

Reliability and Validity of Two Versions of the Upper Extremity Functional Index

Bert M. Chesworth, BA, BScPT, MClScPT, PhD;^{*†‡} Clayton B. Hamilton, MSc;[†]
 David M. Walton, PT, BScPT, MSc, PhD;^{*} Melissa Benoit, PT, BScPT, MClScMT;^{*}
 Tracy A. Blake, PT, BKin, MScPT, MClScMT;^{*} Heather Bredy, PT, BScPT, MClScMT;^{*}
 Cameron Burns, PT, BSc, MClScMT, DPT;^{*} Lianne Chan, PT, BScPT, MClScMT;^{*}
 Elizabeth Frey, PT, BSc, BPHE, MSc, MScPT, MClScMT;^{*} Graham Gillies, PT, BKin,
 MScPT, MClScMT;^{*} Teresa Gravelle, PT, BScPT, MClScMT;^{*} Rick Ho, PT, BScPT,
 MClScMT;^{*} Robert Holmes, PT, BScPT, MClScMT;^{*} Roland L.J. Lavallée, PT, BA, BMRPT,
 MClScMT;^{*} Melanie MacKinnon, PT, HBScKIN, MScPT, MClScMT;^{*} Alishah (Jamal)
 Merchant, PT, BHSc, MScPT, MClScMT;^{*} Tammy Sherman PT, BKin, MScPT, MClScMT;^{*}
 Kelly Spears, PT, BPhEd, BHScPT, MClScMT;^{*} Darryl Yardley, PT, BKin, MScPT, MClScMT^{*}

ABSTRACT

Purpose: To examine the reliability, validity, and sensitivity to change of the 20-item version and the Rasch-refined 15-item version of the Upper Extremity Functional Index (UEFI-20 and UEFI-15, respectively) and to determine the impact of arm dominance on the positive minimal clinically important difference (pMCID). **Methods:** Adults with upper-extremity (UE) dysfunction completed the UEFI-20, Upper Extremity Functional Scale (UEFS), Pain Limitation Scale, and Pain Intensity Scale at their initial physiotherapy assessment (Time 1); 24–48 hours later (Time 2); and 3 weeks into treatment or at discharge, whichever came first (Time 3). Demographics, including working status, were obtained at Time 1. Global ratings of change (GRC) were provided by the treating physiotherapist and patient at Time 3. The UEFI-15 was calculated from relevant items in the UEFI-20. The intra-class correlation coefficient (ICC) and minimal detectable change (MDC) quantified test–retest reliability (Time 1–Time 2). Cross-sectional convergent validity was determined by the association (Pearson's *r*) between Time 1 measures of function and pain. Known-groups validity was evaluated with a one-way ANOVA across three levels of working status. Longitudinal validity was determined by the association (Pearson's *r*) between function and pain change scores (Time 1–Time 3). Receiver operating characteristic (ROC) curves estimated the pMCID using Time 1–Time 3 change scores and average patient/therapist GRC. **Results:** Reliability for the UEFI-20 and UEFI-15 was the same (ICC = 0.94 for both measures). MDC values were 9.4/80 for the UEFI-20 and 8.8/100 for the UEFI-15. Cross-sectional, known-groups, and longitudinal validity were confirmed for both UEFI measures. pMCID values were 8/80 for the UEFI-20 and 6.7/100 for the UEFI-15; pMCID was higher for people whose non-dominant arm was affected. **Conclusions:** Both UEFI measures show acceptable reliability and validity. Arm dominance affects pMCID. The UEFI-15 is recommended because it measures only one dimension: UE function.

Key Words: physiology; outcome assessment; reproducibility of results; ROC curve; upper limb.

From the: ^{*}School of Physical Therapy, [†]Graduate Program in Health and Rehabilitation Sciences, Faculty of Health Sciences, and [‡]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, Ont.

Correspondence to: Bert M. Chesworth, Faculty of Health Sciences, Western University, 1201 Western Rd., London, ON N6G 1H1; bcheswor@uwo.ca.

Contributors: All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared.

Clayton B. Hamilton is supported in part by the Joint Motion Program – A Canadian Institutes of Health Research (CIHR) Training Program in Musculoskeletal Health Research and Leadership, and also by a CIHR Doctoral Research Award.

Acknowledgements: The authors thank the patients who agreed to be study participants and Professor Paul Stratford for allowing us to use the original UEFI questionnaire.

For MB, TAB, HB, CB, LC, EF, GG, TG, RH, RLJL, MM, A(J)M, TS, KS, and DY, components of this research were completed in partial fulfillment of the requirements for the Western University, School of Physical Therapy MClScMT degree.

Physiotherapy Canada 2014; 66(3);243–253; doi:10.3138/ptc.2013-45

RÉSUMÉ

Objectif : Étudier la fiabilité, la validité et la sensibilité au changement des versions à 20 questions et à 15 questions raffinées par Rasch de l'Indice fonctionnel des membres supérieurs (IFMS-20 et IFMS-15, respectivement) et déterminer l'effet du bras dominant sur la différence minimale positive cliniquement importante (pDMCI). **Méthodes :** Les adultes ayant une dysfonction des membres supérieurs (MS) ont répondu au questionnaire IFMS-20, aux questionnaires de l'Échelle fonctionnelle des membres supérieurs (EFMS), de l'Échelle de limitation de la douleur et de l'Échelle de l'intensité de la douleur au cours de leurs premières évaluations en physiothérapie (moment 1); de 24 à 48 heures plus tard (moment 2) et 3 semaines après le début du traitement ou le congé, selon l'échéance la plus rapprochée (moment 3). On a réuni des données démographiques, y compris sur leur état de travailleur, au cours du moment 1. Le physiothérapeute traitant et le patient ont fourni des évolutions globales du changement (EGC) au moment 3. On a calculé le résultat du questionnaire IFMS-15 à partir de questions pertinentes contenues dans la version IFMS-20. Le coefficient de corrélation intracatégorie (CCI) et le changement détectable minimal (CDM) ont quantifié la fiabilité de test–retest (moment 1–moment 2). La validité convergente transversale a été déterminée par le lien (r de Pearson) entre les mesures de fonction et de douleur prises au moment 1. On a évalué la validité des groupes connus au moyen d'une analyse bidirectionnelle des écarts (ANOVA) entre trois niveaux d'état de fonctionnement. La validité longitudinale a été déterminée en fonction du lien (r de Pearson) entre les scores de changement de la fonction et de la douleur (moment 1–moment 3). Des courbes des caractéristiques opérationnelles du récepteur (COR) ont estimé la pDMCI à partir des scores de changement entre le moment 1 et le moment 3 et les EGC moyennes patient/thérapeute. **Résultats :** La fiabilité des questionnaires IFMS-20 et IFMS-15 a été la même (CCI = 0,94 pour les deux mesures). Les valeurs du CDM se sont établies à 9,4/80 pour le questionnaire IFMS-20 et à 8,8/100 pour le questionnaire IFMS-15. La validité transversale, de groupes connus et longitudinale a été confirmée pour les deux mesures de l'IFMS. Les valeurs de la pDMCI s'établissaient à 8/80 pour le questionnaire IFMS-20 et à 6,7/100 pour le questionnaire IFMS-15; la pDMCI était plus élevée chez les personnes dont le bras non dominant était atteint. **Conclusions :** Les deux mesures de l'IFMS montrent une fiabilité et une validité acceptables. Le bras dominant a un effet sur la pDMCI. On recommande le questionnaire IFMS-15 parce qu'il mesure une dimension seulement: la fonction des membres supérieurs.

Physiotherapists are commonly consulted by people with musculoskeletal upper-extremity (UE) problems. These disorders may affect any part of the upper limb or neck, causing numbness, tingling, swelling, pain, loss of coordination or strength, and UE dysfunction that affects work or recreational activities.¹ In 2006, a systematic review of worldwide rates of UE musculoskeletal disorders reported point prevalence as high as 53% (31.9% in Canada) and 12-month prevalence as high as 41% (19.8% in Canada).²

The Upper Extremity Functional Index (UEFI) is a patient-reported outcome measure (PROM) for quantifying UE function³ that has been used in several studies of people with musculoskeletal UE problems.^{4–8} Recently, we performed a Rasch analysis of the tool that informed its modification to a 15-item interval-level PROM (UEFI-15).⁹ Psychometric properties of the UEFI-15 have not been compared to the original version, nor has the positive minimal clinically important difference (pMCID) been determined for either measure.

Lang and colleagues have suggested that arm dominance may influence the pMCID.¹⁰ Briefly, they suggest that limited function may trouble patients more when it affects the dominant limb than when the non-dominant limb is involved, because the dominant arm is used for more skilled movements; therefore, a smaller improvement in ability to use the affected arm may be more significant for the patient. The literature on arm dominance and the pMCID is limited and inconclusive.^{10–12} The impact of limb dominance on the pMCID has not been determined for the original UEFI or for the UEFI-15.

The main objective of our study was to estimate the reliability, validity, sensitivity to change, and pMCID of the original UEFI and the UEFI-15; a secondary objective was to determine the impact of limb dominance on the pMCID.

METHODS

Design

We used a prospective longitudinal study design. Data were collected at the initial physiotherapy assessment (Time 1) for convergent and known-groups validity, 24–48 hours later (Time 2) for test–retest reliability, and at 3 weeks or discharge, whichever came first (Time 3), for longitudinal validity and calculation of the pMCID.

Participants

From October 2007 through March 2010, participants were recruited from 17 physiotherapy clinics across four Canadian provinces. Clients with a UE problem attending their first physiotherapy visit were invited to participate in the study. Inclusion criteria were (1) attendance at physiotherapy treatment for UE dysfunction deemed by the treating physiotherapist to be musculoskeletal in origin and (2) the ability to read and speak English fluently. The study received ethics approval from Western University's Health Sciences Research Ethics Board, and all participants gave informed consent.

Data collection

Time 1 descriptive data were captured by self-administered questionnaire and included age, gender, education, limb dominance, affected limb, symptom location, duration of problem, surgery status (did/did not have surgery), and working status (work not affected, work affected but continuing to work, off work because of problems). At all three time points, the UEFI,³ the Upper Extremity Functional Scale (UEFS),¹³ the Pain Limitation Scale (PLS), and the Pain Intensity Scale (PIS) were administered.¹⁴ Time 3 global ratings of change (GRC) were provided by the treating physiotherapist and the patient: one for function (GRC-function) and one for pain (GRC-pain).¹⁵

Outcome measures

20-item Upper Extremity Functional Index (UEFI-20)

The UEFI, a region-specific PROM of functional status for people with UE dysfunction of musculoskeletal origin,³ was originally developed by creating 105 items from a review of existing questionnaires, responses to the Patient Specific Functional Scale from patients with musculoskeletal UE problems,¹⁴ and input from physiotherapists with experience treating patients with UE dysfunction. The index was reduced to 32 items by combining similar items. A formal item analysis, including factor analysis, generated the 20-item questionnaire as a single domain quantifying UE functional status. Each item uses a 5-point adjectival response scale to rate difficulty in performing UE activities: 0 = extreme difficulty or unable to perform activity, 1 = quite a bit of difficulty, 2 = moderate difficulty, 3 = a little bit of difficulty, and 4 = no difficulty. Summing the items yields a total score from 0 (worst) to 80 (best) points. Before we began our study, the UEFI developer suggested minor changes to the original wording of two items to reflect feedback from clinicians (Stratford PW, 2007, personal communication). The item “lifting a bag of groceries above your head” was changed to “placing an object onto, or removing it from an overhead shelf,” since typically people do not lift groceries as originally described; the item “grooming your hair” was changed to “washing your hair or scalp” to take hair loss from medical or non-medical causes into account. Our study used this updated wording (hereafter UEFI-20).

15-item Upper Extremity Functional Index (UEFI-15)

We used the cohort in this study to produce the UEFI-15 through Rasch analysis of the UEFI-20.⁹ The UEFI-15 retains the UEFI-20 rating scale described above for all items except “doing up buttons,” which was modified to a scale from 0 to 3 points based on Rasch analysis. Its lowest anchor, extreme difficulty or unable to perform activity, has the same weight as the other items (=0), but the next two adjacent response options are weighted equally: quite a bit of difficulty (=1) and moderate difficulty (=1). The last two response options are weighted as follows: a little bit of difficulty (=2) and no difficulty (=3). Raw total scores (0 = worst state; 59 = best state), are transformed to generate an interval-level total score (0 = worst state; 100 = best state). In the current analysis, we extracted the UEFI-15 items from the UEFI-20 and used the interval-level total score. The UEFI-15 is reproduced in the Appendix.

Upper Extremity Functional Scale (UEFS)

The UEFS is an 8-item region-specific PROM of UE function.¹³ Each item is scored from 1 (no problem) to 10 (major problem), for a total score ranging from 8 (best state) to 80 (worst state). We chose the UEFS because it was the comparator used in the original UEFI study.³

Pain scales

Two pain scales originally developed for the Patient-Specific Functional Scale were also included: the Pain Limitation Scale (PLS) and the Pain Intensity Scale (PIS).¹⁴ The initial creation of these two pain scales was guided by the pain questions in the Short-Form 36 (SF-36);¹⁶ both pain scales were intended for use with any patient, regardless of their health condition. The PLS asks, “Over the past 24 hours, has the pain limited you from performing any of your normal daily activities?” Responses vary from 0 (activities have been severely limited) to 10 (activities have not been limited). The PIS asks, “Over the past 24 hours, how bad has your pain been?” Responses vary from 0 (no pain) to 10 (pain as bad as it can be). We chose these two pain scales because they were also comparators used in the original UEFI study.³

Global ratings of change

GRCs were determined using the retrospective transition rating scale method.¹⁵ At Time 3, the participant and physiotherapist were asked whether UE function was the same as, better than, or worse than at Time 1. A response of “no change” was assigned a value of 0; other responses were scored as follows: a tiny bit better/worse, almost the same = 1; a little bit better/worse = 2; somewhat better/worse = 3; moderately better/worse = 4; quite a bit better/worse = 5; a great deal better/worse = 6; a very great deal better/worse = 7. Improvement/deterioration ratings were assigned a positive or negative value, respectively. The result was two GRC-function 15-point scales that documented change in function, from -7 (a very great deal worse) to +7 (a very great deal better). Two similar GRC-pain scales were also created.

Analysis

We summarized sample characteristics by means and frequencies, and calculated all change scores so that positive values represent improvement.

Reliability

For reliability, we calculated a minimum required sample size of 35, based on the following a priori criteria: two test sessions, parameter estimation of an intra-class correlation coefficient (ICC) = 0.85, 95% CI width of 0.20, and 10% loss to follow-up.¹⁷ For test-retest reliability, we believed that clinical status should not change appreciably within 24–48 hours after the first physiotherapy assessment. To test this assumption, we compared the PROMs at these two time points using paired *t*-tests.¹⁸ We calculated the test-retest ICC_{2,1},¹⁸ the standard error of measurement (SEM),¹⁸ and the minimal detectable change at the 90% CI (MDC₉₀) with their 95% CIs.^{18–20} We also determined the 95% limits of agreement between the test-retest values.²¹ We used the Shapiro-Wilk test and visual inspection of the test-retest difference scores and their probability plots to determine whether or not they were normally distributed.^{22,23}

Table 1 Sample Characteristics at Initial Assessment ($n = 298$)

Characteristic	No. (%) of participants*
Mean (SD), min–max age, y ($n = 288$)	48.2 (14.3), 20–83
Female sex	152 (51)
Education ($n = 279$)	
Elementary	2 (1)
Some high school	28 (10)
High school	38 (14)
Some university or college	93 (33)
University	118 (42)
Affected limb ($n = 296$)	
Right	170 (57)
Left	109 (37)
Both	17 (6)
Dominant limb ($n = 296$)	
Right	257 (87)
Left	33 (11)
Ambidextrous	6 (2)
Work affected ($n = 292$)	
No	115 (39)
Yes, but continuing to work	139 (48)
Yes, off work because of problems	38 (13)
Had surgery	52 (18)
Duration of problem ($n = 295$)	
<1 wk	19 (6)
1–3 wk	62 (21)
4–8 wk	90 (31)
9–12wk	29 (10)
>12 wk	95 (32)
Symptom location ($n = 295$)	
Shoulder	154 (52)
Elbow	29 (10)
Wrist/hand	26 (9)
Multiple locations†	86 (29)

*Unless otherwise indicated.

† Multiple patterns of symptom location throughout the neck, shoulder or upper extremity.

Validity and sensitivity to change

Our validity analyses required a minimum sample size of 241 based on a Pearson's product–moment correlation coefficient (r) of 0.50,¹⁸ a 95% CI width of 0.20, and 10% loss to follow-up over time.²⁴ To determine known-groups validity, we compared the function PROMs across three subgroups of working status with a one-way ANOVA and Tukey's test post hoc.¹⁸ A significant difference indicated that a measure could differentiate across known groups of working status. We used Pearson's r to examine concurrent validity among the pain and function measures. A series of hypotheses were established a priori to evaluate evidence of validity: given the direction of the scales and constructs measured, we expected a strong negative correlation ($r \geq -0.70$) between the two UEFI versions and the UEFS, a moderate to strong positive correlation ($r \geq 0.50$) between the UEFI versions and the PLS, and a moderate negative correlation ($r \geq -0.40$) between the UEFI and the PIS.

For our analysis of longitudinal validity we assumed that clinical status would improve over 3 weeks of physio-

therapy treatment. We used Pearson's r to evaluate the relationship among change scores for the pain and function measures. We hypothesized that better UEFI change scores would be associated with better UEFS and pain change scores; moderate ($r \geq 0.40$) positive correlations were expected, since positive change reflects improvement. To analyze sensitivity to change, we used the approach for a heterogeneous sample of individuals, most of whom were expected to change by different amounts.²⁵ We used Spearman's rank correlation coefficient (r_s)¹⁸ to examine the relationship between the change score for each function PROM and the *average* of the patient's and physiotherapist's GRC-function. A moderate ($r \geq 0.40$) positive correlation was anticipated.

Positive minimal clinically important difference (pMCID)

We defined the pMCID as an average GRC-function of "somewhat better" or more ($\geq +3/+7$), which we felt reflected a minimally important change over 3 weeks of physiotherapy treatment. We used receiver operating characteristic (ROC) curve analyses to identify the change score that best discriminated between those who attained the pMCID and those who did not.²⁶ Pre-measurement chance of important change was set at 50%. We repeated the analyses with participants stratified into dominant and non-dominant affected arm groups, excluding both ambidextrous participants and those with bilateral symptoms, as neither could be assigned to a single affected or dominant limb group. To determine the impact of using the average GRC-function as our reference standard, we repeated the analyses with the pMCID defined only by the patient's GRC-function.

To evaluate the reliability of average GRC-function, we calculated the average GRC-pain the same way and used Cronbach's alpha²⁷ to quantify their internal consistency. For validity, we used Spearman's r_s to quantify their association and to evaluate the relationship between average GRC-function and the function PROMs' Time 3 and change score values.

Sensitivity analyses

Using multiple imputation,²⁸ we created five data sets with no missing UEFI, UEFS, or GRC-function values. We visually compared the study findings with the mean of five imputed reliability and validity coefficients and with the unstratified pMCID for each function PROM. All analyses were performed using SAS software, version 9.2 (Cary, NC, USA), except the ROC curve analyses, which were performed using MedCalc software, version 12.5.0 (Ostend, Belgium).

RESULTS

We recruited a total of 298 participants (see Table 1). After removal of 43 participants with missing Time 1 or 2 function PROM scores, 255 participants were available for reliability and cross-sectional validity analyses. An additional 25 participants with missing Time 1 or Time 3 function PROM scores were also removed, leaving

Table 2 Patient-reported Outcome Measures by Testing Occasion

Measure	Testing occasion; mean (SD)				<i>p</i> -value
	Time 1 initial assessment (<i>n</i> = 255)	Time 2* 24–48 hr later (<i>n</i> = 255)	Time 3 3 wk later or discharge† (<i>n</i> = 230)	Change score‡ Time 3 – Time 1 (<i>n</i> = 230)	
UEFI-20 (0–80)	51.2 (16.9)	52.7 (17.1)	61.4 (16.1)	10.7 (13.1)	< 0.001
UEFI-15 (0–100)	59.8 (15.3)	60.8 (15.3)	69.4 (16.1)	10.1 (11.7)	< 0.001
UEFS (8–80)	26.4 (15.1)	25.8 (15.2)	20.1 (13.2)	7.0 (11.2)	< 0.001
PLS (0–10) §	5.8 (2.8)	6.1 (2.7)	7.2 (2.6)	1.6 (3.1)	< 0.001
PIS (0–10) §	4.7 (2.4)	4.6 (2.4)	3.0 (2.3)	1.8 (2.5)	< 0.001

* Difference between Time 1 and Time 2: UEFI-20: $t_{254} = 4.13$, $p < 0.001$. UEFI-15: $t_{254} = 3.31$, $p = 0.001$. UEFS: $t_{254} = 1.65$, $p = 0.10$. Pain Limitation Scale: $t_{253} = -2.86$, $p = 0.005$. Pain Intensity Scale: $t_{253} = 1.57$, $p = 0.12$.

† 3 wk after initial assessment or discharge from physiotherapy, whichever came first.

‡ Positive change score indicates improvement for all measures.

§ Time 1 and 2 ($n = 254$); Time 3 ($n = 229$); Change ($n = 228$).

UEFI-20 = 20-item Upper Extremity Functional Index (higher scores indicate better function); UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (higher scores indicate better function); UEFS = Upper Extremity Functional Scale (lower scores indicate better function); PLS = Pain Limitation Scale (higher scores indicate less pain-related limitation); PIS = Pain Intensity Scale (lower scores indicate less pain intensity).

Table 3 Test–retest Reliability and Agreement Findings ($n = 255$)

Findings	Measure of upper extremity function		
	UEFI-20	UEFI-15	UEFS
Difference between test–retest scores, mean (SD)*	1.4 (5.5)	1.1 (5.2)	0.6 (6.0)
Reliability Parameters			
ICC (95% CI)	0.94 (0.93–0.96)	0.94 (0.92–0.95)	0.92 (0.90–0.94)
SEM (95% CI)	4.0 (3.7–4.4)	3.8 (3.5–4.1)	4.2 (3.9–4.6)
MDC ₉₀ (95% CI)	9.4 (8.6–10.2)	8.8 (8.1–9.5)	9.8 (9.1–10.7)
Agreement Parameter			
95% limits of agreement	–12.3, 9.4	–11.3, 9.1	–11.1, 12.3

* Positive values indicate improvement for all three measures. Shapiro-Wilk test results: UEFI-20, $W = 0.79$; UEFI-15, $W = 0.81$; UEFS, $W = 0.63$; all $p < 0.001$. UEFI-20 = 20-item Upper Extremity Functional Index (0–80, higher scores indicate better function); UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (0–100, higher scores indicate better function); UEFS = Upper Extremity Functional Scale (8–80, lower scores indicate better function); ICC = intra-class correlation coefficient; SEM = standard error of measurement; MDC₉₀ = minimal detectable change at the 90% CI.

a sample of 230 for longitudinal validity. Mean PROM values by time are shown in Table 2.

A total of 19 participants were missing a GRC-function, leaving 211 for the determination of the pMCID; of these, 14 were ambidextrous or presented with bilateral symptoms, leaving 197 participants for determination of the pMCID by arm dominance.

Reliability

All ICC_{2,1} values were >0.9 (see Table 3). Shapiro-Wilk tests rejected the null hypothesis of normally distributed test–retest difference scores. On visual inspection, difference scores for the UEFI-20 and UEFI-15 were symmetric, with many observations clustered near the mean (i.e., leptokurtic) and probability plots approximating a normal distribution. The UEFS difference scores were skewed left.

Validity and sensitivity to change

For known-groups validity (see Table 4), the UEFI-20, UEFI-15, and UEFS scores differed among work status

categories; all post hoc pairwise comparisons reached statistical significance. For cross-sectional and longitudinal validity (see Table 5), the absolute values of all function correlations were ≥ 0.6 ($p < 0.001$) and ≥ 0.4 ($p < 0.001$) for the relationship between the function and pain measures. For sensitivity to change, correlations between change scores and the average GRC-function are given in Table 5.

pMCID

For change thresholds up to three units larger than the pMCID, the UEFI-20 post-measure chance of improvement increased, whereas the UEFI-15 values were stable (see Table 6). The pMCID decreased by 1 and 0.4 units for the UEFI-20 and UEFI-15, respectively, when the patient GRC-function alone defined important change.

For both measures of function, the pMCID was higher for patients whose non-dominant arm was affected (see Table 7). Post-measure probabilities for important change were generally higher for those with an affected non-dominant arm.

Table 4 Patient-reported Outcome Measures by Working Status: Known-groups Validity ($n = 250$)

Measure	Working status; mean (SD)*		
	Work not affected	Work affected – continuing work	Off work because of problems
UEFI-20 (0–80)	58.3 (14.4)	49.9 (15.8)	35.1 (15.7)
UEFI-15 (0–100)	66.6 (14.8)	57.7 (13.3)	46.8 (13.6)
UEFS (8–80)	21.6 (13.4)	27.0 (14.1)	38.3 (16.4)

*ANOVA ($F_{2,247}$) for UEFI-20 (27.34), UEFI-15 (25.71), and UEFS (16.25), all $p < 0.001$. All post hoc pairwise comparisons $p < 0.05$.

UEFI-20 = 20-item Upper Extremity Functional Index (higher scores indicate better function); UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (higher scores indicate better function); UEFS = Upper Extremity Functional Scale (lower scores indicate better function).

Table 5 Association among Upper Extremity Measures of Function and Pain

Type of validity	Measures of upper extremity function; Pearson correlation coefficient (95% CI)*		
	UEFI-20	UEFI-15	UEFS
Cross-sectional†			
UEFI-20	–	0.95 (0.94–0.96)	–0.81 (–0.85 to –0.77)
UEFI-15	–	–	–0.79 (–0.83 to –0.74)
PLS‡	0.54 (0.45–0.62)	0.51 (0.41–0.60)	–0.44 (–0.53 to –0.33)
PIS‡	–0.44 (–0.54 to –0.34)	–0.42 (–0.52 to –0.32)	0.42 (0.31–0.52)
Longitudinal§			
UEFI-20	–	0.86 (0.83–0.89)	0.67 (0.59–0.73)
UEFI-15	–	–	0.57 (0.48–0.65)
PLS¶	0.51 (0.40–0.60)	0.46 (0.35–0.56)	0.39 (0.28–0.50)
PIS¶	0.50 (0.40–0.59)	0.45 (0.34–0.55)	0.46 (0.35–0.56)
Sensitivity to change**	0.57 (0.47–0.65)	0.58 (0.48–0.66)	0.43 (0.31–0.53)

*Unless otherwise indicated, all $p < 0.001$.

†Correlation between Time 1 values ($n = 255$).

‡ $n = 254$.

§Correlation between change scores ($n = 230$).

¶ $n = 228$.

**Spearman correlation coefficient between change scores and average Time 3 patient and physiotherapist global ratings of change in function ($n = 211$).

All $p < 0.001$.

UEFI-20 = 20-item Upper Extremity Functional Index (0–80, higher scores indicate better function); UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (0–100, higher scores indicate better function); UEFS = Upper Extremity Functional Scale (8–80, lower scores indicate better function); PLS = Pain Limitation Scale (0–10, higher scores indicate less pain-related limitation); PIS = Pain Intensity Scale (0–10, lower scores indicate less pain intensity).

For GRC reliability, Cronbach's alpha for the average GRC-function and GRC-pain was 0.97. For GRC validity, Spearman's s_r for their association was 0.93. Spearman's s_r for the association between the average GRC-function and Time 3 function PROMs were 0.40, 0.36, and –0.37 for the UEFI-20, UEFI-15, and UEFS, respectively. Table 5 shows correlations between average GRC-function and the function PROMs' change scores.

Sensitivity analyses

Mean imputed reliability coefficients were 1% higher than the values given in Table 3. Mean imputed correlation coefficients were $\pm 2\%$ of Table 5 values and up to 4% higher for sensitivity to change correlations. Mean imputed pMCIDs for the UEFI-20 and UEFI-15 were 7 and 6.3, respectively; their mean imputed post-measure chance of improvement was 2% less than Table 6 values.

Table 6 Results of Receiver Operating Characteristic (ROC) Curve Analysis, Adjusted for 50% Pre-measure Chance of Improvement ($n = 211$)

pMCID +/- change units (alternate threshold)	Sensitivity, %	Specificity, %	PLR	NLR	PPV, %	NPV, %	Post-measure chance of improvement,* %
UEFI-20							
-3 (5)	80	68	2.5	0.3	72	77	72
-2 (6)	78	78	3.5	0.3	78	78	78
-1 (7)	76	80	3.8	0.3	79	77	79
pMCID = 8†	72 (64–79)	84 (71–93)	4.5 (2.4–8.6)	0.3 (0.3–0.4)	82 (72–89)	75 (66–83)	82 (71–90)
+1 (9)	66	86	4.8	0.4	83	72	83
+2 (10)	60	92	7.5	0.4	88	70	88
+3 (11)	58	92	7.2	0.5	88	69	88
UEFI-15							
-3 (3.7)	83	56	1.9	0.3	65	77	65
-2 (4.7)	82	64	2.3	0.3	70	78	70
-1 (5.7)	76	72	2.7	0.3	73	75	73
pMCID = 6.7†	73 (66–80)	80 (66–90)	3.7 (2.1–6.4)	0.3 (0.2–0.4)	79 (69–86)	75 (66–83)	79 (71–86)
+1 (7.7)	70	80	3.5	0.4	78	73	78
+2 (8.7)	65	82	3.6	0.4	78	70	78
+3 (9.7)	58	84	3.7	0.5	79	67	79

*Chance that patients with change \geq the pMCID would report improvement of 'somewhat better' or more ($\geq +3/+7$).

† point estimate (95% CI).

pMCID = positive minimal clinically important difference defined by average patient and physiotherapist global rating of change in function of "somewhat better" or more ($\geq +3/+7$); PLR = positive likelihood ratio (no units); NLR = negative likelihood ratio (no units); PPR = positive predictive value; NPR = negative predictive value; UEFI-20 = 20-item Upper Extremity Functional Index (0–80, higher scores indicate better function). Area under ROC curve (95% CI): 0.83 (0.77–0.88), $p < 0.001$; UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (0–100, higher scores indicate better function). Area under ROC curve (95% CI): 0.79 (0.72–0.84), $p < 0.001$.

Table 7 Results of ROC Curve Analysis, Stratified by Dominance of Affected Limb, Adjusted for 50% Pre-measure Chance of Improvement ($n = 197$)

Affected limb dominance*	pMCID	AUC†	Sensitivity, %	Specificity, %	PLR	PPV, %	Post measure chance of improvement,‡ %
UEFI-20§							
Dominant	7	0.85 (0.78–0.91)	76 (66–84)	83 (66–93)	4.4 (2.1–9.2)	82 (70–90)	81 (68–90)
Non-dominant	10	0.74 (0.61–0.84)	62 (48–75)	90 (56–99)	6.2 (1.0–40.0)	86 (65–97)	86 (50–98)
UEFI-15§							
Dominant	5.7	0.78 (0.70–0.85)	77 (67–85)	74 (57–88)	3.0 (1.7–5.3)	75 (63–85)	75 (63–84)
Non-dominant	9.1	0.80 (0.68–0.89)	64 (50–77)	100 (69–100)	64.0¶ (1.6–77.0)	100 (83–100)	98¶ (62–99)

*Number (n) attaining a pMCID: Dominant affected, $n = 99/134$; Non-dominant affected, $n = 53/63$.

† All $p < 0.001$.

‡ Chance that patients with change \geq the pMCID would report improvement of 'somewhat better' or more ($\geq +3/+7$).

§ point estimate (95% CI).

¶ PLR cannot be calculated when specificity = 100%. To estimate this PLR and post-measure chance of improvement, specificity was set to 99% for point estimate and upper confidence limit.

pMCID = positive minimal clinically important difference defined by average patient and physiotherapist global rating of change in function of "somewhat better" or more ($\geq +3/+7$); AUC = area under receiver operator characteristic curve (no units); PLR = positive likelihood ratio (no units); PPV = positive predictive value; UEFI-20 = 20-item Upper Extremity Functional Index (0–80, higher scores indicate better function); UEFI-15 = Rasch-reduced 15-item Upper Extremity Functional Index (0–100, higher scores indicate better function).

Note: negative likelihood ratios/predictive values not included because they were unchanged from Table 6.

DISCUSSION

Our study found that the UEFI-20 and UEFI-15 demonstrated comparable reliability, validity, and sensitivity to change. The pMCID analyses were also comparable; both measures required more change to define important improvements in people with an affected non-dominant arm.

Reliability

Our test-retest reliability results are consistent with those published in the original UEFI study.³ We believe that our evaluation of the test-retest scores supports the two key assumptions for calculating the MDC—no systematic change between test occasions and consistency with a normal distribution²⁹—because the UEFI-20 test-

retest difference score (1.4) was similar to the mean change (1.8) of patients with a shoulder problem who rated their response to physiotherapy as unchanged.⁸ With respect to the distribution of the test–retest scores, we note that the Shapiro-Wilks test was intended to supplement rather than replace visual inspection of normal probability plots²²; the test statistic is a summary of the data, whereas the visual plot shows all the data.³⁰ Given that our normal probability plots were symmetric and that the requirement of a normal distribution may be less critical for test statistics derived from *F* tests,³⁰ we conclude that the test–retest scores sufficiently approximated a normal distribution.

Validity

Cross-sectional and longitudinal validity of both UEFI measures was supported by their correlation with the UEFS: all relationships were as strong as or stronger than our a priori hypotheses. While the directions of some of our scales were opposite to those reported by Stratford and colleagues,³ the absolute values of all our correlation CIs overlapped theirs, and our point estimates were either equal to or within 0.2 points of their reported values.

External validity of our UEFI findings is supported by the external validity of the UEFI values obtained in our study. Our mean UEFI-20 scores at Time 1 (51.2) reveal a sample of participants whose dysfunction was less severe than those of participants in Stratford and colleagues' study (43.2).³ However, our mean 3-week change score differs from theirs by only 0.1 UEFI-20 units, which may reflect the similarity in study design. Furthermore, our Time 1 UEFI-20 score compares favourably to the value (54.2) obtained from patients with a shoulder problem on their first visit to physiotherapy.⁸ Our mean Time 3 UEFI-20 score (61.4) compares favourably to the mean UEFI value obtained from patients attending physiotherapy for rotator cuff disease (65.2).⁵

pMCID

The UEFI-20 and the UEFI-15 pMCID estimates possess similar post-measure chances of improvement (see Table 6). While there may be differences between these measures when a different threshold value for the pMCID is used, it is important to keep in mind that the reason for these differences may be that the UEFI-20 is measuring other constructs in addition to UE function, such as the effects of sleeping on the affected shoulder—one of the items removed during the Rasch analysis.⁹

Our study found that pMCID was smaller for participants whose dominant arm was affected than for those whose non-dominant arm was affected. This finding is similar to pMCID findings reported for a pain visual analogue scale (VAS) among people with rotator cuff disease,¹¹ although in the same sample, arm dominance showed no impact on the pMCID for two shoulder PROMs.¹² Our finding is also similar to preliminary pMCID

findings reported for selected performance-based UE measures among people with hemiparesis following stroke.¹⁰ Different pMCID values for participants with non-dominant and dominant affected arms support the view that the pMCID value is context-specific.³¹ Independent verification of the current findings is recommended in the future.

To address concerns about the validity and reliability of GRC transition ratings,³² we chose a conservative strategy for calculating the pMCID by averaging patients' and physiotherapists' GRC. The fact that these average GRCs had a stronger association with the UEFI change scores than with the Time 3 UEFI values supports their validity. The internal consistency of the pain and function GRCs support the reliability of GRC-function, because Cronbach's alpha was in the range of expected values when two items have an inter-item correlation coefficient of the magnitude found in our study.²⁷

pMCID versus MDC

A reasonable question after reviewing our study results is whether the UEFI measures will identify meaningful change if the margin for variation in measurement (i.e., MDC_{90}) is larger than what patients perceive to be an important improvement (i.e., pMCID). In other words, can the UEFI-20 identify an important improvement of 8 units if 90% of patients whose UE function is truly unchanged over repeated testing exhibit random variation greater than this (i.e., 9.4 units)? Can the UEFI-15 identify an important improvement of 6.7 units if 90% of patients whose UE function is truly unchanged over repeated testing show random variation greater than this (i.e., 8.8 units)? Stratford and Riddle have provided insight into this apparent paradox by reminding us that the MDC_{90} is calculated from patients whose condition has not changed over repeated testing and therefore cannot provide accurate information about the chance that a patient who has attained a given change score has actually improved.²⁹ Table 6 shows alternative change-score thresholds for the pMCID; note that the higher threshold values do not negatively affect post-measure chances of improvement. Alternative pMCID values that are greater than or about equal to the MDC_{90} are 10 or 11 for the UEFI-20 and 8.7 or 9.7 for the UEFI-15. Clinicians may feel more comfortable using these alternative thresholds for the pMCID because, combined with the MDC_{90} value, they address the concern that a key threshold for defining important change lies within the bounds of random variation among patients who have truly shown no change.

Clinical implications

The implications of our study for clinicians are two-fold. First, although both UEFI versions have similar measurement characteristics of reliability, validity, and sensitivity to change, the UEFI-15 may be a sounder measure because of its unidimensionality. Sick has ex-

pressed the importance of this measurement quality in a manner that clinicians can readily appreciate: “Clear unidimensional variables help us to form conclusions and make decisions free of confounding interpretations.”^{33(p.23)} Because the UEFI-15 does not include specific items related to symptom intensity,⁹ an additional symptom-specific scale, such as the P4 pain instrument,³⁴ is recommended for a more complete picture of the dysfunction. Given that pain can also impact UE function, the takeaway message is that the UEFI-15 should always be used in conjunction with a separate measure of pain. The second implication is that clinicians can use our results when interpreting the importance of change scores in their patients. Considering the caveats outlined in the previous section, clinicians can confidently use these scores as a proxy for patient improvement.

Our study has several limitations. First, the findings are generalizable only to adults with UE problems of musculoskeletal origin and who present to physiotherapy clinics. Second, data were missing for the primary measures of interest: while sensitivity analyses confirmed our findings, they may have been influenced by the exclusion of patients with missing data. Third, missing values for our known-groups validity analysis may have resulted from the restricted response options for working status: participants who were not workers (i.e., retired, students, homemakers) may have been excluded from this analysis, which would limit the generalizability of these findings. Fourth, our mean UEFI and UEFS values suggest that severely affected people were underrepresented in our sample, which prevented us from evaluating the impact of baseline severity on the pMCID. Finally, the cohort used in our analysis was also used to generate the Rasch-reduced UEFI-15.⁹ While this allowed comparison of the two UEFI versions under similar circumstances, the fitting process that generated the UEFI-15 ensured that the accuracy of the fit to the Rasch model was as high as possible. The subsequent comparison of reliability and validity of the two UEFI versions with the same participant cohort may have produced an overly optimistic view of the performance of the newly developed UEFI-15. Therefore, a future reliability and validity study is warranted in which both UEFI versions are administered to an independent sample.

CONCLUSIONS

The UEFI-20 and the UEFI-15 have comparable reliability and validity. Overall, we believe that clinicians can confidently use the shorter UEFI-15 in routine clinical practice to evaluate functional change in people with UE musculoskeletal disorders. The UEFI-15 is recommended for measuring UE function because of its unidimensionality.

Future research should consider including participants with more severe functional limitations, perhaps drawn from populations outside of visiting physiotherapy clinics.

Given our finding that change-score thresholds vary by affected arm (dominant vs. non-dominant) and in light of the Rasch-transformed scores (available on the UEFI-15 in the Appendix), it is logical to consider that the magnitude of important change will vary by baseline score.

KEY MESSAGES

What is already known on this topic

The original Upper Extremity Functional Index (UEFI) is reliable and valid. A shortened version, the UEFI-15, has been shown to be unidimensional, measuring only UE function. The psychometric properties of these measures have not been compared, and the positive minimal clinically important difference (pMCID) has not been ascertained. The impact of arm dominance on pMCID has not been investigated.

What this study adds

The UEFI-15 has comparable measurement properties to the original questionnaire. Its pMCID is 6.7 / 100 units. An approach to reconciling differences between the minimal detectable change and the pMCID is provided. The UEFI-15 pMCID is larger (9.1) for patients with a non-dominant affected arm, and smaller (5.7) for those with a dominant affected arm. The UEFI-15 is recommended for use in clinical and research settings.

REFERENCES

1. Staal JB, de Bie RA, Hendriks EJ. Aetiology and management of work-related upper extremity disorders. *Best Pract Res Clin Rheumatol.* 2007;21(1):123–33. <http://dx.doi.org/10.1016/j.berh.2006.09.001>. Medline:17350548
2. Huisstede BM, Bierma-Zeinstra SM, Koes BW, et al. Incidence and prevalence of upper-extremity musculoskeletal disorders. A systematic appraisal of the literature. *BMC Musculoskelet Disord.* 2006;7(1):7. <http://dx.doi.org/10.1186/1471-2474-7-7>. Medline:16448572
3. Stratford PW, Binkley JM, Stratford DM. Development and initial validation of the upper extremity functional index. *Physiother Can.* 2001;53:259–67.
4. Razmjou H, Bean A, van Osnabrugge V, et al. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. *BMC Musculoskelet Disord.* 2006;7(1):26. <http://dx.doi.org/10.1186/1471-2474-7-26>. Medline:16533405
5. Razmjou H, Bean A, Macdermid JC, et al. Convergent validity of the constant-murley outcome measure in patients with rotator cuff disease. *Physiother Can.* 2008;60(1):72–9. <http://dx.doi.org/10.3138/physio/60/1/72>. Medline:20145743
6. Lehman LA, Sindhu BS, Shechtman O, et al. A comparison of the ability of two upper extremity assessments to measure change in function. *J Hand Ther.* 2010;23(1):31–9, quiz 40. <http://dx.doi.org/10.1016/j.jht.2009.09.006>. Medline:19944563
7. Kingston G, Tanner B, Gray MA. The functional impact of a traumatic hand injury on people who live in rural and remote locations. *Disabil Rehabil.* 2010;32(4):326–35. <http://dx.doi.org/10.3109/09638280903114410>. Medline:20055571
8. Hefford C, Abbott JH, Arnold R, et al. The patient-specific functional scale: validity, reliability, and responsiveness in patients with upper extremity musculoskeletal problems. *J Orthop Sports Phys Ther.* 2012;42(2):56–65. <http://dx.doi.org/10.2519/jospt.2012.3953>. Medline:22333510

9. Hamilton CB, Chesworth BM. A Rasch-validated version of the upper extremity functional index for interval-level measurement of upper extremity function. *Phys Ther.* 2013;93(11):1507–19. <http://dx.doi.org/10.2522/ptj.20130041>. Medline:23813086
10. Lang CE, Edwards DF, Birkenmeier RL, et al. Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Arch Phys Med Rehabil.* 2008;89(9):1693–700. <http://dx.doi.org/10.1016/j.apmr.2008.02.022>. Medline:18760153
11. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg.* 2009;18(6):927–32. <http://dx.doi.org/10.1016/j.jse.2009.03.021>. Medline:19535272
12. Tashjian RZ, Deloach J, Green A, et al. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am.* 2010;92(2):296–303. <http://dx.doi.org/10.2106/JBJS.H.01296>. Medline:20124055
13. Pransky G, Feuerstein M, Himmelstein J, et al. Measuring functional outcomes in work-related upper extremity disorders. Development and validation of the Upper Extremity Function Scale. *J Occup Environ Med.* 1997;39(12):1195–202. <http://dx.doi.org/10.1097/00043764-199712000-00014>. Medline:9429173
14. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther.* 1998;27(5):331–8. <http://dx.doi.org/10.2519/jospt.1998.27.5.331>. Medline:9580892
15. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407–15. [http://dx.doi.org/10.1016/0197-2456\(89\)90005-6](http://dx.doi.org/10.1016/0197-2456(89)90005-6). Medline:2691207
16. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30(6):473–83. <http://dx.doi.org/10.1097/00005650-199206000-00002>. Medline:1593914
17. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21(9):1331–5. <http://dx.doi.org/10.1002/sim.1108>. Medline:12111881
18. Portney LG, Watkins MP. *Foundations of clinical research: Applications to practice.* 2nd ed. Upper Saddle River (NJ): Prentice Hall Health; 2000.
19. Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther.* 1996;76(4):359–65, discussion 366–8. Medline:8606899
20. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther.* 1997;77(7):745–50. Medline:9225846
21. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* 2003;22(1):85–93. <http://dx.doi.org/10.1002/uog.122>. Medline:12858311
22. Shapiro SS, Wil MB. An analysis of variance test for normality (complete samples). *Biometrika.* 1965;52(3&4):591–611.
23. Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research.* 4th ed. Oxford: Blackwell Science; 2001.
24. Bonett DG, Wright TA. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika.* 2000;65(1):23–8. <http://dx.doi.org/10.1007/BF02294183>.
25. Stratford PW, Riddle DL. Assessing sensitivity to change: Choosing the appropriate change coefficient. *Health Qual Life Out.* 2005;3(24). <http://dx.doi.org/10.1186/1477-7525-3-23>
26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29–36. Medline:7063747
27. Carmines EG, Zeller RA. *Reliability and validity assessment.* Newbury Park (CA): Sage; 1979.
28. Institute for Digital Research and Education, UCLA. *Statistical computing seminars: multiple imputation in SAS, part 1* [Internet]. Los Angeles: The Institute [cited 2013 Mar 12]. Available from: http://www.ats.ucla.edu/stat/sas/seminars/missing_data/part1.htm.
29. Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test-based threshold change value for an outcome measure: resolving the conflict. *Phys Ther.* 2012;92(10):1338–47. <http://dx.doi.org/10.2522/ptj.20120002>. Medline:22767887
30. Royston P. A toolkit for testing for non-normality in complete and censored samples. *Statistician.* 1993;42(1):37–43. <http://dx.doi.org/10.2307/2348109>.
31. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14(2):109–14. <http://dx.doi.org/10.1097/00002281-200203000-00006>. Medline:11845014
32. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol.* 1997;50(8):869–79. [http://dx.doi.org/10.1016/S0895-4356\(97\)00097-8](http://dx.doi.org/10.1016/S0895-4356(97)00097-8). Medline:9291871
33. Sick J. Rasch measurement education. Part 5: assumptions and requirements of Rasch measurement. SHIKEN: JALT Testing & Evaluation SIG Newsletter. 14(2);2010 [cited 2013 Nov 5]:23–9. Available from: <http://jalt.org/test/PDF/Sick5.pdf>.
34. Spadoni GF, Stratford PW, Solomon PE, et al. Development and cross-validation of the P4: a self-report pain intensity measure. *Physiother Can.* 2003;55(1):32–8. <http://dx.doi.org/10.2310/6640.2003.35217>.

APPENDIX**Upper Extremity Functional Index-15**

Patient's name (or ID#) _____ Date _____

We are interested in knowing whether you are having any difficulty at all with the activities listed below because of your upper limb problem for which you are currently seeking attention. Please provide an answer for **each** activity.

Today, do you or would you have any difficulty at all with: (Circle one number on each line)

Activities	Extreme Difficulty / Unable to Do	Quite a Bit of Difficulty	Moderate Difficulty	A Little Bit of Difficulty	No Difficulty
1 Any of your usual work, housework, or school activities	0	1	2	3	4
2 Lifting a bag of groceries to waist level	0	1	2	3	4
3 Placing an object onto, or removing it from, an overhead shelf	0	1	2	3	4
4 Washing your hair or scalp	0	1	2	3	4
5 Pushing up on your hands (e.g., from bathtub or chair)	0	1	2	3	4
6 Preparing food (e.g., peeling, cutting)	0	1	2	3	4
7 Driving	0	1	2	3	4
8 Vacuuming, sweeping, or raking	0	1	2	3	4
9 Doing up buttons (Note: response numbering is correct)	0	1	1	2	3
10 Using tools or appliances	0	1	2	3	4
11 Opening doors	0	1	2	3	4
12 Cleaning	0	1	2	3	4
13 Laundering clothes (e.g., washing, ironing, folding)	0	1	2	3	4
14 Opening a jar	0	1	2	3	4
15 Carrying a small suitcase with your affected limb	0	1	2	3	4
Column Totals:					

Clinician: sum column totals for raw score: _____ / 59, then use table below for a final score _____ / 100.

Raw Score	Final Score	Raw Score	Final Score	Raw Score	Final Score	Raw Score	Final Score	Raw Score	Final Score	Raw Score	Final Score
0	0.0	10	33.1	20	43.5	30	51.5	40	59.4	50	69.9
1	8.5	11	34.4	21	44.4	31	52.3	41	60.2	51	71.3
2	14.4	12	35.6	22	45.2	32	53.0	42	61.1	52	72.9
3	18.6	13	36.7	23	46.0	33	53.8	43	62.0	53	74.8
4	21.7	14	37.8	24	46.9	34	54.6	44	63.0	54	76.8
5	24.3	15	38.9	25	47.6	35	55.3	45	64.0	55	79.3
6	26.5	16	39.9	26	48.4	36	56.1	46	65.0	56	82.3
7	28.4	17	40.8	27	49.2	37	56.9	47	66.1	57	86.2
8	30.1	18	41.8	28	50.0	38	57.7	48	67.3	58	91.8
9	31.7	19	42.7	29	50.7	39	58.5	49	68.5	59	100.0

Copyright of Physiotherapy Canada is the property of University of Toronto Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.