, Balzer WK, Smith PC. Issues and strategies for reducing the length of self-report scales. Personnel Psychol 2002;55:167–94.
[22] Coste J, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 1997;50:247–52.
[23] Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J-B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual life Outcomes 2014;12:1–10.
[24] Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012;24:69–71.
[25] Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960;20(1):141–51.
[26] DeMars CE. Infit and Outfit: Interpreting Statistical Significance and Magnitude of Misfit in Conjunction. J Appl Meas 2017;18:163–77.
[27] Streiner DL. Figuring out factors: the use and misuse of factor analysis. Can J Psychiatry 1994;39:135–40.
[28] Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
[29] Fermanian J. [Validation of assessment scales in physical medicine and rehabilitation: how are psychometric properties determined?] Ann Readapt Med Phys 2005;48:281–7. http://dx.doi.org/10.1016/j.annrmp.2005.04.004 [published Online First: Epub Date].
[30] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.
[31] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[32] de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033–9.
[33] https://www.ncss.com/sofware/ncss-documentation//#itemanalysis. Secondary.
[34] Linacre J. [WINSTEPS Rasch measurement. Computer software]. Chicago, IL: Winsteps. com; 2005.
[35] Noh J-W, Kwon YD, Park J, Oh I-H, Kim J. Relationship between physical disability and depression by gender: a panel regression model. PloS one 2016;11:e0166238.
[36] Marshall PW, Schabrun S, Knox MF. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related disability in people with chronic low back pain. PloS one 2017;12:e0180788.
[37] Costa J, Pinto-Gouveia J, Marôco J. Chronic pain experience on depression and physical disability: The importance of acceptance and mindfulness-based processes in a sample with rheumatoid arthritis. J Health Psychol 2019;24:153–65.
[38] Deschênes SS, Burns RJ, Schmitz N. Associations between depression, chronic physical health conditions, and disability in a community sample: a focus on the persistence of depression. J Affect Disord 2015;179:6–13.
[39] McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoeconomics Outcomes Res 2011;11:163–9.
[40] Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J Man Manipulative Ther 2012;20:160–6.
[41] Diehr P, Chen L, Patrick D, Feng Z, Yasui Y. Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials. Contemp Clin Trials 2005;26:45–58.