



Structural validity and construct validity of the Dutch-Flemish PROMIS[®] physical function-upper extremity version 2.0 item bank in Dutch patients with upper extremity injuries

Suus G. J. van Bruggen, Charlotte M. Lameijer & Caroline B. Terwee

To cite this article: Suus G. J. van Bruggen, Charlotte M. Lameijer & Caroline B. Terwee (2019): Structural validity and construct validity of the Dutch-Flemish PROMIS[®] physical function-upper extremity version 2.0 item bank in Dutch patients with upper extremity injuries, *Disability and Rehabilitation*, DOI: [10.1080/09638288.2019.1651908](https://doi.org/10.1080/09638288.2019.1651908)

To link to this article: <https://doi.org/10.1080/09638288.2019.1651908>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Aug 2019.



[Submit your article to this journal](#)



Article views: 1730



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

Structural validity and construct validity of the Dutch-Flemish PROMIS[®] physical function-upper extremity version 2.0 item bank in Dutch patients with upper extremity injuries

Suus G. J. van Bruggen^a, Charlotte M. Lameijer^a and Caroline B. Terwee^b

^aDepartment of Trauma Surgery, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; ^bDepartment of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

ABSTRACT

Introduction: Aim of this study was to validate the Dutch-Flemish Patient-Reported Outcomes Measurement Information System Physical Function – Upper Extremity version 2.0 item bank in patients with upper extremity injuries.

Materials and methods: Cross-sectional study. Structural validity was assessed using Confirmatory Factor Analysis examining unidimensionality. In addition, a bi-factor model was fitted. Internal consistency was assessed by Cronbach's alpha. Construct validity was examined by assessing correlations with legacy instruments Disability of Arm Shoulder and Hand, Patient Reported Wrist Evaluation and Michigan Hand Questionnaire subscale Activities in Daily Life.

Results: A total of 303 patients (144 female) with mean age of 50 years (standard deviation 18) were included. Confirmatory Factor Analysis showed Comparative Fit Index of 0.94, a Tucker Lewis Index of 0.93, a Root Mean Square Error of Approximation of 0.12 and a Standardized Root Mean Residual of 0.09. Factor loadings were all above 0.70. Bifactor analysis showed an omega-H of 0.79 and Explained Common Variance of 0.67. The correlations with the legacy instruments were as expected or higher than expected.

Conclusion: The Dutch-Flemish Patient-Reported Outcomes Measurement Information System Physical Function – Upper Extremity version 2.0 item bank measures a unidimensional trait and sufficient construct validity was found.

ARTICLE HISTORY

Received 9 April 2019

Revised 30 July 2019

Accepted 31 July 2019

KEYWORDS

Patient reported outcomes; upper extremity; validation studies



► IMPLICATIONS FOR REHABILITATION

- Completing Patient Reported Outcomes is time-consuming for patients and interpretability of outcomes is sometimes unclear due to some variation in psychometric properties.
- Computerized Adaptive Testing reduces the burden for patients by using an algorithm which decreases the amount of questions that need to be answered to 4 to 7 items.
- The Dutch-Flemish Patient-Reported Outcomes Measurement Information System Physical Function – Upper Extremity version 2.0 item bank measures a unidimensional trait and has sufficient structural validity, internal consistency and construct validity.
- After calibration of the Patient-Reported Outcomes Measurement Information System Physical Function – Upper Extremity version 2.0, the item bank is operable to use with Computerized Adaptive Testing.

Introduction

Every year many people suffer from musculoskeletal injuries. A recent study performed in the Netherlands showed a prevalence varying from 20 to 56% [1]. Especially, upper extremity injuries form a major problem for society. They have a lot of impact on physical health, but also on work, daily activities, participation, and health care costs. Huisstede *et al.* showed that the prevalence of upper extremity injuries vary between 1.6 and 53% [2]. Although exact prevalence numbers are lacking, probably due to different ways of defining injuries of the upper extremity and the differences in rehabilitation strategies [2].

In daily clinical practice, patients' rehabilitation outcome after suffered upper extremity injuries is objectified using clinical measurements, e.g., grip strength, range of motion, and radiological parameters. Other aspects such as pain, activity limitations, and participation restrictions are not being taken into account by these traditional methods [3]. Nowadays, Patient-Reported Outcomes (PROs) are used more frequently, to consider both patient- and expert opinion-based outcomes. Using PROs improve communication between patient and expert, which improves treatment and rehabilitation strategies [4,5].

CONTACT Charlotte M. Lameijer  c.lameijer@amsterdamumc.nl  Department of Trauma Surgery, Amsterdam UMC, De Boelelaan 1117, 1081 HV, Amsterdam, the Netherlands

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Most frequently used PROs for upper extremity injuries include the Disability of the Arm, Shoulder and Hand (DASH) [6] questionnaire, the Patient-Rated Wrist Evaluation (PRWE) questionnaire [7], and the Michigan Hand Outcomes Questionnaire (MHQ) [8]. In general, there is some variation in the psychometric properties, and the concepts measured are not always well defined [9–12]. In addition, completing PROs is time-consuming for patients, and interpretability of the outcomes is sometimes unclear [9–12].

Because of these challenges, the Patient-Reported Outcomes Measurement Information System (PROMIS) was initiated by six US research centers and the National Institutes of Health. PROMIS contains a series of item banks with items focused on function or pain, to measure outcomes from the patients' perspective across medical conditions [13]. The goal was to improve measurement quality and comparability of PROs and reduce patients' burden. Item banks were developed and validated for measuring specific symptoms and health status domains [13,14].

The PROMIS Psychological Function item bank has been developed, validated and calibrated in the English language [15,16]. To use the PROMIS item bank in the Netherlands and Belgium, the item bank was translated into the Dutch-Flemish language by the Dutch-Flemish PROMIS Group [17]. The first version of the Dutch-Flemish PROMIS Physical Function item bank (v1.2) contained 121 items [15–17]. This item bank has been validated in Dutch patients with chronic pain, patients undergoing physiotherapy, and patients with rheumatoid arthritis [18–20]. The PROMIS Group continued to improve the item bank, and a separate item bank for upper extremity injuries was developed, containing 46 items; the PROMIS[®] Physical Function – Upper Extremity (UE) v2.0 item bank [21,22]. Eventually, the item bank will be used as a Computerized Adaptive Test (CAT). The CAT system uses an algorithm that selects questions from the item bank based on the patients' response to the previous questions. When a predefined precision is reached, the system automatically stops asking questions. The benefit of a CAT is that the number of questions that need to be asked can be reduced to between 4 and 7 items [23].

The aim of this study was to validate the Dutch Flemish PROMIS UE v2.0 item bank. Structural validity and internal consistency must be sufficient before the item bank can be used as CAT. This ensures incorporating the patients' perspective on outcome following upper extremity injury and rehabilitation.

Materials and methods

Design

For this cross-sectional study, patients with upper extremity injuries were recruited at the outpatient clinic of a level 1 traumacenter in the Netherlands from May until July 2018. All patients were treated, conservatively or surgically. Online written informed consent was obtained.

Inclusion criteria were: patients having an injury of the upper extremity, age of 18 years or older at the moment of completing the questionnaire and sufficient understanding of the Dutch language in reading and writing. *Exclusion criteria* were: no sufficient knowledge of the Dutch language. Uncompleted questionnaires were not included in the analysis.

Dispensation for medical ethical approval was granted by the local Medical Ethics Committee [2018.259]. In addition, the study was performed in compliance with the principles outlined in the Declaration of Helsinki on ethical principles for medical research involving human subjects [24].

Methods of measurement

The international *PROMIS guidelines* for instrument development and validation were used, which serves as the scientific foundation for questionnaire development and validation [25]. The guideline follows the following structure: First, translation has to take place. Second, cognitive debriefing needs to be performed to ensure understanding and readability of the translated questions. Third, validation should be performed [25].

Procedure

Patients were requested to complete an online questionnaire, containing 4 questionnaires (108 items in total): the Dutch-Flemish PROMIS UE v2.0 item bank (containing 46 items), the DASH (containing 30 items), the PRWE (containing 15 items) and the MHQ subscale Activities in Daily Life (MHQ-ADL, containing 17 items). The complete questionnaire was built in "Survalyzer[®]", which is an online survey software program.

Following informed consent, the included patients completed the questionnaires on an electronic device (e.g., iPad, smartphone, computer), during their visit to the outpatient clinic or at home through an email with the link to the web-based questionnaire. Patients who were unable to use an electronic device could complete a paper version of the questionnaire. If applicable, a reminder by email was sent following the initial invitation. The estimated time to complete the online questionnaire was calculated to be about 13 min, based on an average expected response time per item of 7 s (108×7 s) [26].

Measurements

The online questionnaire also contained questions addressing demographic and clinical characteristics. Demographic characteristics asked were age, gender, country of birth, expected duration of rehabilitation, and educational level. In addition, if present a pending compensation claim was recorded. Clinical characteristics that were asked were dominance, trauma intensity (mono-trauma or poly-trauma), disease duration at the moment of inclusion and the presence of other pain complaints.

The questionnaire of specific interest was the Dutch-Flemish PROMIS UE v2.0 item bank. The first version (v1.2) of the Physical Function item bank measures general physical function, of both upper and lower extremity and health condition together [16]. A specific item bank focusing on upper extremity limitations was developed. Forty-two relevant items of the v1.2 item bank were reused, and 4 new upper extremity functioning questions were developed and translated into Dutch-Flemish. Following translation, to evaluate the comprehensibility and relevance of the items, cognitive debriefing was performed. Cognitive interviews were

Table 1. PROMIS physical function - upper extremity v2.0 item bank.

PROMIS upper extremity	46 items <ul style="list-style-type: none"> • 42 items specifically addressed to UE function. • 4 recently added items addressing UE function.
	Timeframe: none, but current status is inferred.
	Two different 5-point Likert response scales: <ul style="list-style-type: none"> • Unable to do/With much difficulty/With some difficulty/With a little difficulty/Without any difficulty • Cannot do/Quite a lot/Somewhat/Very little/Not at all
	Higher scores indicate better function: <ul style="list-style-type: none"> 0 (not able to do anything) to 100 (no disability at all)

conducted for all 45 items with at least 5 native Dutch and at least 5 native Flemish patients and people from the general population (submitted for publication).

The current version of the PROMIS UE v2.0 item bank contains 46 items focusing on limitations in activities that require the upper extremity, with two different 5-point Likert response scales (Table 1). Scores of all PROMIS measures are expressed as T-scores, where a score of 50 represents the average of the (US) general population, with a standard deviation (SD) of 10. Higher scores indicate better function.

As mentioned earlier, 3 disease specific legacy instruments were used in this study; DASH, PRWE and MHQ-ADL Table 2.

The *DASH questionnaire* contains 30 items, specifically addressed to disabilities and symptoms in injuries of the upper extremity (Table 2) [6]. The questions use 5-point Likert response scales, ranging from no problems with functioning at all (1 point) until completely unable to function (5 points). The total score ranges from 0 (no disability) to 100 (most severe disability). The timeframe for the items is “during the past week.” Both the primary English DASH questionnaire and the official Dutch translation have acceptable psychometric properties [11,12,27–31].

The *Patient Rated Wrist Evaluation (PRWE) questionnaire* contains 15 items, focused on wrist injury (Table 2) [7]. Five items are specifically addressed to pain and 10 items to function, divided into specific activities and usual activities. The timeframe for the items is “during the past week.” The items assessing pain are rated from no pain (0) to unbearable pain (10), and the function scale is rated from no disability (0) to most disability (10). Higher scores imply worse outcome. Both the primary English PRWE

questionnaire and the Dutch translation have acceptable psychometric properties [3,7,32].

The *Michigan Hand Outcomes Questionnaire subgroup Activities in Daily Life (MHQ-ADL)*. The complete MHQ contains 57 items [8], divided into 6 subgroups; activities of daily living (ADL), overall hand function, pain, work performance, esthetics and patient satisfaction with their hand function. For this study, the subscale MHQ-ADL was chosen, because the ADL subscale is expected to measure the same construct as the Dutch-Flemish PROMIS UE v2.0 item bank (Table 2). The MHQ-ADL contains 17 items, focused on hand or wrist injury [8]. Five items for the right hand, 5 items for the left hand and 7 items for both hands, all addressed to activities of daily living. Scores of the injured side were used for calculations. A 5-point Likert response scale is used: Not difficult at all/A little difficult/Somewhat difficult/Moderately difficult/Very difficult. The timeframe for the items is “during the past week.” The total score per scale is converted to a score ranging from 0 to 100. Higher scores indicate less disability. Both the primary English MHQ and the official Dutch translation have good psychometric properties [8,33–38].

Analysis

Sample size

There are varying views and guidelines to determine the sample size for validating questionnaires [39]. The required sample size for Confirmatory Factor Analysis (CFA) was estimated as at least 300 participants, based on recommendations by Comrey and Lee [40].

Structural validity

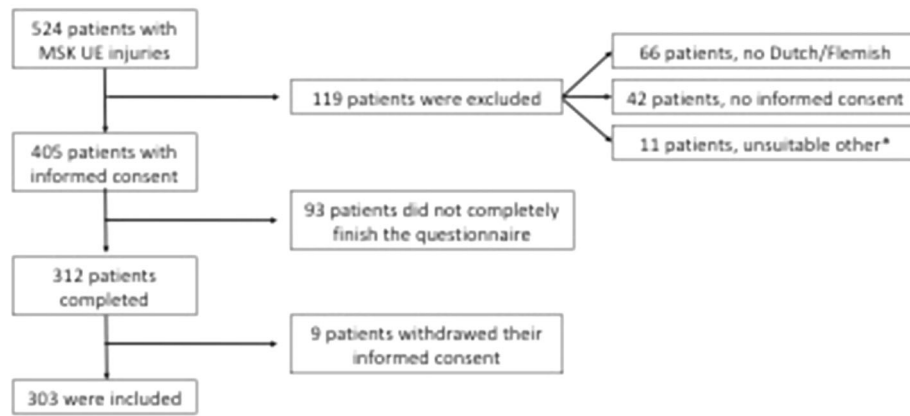
Structural validity measures the degree to which the scores of a health-related PRO instrument are an adequate reflection of the dimensionality of the construct to be measured [41]. In this study, the structural validity of the Dutch-Flemish PROMIS UE v2.0 item bank was assessed by confirmatory factor analyses (CFA). A single factor model of the Dutch-Flemish PROMIS UE v2.0 item bank was tested.

Because the item bank is supposed to measure one construct (upper extremity physical function), we expect that all items load on a single factor. Unidimensionality was examined by CFA on the polychoric correlation matrix with Weighted Least Squares with Mean and Variance adjustment (WLSMV) estimation. The Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Residual (SRMR) evaluate model fit. We report scaled fit indices, which are considered more exact than unscaled indices [42]. Following the PROMIS analysis plan [23] and recommendations by Hu and Bentler [43] we considered sufficient evidence for unidimensionality and thus adequate model fit if CFI was close to 0.95 or higher, a TLI close to 0.95 or higher, a RMSEA close to 0.06 or less and a SRMR close to 0.08 or less.

If the model did not fit well, a bi-factor model was used to examine if the scale is unidimensional enough for future IRT analyses. To evaluate the influence of multidimensionality a bi-factor model was fitted, and omega-H and Explained Common Variance (ECV) were calculated. A high omega H value indicates that a composite score is reflected by a single common source, for example one common factor underlies item responses [44,45]. The Explained Common Variance (ECV) was calculated, which is the ratio of the variance explained by the general factor, divided by the variance explained by the general factor and the group factors. A high coefficient omega (>0.80) [44] and a high ECV

Table 2. Legacy instruments.

DASH	30 items (addressed to disabilities and symptoms in musculoskeletal disorders of the upper limbs). Timeframe: during the last week. Six different 5-point Likert response scales: <ul style="list-style-type: none"> • No difficulty/Mild difficulty/Moderate difficulty/Severe difficulty/Unable • Not at all/Slightly/Moderately/Quite a bit/Extremely • Not limited at all/Slightly limited/Moderately limited/Very limited/Unable • None/Mild/Moderate/Severe/Extreme • No difficulty/Mild difficulty/Moderate difficulty/Severe difficulty/So much difficulty that I can't sleep • Strongly disagree/Disagree/Neither agree or disagree/Agree/Strongly agree. <p>Higher scores imply more disability: 0 (no disability) to 100 (most severe disability).</p>
PRWE	15 items (addressed to pain and function in patients with wrist fractures). The function subscale contains two sections; specific activities and usual activities. Timeframe: during the last week. Two different 11-point response scales: <ul style="list-style-type: none"> • Pain: 0 (no pain) to 10 (unbearable pain) • Function: 0 (no disability) to 10 (most disability) <p>Higher scores imply worse outcome: 0 (no disability and no pain) to 100 (most disability and unbearable pain).</p>
MHQ-ADL	17 items (addressed to activities of daily living). Timeframe: during the last week. One 5-point Likert response scale: <ul style="list-style-type: none"> • Not difficult at all/A little difficult/Somewhat difficult/Moderately difficult/Very difficult. <p>Higher scores imply less disability: 0 (Very difficult to do) to 100 (not difficult to do at all).</p>



*: 11 patients were excluded because of psychiatric disorders, e.g. dementia, depression and psychosis.

Figure 1. Sample size, inclusions and exclusions.

(>0.60) [45] indicate that the risk of biased parameters when fitting multidimensional data into a unidimensional model is low.

In addition, an Exploratory Factor Analysis (EFA) with WLSMV estimation procedures using the R package Psych (version 1.7.5) was performed [46]. The first factor in EFA should account for at least 20% of the variability, and the ratio of the variance explained by the first to the second factor needs to be greater than four [23,47].

Factor loadings were calculated to give a representation of the relationship of each item to the underlying factor. The factor loading is the correlation between the observed score and the latent score. Factor loadings had to be higher than the criterion of >0.50 [40,48,49].

Internal consistency

Internal consistency measures the degree of the interrelatedness among the items [41,50]. The internal consistency of the full item bank as well as the standard 7-item Short Form was determined after conducting a factor analysis. Internal consistency was assessed by calculating Cronbach's alpha. A Cronbach's alpha of >0.70 was considered sufficient evidence for internal consistency [51,52].

Construct validity

T-scores were calculated for the Dutch-Flemish PROMIS UE v2.0 item bank, based on response pattern scoring using the US item parameters. The T-scores were correlated (Pearson correlations) to the scores of the legacy instruments. For assessing convergent validity, we hypothesized that the Dutch-Flemish PROMIS UE v2.0 item bank score has a:

- Hypothesis 1: Strong negative correlation ($r \geq -0.50$) with the DASH, given the fact that both instruments are supposed to measure related constructs (UE related physical function and UE physical function and symptoms).
- Hypothesis 2: Moderately strong negative correlation ($-0.50 \leq r \leq -0.30$) with the PRWE function, because both instruments are supposed to measure hand and wrist related activities and function. Because of this expected correlation, we hypothesized that the PRWE function is stronger correlated to the Dutch-Flemish PROMIS UE v2.0 item bank than PRWE pain.
- Hypothesis 3: Moderately strong positive correlation ($0.30 \leq r \leq 0.50$) with the MHQ-ADL, because both instruments are supposed to measure related constructs of upper extremity related physical function and hand related daily activities.

Table 3. Demographic and clinical characteristics ($n = 303$).

Age	Mean	50.07	(SD 17.5)
	Range	18–93	
		N	Percentage
Gender			
	Male	159	52%
	Female	144	48%
Employment status			
	Fulltime	141	47%
	Part-time	55	18%
	Student	20	7%
	Unpaid, volunteer or household	13	4%
	Retired	49	16%
	Unemployed	6	2%
	Other	19	6%
Injury influence on employment status			
	Remained the same, no large influence	174	57%
	Modified work, function remained	59	20%
	Other function	7	2%
	Unemployed	31	10%
	Called in sick at work	32	11%
Current compensation claim			
	Yes	50	17%
	No	253	83%
Dominant hand			
	Left-handed	41	14%
	Right-handed	258	85%
	Ambidextrous	4	1%
Treatment			
	Conservative	217	72%
	Surgery	86	28%

Construct validity was considered sufficient if at least 75% of the correlations were as expected [52].

Results

Study participants

A total of 524 patients with upper extremity injuries were approached (Figure 1). A total amount of 405 patients were eligible to participate in this study and gave informed consent, of which 303 (74.8%) completed the online questionnaires (Figure 1). There were no missing values in the completed questionnaires.

Demographic and clinical characteristics are presented in Table 3. Of the 303 included patients, 159 were male (52%) and 144 females (48%). The mean age was 50.1 years (SD= 17.5), ranging from 18 until 93 years. An amount of 276 (91%) patients were born in the Netherlands.

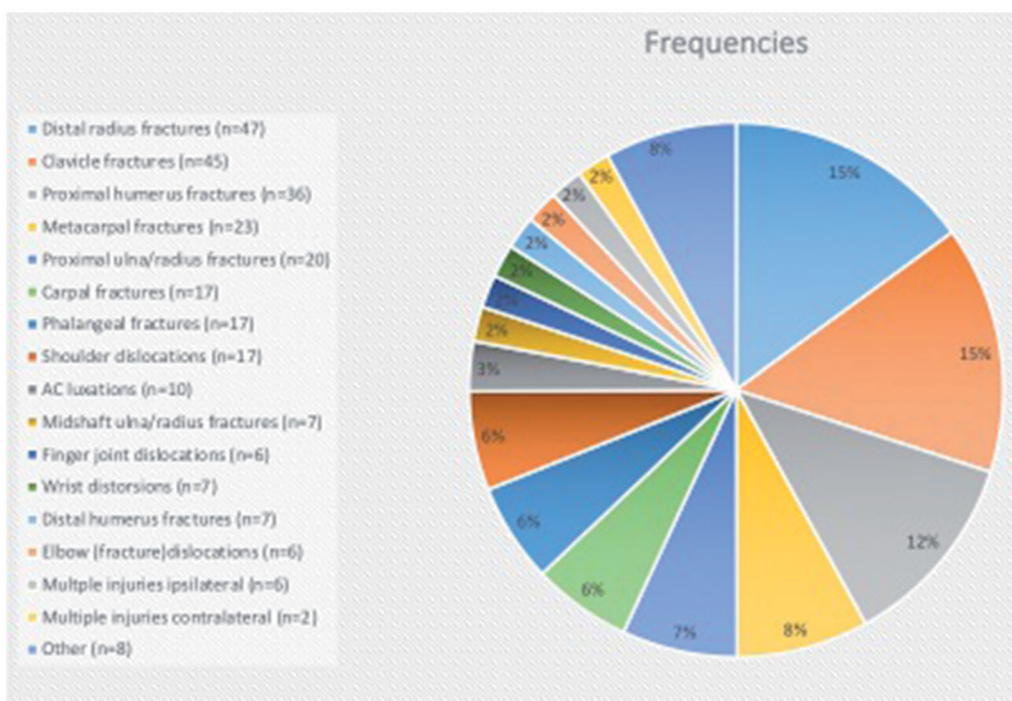


Figure 2. Types of injury, pie-chart ratio.

Regarding educational levels, 68% had at least a high school degree. The largest group (122 patients, 40%) had achieved a college degree and only 7 patients (2%) had only a primary school degree. After injury, 174 patients (57%) were able to remain their job without major adjustments. About 10% declared to be unemployed, and 11% was reported ill, due to their injury. Of the total group, 50 patients (17%) had a pending (injury) claim (Table 3). Most included patients had injuries with an acute or semi acute onset, with a follow-up duration of 1 until 4 months in 30%.

Most reported physical injuries were fractures, tendon injuries and muscle injuries.

Of all injuries, 72% were fractures, mainly distal radius fractures (16%), clavicle fractures (15%) and proximal humeral fractures (12%) (Figure 2). Conservative treatment was maintained in 217 injured patients (72%). The other 86 patients (28%) needed surgical treatment (Table 3).

Statistics analysis

Structural validity

With CFA we found a CFI of 0.94, a TLI of 0.93, a RMSEA of 0.10 and a SRMR of 0.09, which was near the reference criteria [23,43–47]. The factor loadings for this model were 0.73 or higher (mean 0.83, and range 0.73 to 0.94), which were all above the criteria of >0.50 [40,48]. A bi-factor model was investigated, to examine if the scale is unidimensional enough for future IRT analyses. Omega-H and ECV were 0.79 and 0.67 respectively, which indicated sufficient unidimensionality. In addition, in EFA three factors with eigenvalues greater than 1 were identified. The eigenvalue of the first factor was 29.6, the eigenvalue of the second and third factors were 2.7 and 2.3, respectively. The ratio of the first to the second factor was 11.0, which is larger than the criterion of 4.

Internal consistency

Cronbach’s alpha for the full Dutch-Flemish PROMIS UE v2.0 item bank was 0.98, which is above the criterion of >0.70 [51]. The

Table 4. T-scores (n = 303).

	Mean	SD
PROMIS UE v2.0	33.4	9.1
DASH	35.5	22.1
PRWE Function	26.0	15.8
PRWE Pain	20.9	14.0
PRWE Total	46.9	27.1
MHQ-ADL ^a	57.4	34.8

^aMHQ-ADL total scores per injured side (right sided, left sided or both sides injured).

Cronbach’s alpha of the PROMIS UE Short Form 7a was 0.90, which is also above the criterion of >0.70.

Construct validity

Mean (SD) T-scores for the PROMIS UE v2.0 were 33.4 (9.1), for the DASH 35.5 (22.1), for the PRWE function 26.0 (15.8), for the PRWE pain 20.9 (14.0), for the PRWE total 46.9 (27.1) and for the MHQ-ADL total (per injured side) 57.4 (34.8) (Table 4). The correlations of T-scores the Dutch-Flemish PROMIS UE v2.0 item bank with the T-scores of the DASH, PRWE function, PRWE pain, PRWE total and MHQ-ADL were, -0.84, -0.75, -0.59, -0.74, and 0.73, respectively, with all p values <.001. Only the correlation of the Dutch-Flemish PROMIS UE v2.0 item bank with the DASH met our hypothesis, a strong negative correlation. The correlations between the Dutch-Flemish PROMIS UE v2.0 item bank with PRWE function, PRWE pain, PRWE total and MHQ-ADL were higher than expected.

Discussion

The aim of this study was to examine the structural validity, internal consistency, and construct validity of the Dutch-Flemish PROMIS UE v2.0 item bank, in order to make it applicable in the outpatient clinic setting. The results show that the Dutch-Flemish PROMIS UE v2.0 item bank measures a unidimensional trait and

has sufficient structural validity, internal consistency, and construct validity.

The CFA analyses showed a CFI and TLI of 0.94 and 0.93, which are lower than, but near the criterion of >0.95 . A RMSEA of 0.10 and a SRMR of 0.09 were found, which are also near the criteria of <0.06 and <0.08 respectively. The RMSEA was higher than the maximum criterion of <0.06 [43]. Inconsistent results have been found for other versions of the PROMIS Physical Function item bank. Rose et al. reported a RMSEA of about 0.08 for subsets of items from the original English PF v1.2 item bank [16]. The Dutch-Flemish PROMIS Physical Function v1.2 item bank showed a RMSEA of 0.122 [19] and 0.045 [20] in previous studies, respectively. For the Spanish population in the US, a RMSEA of 0.052 was found for the PROMIS Physical Function v1.2 item bank [53]. The RMSEA was not reported for the US PROMIS UE population [54,55]. It has been suggested that traditional cutoffs and standards for CFA fit statistics are not suitable to establish unidimensionality of item banks measuring health concepts [56] and that the RMSEA is sensitive to model complexity (number of estimated parameters) and skewed data distributions [56], the latter being the case in health concepts. Reise et al. have stated that the RMSEA statistic is problematic for assessing unidimensionality of health concepts, and considered the SRMR more promising to determine whether a scale is “unidimensional enough” [45]. The SRMR was 0.09 in our study, slightly higher than the criterion of 0.80. None of the studies on the US PROMIS UE [54,55] or Spanish PROMIS UE item bank [53] reported the SRMR.

All factor loadings for this model were above the criterion of at least 0.50 or higher, with a smallest factor loading of 0.73 [40]. Paz et al. found factor loadings all above the criterion of 0.70 for the Spanish PROMIS Physical Function v1.2 item bank, except for 2 items (PFC7 and PFA19) but these items are not included in the UE item bank [53]. Hays et al. found PROMIS UE factor loadings all above 0.70 as well in a US population, with a smallest factor loading of 0.85 [57]. These results support the hypothesis of unidimensionality of the PROMIS UE v2.0 item bank. Other authors have recommended fitting a bi-factor model and consider the Omega H and ECV when the RMSEA does not fit the criterion of <0.06 [45,53–55]. The omega-H in our study was 0.79, which was just beneath, but very close to the criterion of >0.80 , which suggests that a composite score is reflected by a single common source. The ECV was 0.67, which is higher than the criterion of >0.60 . Together this suggests that the risk of biased parameters when fitting multidimensional data into a unidimensional model was low.

Finally, in EFA analyses the first factor accounted for more than 20% of the variability and the ratio of the variance explained by the first to the second factor was 11.0. All these results together suggest enough evidence for unidimensionality of the PROMIS UE v2.0 item bank.

Internal consistency

Evidence for sufficient internal consistency was indicated by a Cronbach’s alpha for the entire item bank of 0.98, which is higher than the criterion of >0.70 [41,50,51]. Kaat et al. found an average marginal reliability of 0.90 for the PROMIS UE item bank [54]. Beckman et al. and Paz et al. both found a Cronbach’s alpha of 0.99 for the PROMIS UE item bank [21,53]. This suggests that the internal consistency found in this study, was comparable to the internal consistency found in previous studies. The Cronbach’s alpha of the entire item bank might be this high because the scale includes 46 items. A very high Cronbach’s alpha might

suggest redundancy of items, but since the entire item bank will not be used in clinical practice, this seems not to be problem. The Cronbach’s alpha of the PROMIS UE Short Form 7a is possibly more relevant, because this short form will be used in clinical practice, instead of the entire item bank. The Cronbach’s alpha of the PROMIS UE Short Form 7a was 0.90. Chung et al. found a Cronbach’s alpha of 0.97 for the MHQ-ADL [8]. For the DASH a Cronbach’s alpha of 0.95 was found by van Eck et al. [31]. Cronbach’s alpha was also calculated by El Moumni et al. for the PRWE total, PRWE pain and PRWE function, which was 0.97, 0.94 and 0.96, respectively [3]. The Cronbach alpha’s found for the DASH, PRWE and MHQ-ADL, were all relatively high, but comparable in terms of meeting the criterion of > 0.70 to the results found for the Dutch-Flemish PROMIS UE v2.0 Short Form 7a in this study. This suggests that evidence for the internal consistency was sufficient, and comparable to the legacy instruments.

Construct validity

Our first hypothesis regarding the correlation between the Dutch-Flemish PROMIS UE v2.0 item bank and the DASH questionnaire was met; there was a strong negative correlation (-0.84). This result was comparable to results from Beckmann et al. [21]. They found a correlation of -0.80 between the US PROMIS UE v2.0 item bank and the DASH [21]. Kaat et al. found a correlation of -0.82 , and Doring et al. found a correlation of -0.81 between the PROMIS UE v2.0 item bank and the Quick DASH [54,58]. Overall, this suggests that the PROMIS UE v2.0 item bank and DASH measure similar constructs.

Our second hypothesis regarding the correlation between the Dutch-Flemish PROMIS UE v2.0 item bank and the PRWE total questionnaire was not met. We found a strong negative correlation (-0.74), instead of the expected moderately strong negative correlation ($-0.50 \leq r \leq -0.30$). For the PRWE pain and PRWE function subscales, correlations of -0.59 and -0.75 were found, supporting the hypothesis that the Dutch-Flemish PROMIS UE v2.0 item bank has a stronger correlation with the PRWE function subscale than with the PRWE pain subscale.

Our third hypothesis regarding the correlation between the Dutch-Flemish PROMIS UE v2.0 item bank and the MHQ-ADL was not met as well. A moderately strong positive correlation was expected ($0.30 \leq r \leq 0.50$), though we found a strong positive correlation (0.73). Apparently, the Dutch-Flemish PROMIS UE v2.0 item bank has more content overlap with the PRWE and MHQ-ADL than we expected. This could actually be considered a positive finding because it shows that the Dutch-Flemish PROMIS UE v2.0 item bank is capable of measuring upper extremity related physical function, as well as hand- and wrist related function, comparable to the DASH, PRWE and MHQ-ADL.

Strengths

A sample size of at least 300 participants was achieved, meeting the recommendations by Comrey and Lee [40]. Patients of all ages and with all kind of upper extremity injuries were included, which supports the representativeness of the study population. The main experienced upper extremity injuries were distal radius fractures, clavicle fractures and proximal humeral fractures. A study by Beerekamp *et al.*, performed in the Netherlands, estimated the prevalence of extremity fractures in general [59]. The most commonly reported fractures, were hand and finger fractures ($n=34.144$), wrist fractures ($n=25.432$) and clavicle and shoulder fractures ($n=13.264$) [59]. Hand and finger fractures

included carpal, metacarpal and phalangeal fractures together [59]. In our sample the sum of these injuries was 20% ($n=57$). These results were comparable to the results of the Beerekamp sample.

This study was conducted according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias-checklist. Adequate study design for cross-sectional validity and construct validity were ensured [51,60]. Besides, the international PROMIS guidelines for instrument development and validation were followed [25].

Weaknesses

Because the hospital the patients were recruited in is a level 1 trauma center, patients with severe and multiple injuries (poly-trauma patients) are overrepresented. In community hospitals, there are less severely injured patients, the mono-trauma patients. These differences can cause bias in the sample and have effect on the representativeness of the study sample. Severely injured patients, often have multiple fractures and corresponding soft-tissue injuries, which has a negative effect on outcome. This might explain a higher amount of severely injured patients with a higher amount of worse outcome in level 1 trauma centers. The implementation of the Dutch-Flemish PROMIS UE v2.0 in outpatient clinical environment might be challenging, due to the need of mobile devices and accessibility to internet. However, almost all patients own a mobile device and internet-based PROs have been implemented worldwide. We therefore think that once an internet-based questionnaire including the Dutch-Flemish PROMIS UE v2.0 has been implemented in your outpatient clinic, it can decrease the burden for the patient and improve interpretation of outcome tremendously.

Clinical interpretation

For daily practice, the Dutch-Flemish PROMIS UE v2.0 item bank is a suitable instrument to measure rehabilitation progress, in comparison with the legacy instruments (DASH, PRWE and MHQ-ADL). The benefit of the Dutch-Flemish PROMIS UE v2.0 item bank is that it can be applied across patient populations, enabling comparison of scores, and it can be used as CAT, which reduces response burden for patients and increases the usability of the Dutch-Flemish PROMIS UE v2.0 item bank in daily practice. Therefore, we recommend future studies to consider using PROMIS instead of the legacy instruments.

Conclusion

This study showed that the Dutch-Flemish PROMIS UE v2.0 item bank measures a unidimensional trait and sufficient structural validity, internal consistency and construct validity were found. Further studies should assess further validation and calibration by IRT analysis, as well as other measurement properties, such as test-retest reliability, measurement error, and responsiveness. After successful IRT analysis, the Dutch-Flemish PROMIS UE v2.0 CAT will be operable to use.

Disclosure statement

One of the authors is president of the PROMIS Health Organization, but receives no compensation for this position.

References

- [1] van der Zee-Neuen A, Putrik P, Ramiro S, et al. Impact of chronic diseases and multimorbidity on health and health care costs: the additional role of musculoskeletal disorders. *Arthritis Care Res (Hoboken)*. 2016;68:1823–1831.
- [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;7:7.
- [3] El Mounni M, Van Eck ME, Wendt KW, et al. Structural validity of the Dutch version of the patient-rated wrist evaluation (PRWE-NL) in patients with hand and wrist injuries. *Phys Ther*. 2016;96:908–916.
- [4] Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” - a practical guideline. *Trials*. 2016;17(1):449.
- [5] Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012;21:1305–1314.
- [6] Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: The DASH (Disabilities of the Arm, Shoulder, and Hand) (vol 29, pg 602, 1996). *Am J Ind Med*. 1996;30:372.
- [7] MacDermid JC, Turgeon T, Richards RS, et al. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma*. 1998;12:577–586.
- [8] Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the Michigan hand outcomes questionnaire. *J Hand Surg Am*. 1998;23:575–587.
- [9] Huang H, Grant JA, Miller BS, et al. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med*. 2015;43:2572–2582.
- [10] Thoomes-de Graaf M, Scholten-Peeters GGM, Schellingerhout JM, et al. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and “activity limitations”: a systematic review. *Qual Life Res*. 2016;25:2141–2160.
- [11] Bot SD, Terwee CB, van der Windt DA, et al. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis*. 2004;63:335–341.
- [12] Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61:623–632.
- [13] Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63:1179–1194.
- [14] Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) Progress of an NIH roadmap cooperative group during its first two years. *Med Care*. 2007;45:S3–S11.
- [15] Rose M, Bjorner JB, Becker J, et al. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61:17–33.
- [16] Rose M, Bjorner JB, Gandek B, et al. The PROMIS Physical Function item bank was calibrated to a standardized metric

- and shown to improve measurement efficiency. *J Clin Epidemiol.* 2014;67:516–526.
- [17] Terwee CB, Roorda LD, de Vet HC. Dutch-Flemish translation of 17 item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Qual Life Res.* 2014;23:1733–1741.
- [18] Oude Voshaar MAH, ten Klooster PM, Glas CAW, et al. Calibration of the PROMIS physical function item bank in Dutch patients with rheumatoid arthritis. *PLoS One.* 2014;9:e92367.
- [19] Crins MHP, Terwee CB, Klausch T, et al. The Dutch-Flemish PROMIS Physical Function item bank exhibited strong psychometric properties in patients with chronic pain. *J Clin Epidemiol.* 2017;87:47–58.
- [20] Crins MHP, van der Wees PJ, Klausch T, et al. Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. *PLoS One.* 2018;13:e0192187.
- [21] Beckmann JT, Hung M, Voss MW, et al. Evaluation of the patient-reported outcomes measurement information system upper extremity computer adaptive test. *J Hand Surg Am.* 2016;41:739–744 e4.
- [22] Hung M, Voss MW, Bounsanga J, et al. Examination of the PROMIS upper extremity item bank. *J Hand Ther.* 2017;30:485–490.
- [23] Reeve BB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007;45(5): S22–S31.
- [24] [The Helsinki Declaration of the World Medical Association (WMA). Ethical principles of medical research involving human subjects]. *Pol Merkur Lekarski.* 2014;36:298–301.
- [25] PROMIS. PROMIS Instrument Development and Validation Scientific Standards Version 2.0 (revised May 2013); 2013.
- [26] Van Son MAC, De Vries J, Roukema JA, et al. Health status and (health-related) quality of life during the recovery of distal radius fractures: a systematic review. *Qual Life Res.* 2013;22:2399–2416.
- [27] Changulani M, Okonkwo U, Keswani T, et al. Outcome evaluation measures for wrist and hand: which one to choose?. *Int Orthop.* 2008;32:1–6.
- [28] Hoang-Kim A, Pegreff F, Moroni A, et al. Measuring wrist and hand function: common scales and checklists. *Injury.* 2011;42:253–258.
- [29] 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; quiz 236.
- [30] Veehof MM, Slegers EJA, van Veldhoven NHMJ, et al. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther.* 2002;15:347–354.
- [31] van Eck ME, Lameijer CM, El Moumni M. Structural validity of the Dutch version of the disability of arm, shoulder and hand questionnaire (DASH-DLV) in adult patients with hand and wrist injuries. *BMC Musculoskelet Disord.* 2018;19:207.
- [32] Brink SM, Voskamp EG, Hout P, et al. Psychometric properties of the Patient Rated wrist/hand evaluation - Dutch Language Version (PRWH/E-DLV). *J Hand Surg Eur Vol.* 2009;34:556–557.
- [33] Chung KC, Hamill JB, Walters MR, et al. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Ann Plast Surg.* 1999;42:619–622.
- [34] Chung BT, Morris SF. Confirmatory factor analysis of the Michigan Hand Questionnaire. *Ann Plast Surg.* 2015;74:176–181.
- [35] van der Giesen FJ, Nelissen RG, Arendzen JH, et al. Responsiveness of the Michigan Hand Outcomes Questionnaire-Dutch language version in patients with rheumatoid arthritis. *Arch Phys Med Rehabil.* 2008;89:1121–1126.
- [36] Chung BT, Morris SF. Reliability and internal validity of the Michigan hand questionnaire. *Ann Plast Surg.* 2014;73:385–389.
- [37] Maia MVDP, de Moraes VY, dos Santos JBG, et al. Minimal important difference after hand surgery: a prospective assessment for DASH, MHQ, and SF-12. *Sicot-J.* 2016;2:32.
- [38] van de Ven-Stevens LA, Munneke M, Terwee CB, et al. Clinimetric properties of instruments to assess activities in patients with hand injury: a systematic review of the literature. *Arch Phys Med Rehabil.* 2009;90:151–169.
- [39] Hogarty KY, Hines CV, Kromrey JD, et al. The quality of factor solutions in exploratory factor analysis: the influence of sample size, communalities, and overdetermination. *Educ Psychol Meas.* 2005;65:202–226.
- [40] Comrey AL, Lee HB. *First course in factor analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1992.
- [41] 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;63:737–745.
- [42] Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika.* 2001;66:507.
- [43] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1–55.
- [44] Rodriguez A, Reise SP, Haviland MG. Applying bifactor statistical indices in the evaluation of psychological measures. *J Pers Assess.* 2016;98:223–237.
- [45] Reise SP, Scheines R, Widaman KF, et al. Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educ Psychol Meas.* 2013;73:5–26.
- [46] Revelle W. *Psych: procedures for personality and psychological research;* 2017. Software
- [47] Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107:238–246.
- [48] Stevens J. *Applied multivariate statistics for the social sciences.* 2nd ed. Hillsdale, Michigan: L. Erlbaum Associates; 1992.
- [49] Peterson RA. A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Market Lett.* 2000;11:261–275.
- [50] Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol.* 1993;78:98–104.
- [51] Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
- [52] Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27:1147–1157.
- [53] Paz SH, Spritzer KL, Morales LS, et al. Evaluation of the Patient-Reported Outcomes Information System (PROMIS((R))) Spanish-language physical functioning items. *Qual Life Res.* 2013;22:1819–1830.

- [54] Kaat AJ, Rothrock NE, Vrahas MS, et al. Longitudinal validation of the PROMIS physical function item bank in upper extremity trauma. *J Orthop Trauma*. 2017;31:e321–e326.
- [55] Gausden EB, Levack AE, Sin DN, et al. Validating the Patient Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests for upper extremity fracture care. *J Shoulder Elbow Surg*. 2018;27:1191–1197.
- [56] Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res*. 2009;18:447–460.
- [57] Hays RD, Spritzer KL, Amtmann D, et al. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Arch Phys Med Rehabil*. 2013;94:2291–2296.
- [58] Döring A-C, Nota SPFT, Hageman MGJS, et al. Measurement of upper extremity disability using the patient-reported outcomes measurement information system. *J Hand Surg Am*. 2014;39:1160–1165.
- [59] Beerekamp MSH, de Muinck Keizer RJO, Schep NWL, et al. Epidemiology of extremity fractures in the Netherlands. *Injury*. 2017;48:1355–1362.
- [60] Mookink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1171–1179.