RELIABILITY, VALIDITY AND RESPONSIVENESS OF THE SLOVENIAN VERSION OF THE PATIENT EVALUATION MEASURE (PEM-Slo) IN PATIENTS WITH WRIST AND HAND DISORDERS

ZANESLJIVOST, VELJAVNOST IN ODZIVNOST SLOVENSKE RAZLIČICE VPRAŠALNIKA BOLNIKOVA OCENA STANJA (PEM-Slo) PRI BOLNIKIH Z OKVARO ZAPESTJA IN ROKE

Vida BOJNEC [1, 2*], Dragan LONZARIĆ [1, 2], Živa KLARER REBEC [3, 4]

[1]University Medical Centre Maribor, Institute for Physical and Rehabilitation Medicine, Ljubljanska ulica 5, 2000 Maribor, Slovenia
[2]Faculty of Medicine, University of Maribor, Taborska ulica 8, 2000 Maribor, Slovenia
[3]Community Healthcare Centre Celje, Physical Medicine and Rehabilitation Unit, Gregorčičeva ulica 5, 3000 Celje, Slovenia
[4]Thermana Laško Spa Resort, Thermana d. d., Zdraviliška cesta 6, 3270 Laško, Slovenia

ABSTRACT

Introduction: The Patient Evaluation Measure (PEM) is a region-specific patient reported outcome measure (PROM) for hand and wrist disorders, first introduced in English for patients with hand surgery in 1995. The purpose of the study was to assess the psychometric properties of the translated and cross-culturally adapted Slovenian version of PEM (PEM-Slo).

Methods: The study was designed as a single-centre observational prospective study conducted from July 2020 to March 2021. The psychometric evaluation was performed on fifty-one patients with miscellaneous hand and wrist disorders. Reliability was tested for internal consistency and test-retest reliability. Convergent and divergent validity, responsiveness, floor and ceiling effect, and interpretability with the determination of minimal detectable change (MDC) and minimal clinically important difference (MCID) were assessed.

Results: The PEM-Slo has excellent internal consistency (Cronbach's α 0.932) and good to excellent test-retest reliability (intraclass correlation coefficient=0.874). Convergent validity was proved with high to moderate correlations of PEM-Slo with DASH, grip strength and self-care, usual activities, and pain EQ-5D-5L subscales, whereas no correlation of PEM-Slo with EQ-5D-5L mobility and anxiety/depression subscale confirmed divergent validity. The PEM-Slo responsiveness was high (standardised response mean=1.42, effect size=1.25). MDC was 18.01 and MCID was 17.31. No floor or ceiling effect was found.

Conclusion: The PEM-Slo is a reliable, valid and responsive PROM for Slovenian-speaking patients with hand and wrist disorders.

IZVLEČEK

Uvod: *Vprašalnik Bolnikova ocena stanja (angl. Patient Evaluation Measure – PEM) je samoocenjevalni regijsko specifičen vprašalnik za bolnike z okvaro zapestja in roke, ki vsebuje 18 postavk, razdeljenih v 3 dele. Prvič je bil uporabljen v angleškem jeziku v Veliki Britaniji leta 1995 pri bolnikih po operativnem posegu na roki, kasneje se je njegova uporaba razširila na bolnike z različnimi okvarami roke. Namen raziskave je bil oceniti psihometrične lastnosti predhodno prevedene in medkulturno prilagojene slovenske različice vprašalnika (PEM-Slo).*

Metode: *Raziskava je potekala v času od julija 2020 do marca 2021 in je bila zasnovana kot prospektivna observacijska študija. Prevod vprašalnika PEM je bil izdelan po veljavnih priporočilih za dvosmerno prevajanje. Razumevanje prevoda je bilo preizkušeno na 12 preiskovancih. V psihometrično analizo vprašalnika PEM je bilo vključenih 51 preiskovancev z različnimi okvarami zapestja in roke. Ob prvem ocenjevanju (T1) so preiskovanci izpolnili vprašalnike PEM, DASH in EQ-5D-5L, izmerjena je bila mišična moč stiska roke. Ob drugem ocenjevanju po 2 do 4 dneh (T2) so ponovno izpolnili vprašalnik PEM. Zadnja ocena (T3) je bila enaka oceni ob T1. Zanesljivost vprašalnika PEM je bila ocenjena z notranjo skladnostjo in zanesljivostjo ponovljene meritve. Ocenili smo tudi standardno napako merjenja in najmanjšo zaznavno spremembo, konstruktno veljavnost s konvergentno in divergentno veljavnostjo glede na povezanost podobnih in različnih konstruktov vseh treh vprašalnikov ter povezanost ocene PEM z mišično močjo stiska roke. Ocenjen je bil tudi učinek tal in stropa vprašalnika PEM in njegova najmanjša klinično pomembna sprememba.*

Rezultati: *Povprečna starost 51 preiskovancev (26 moških) je bila 53 let. Slovenski prevod so preiskovanci sprejeli kot dobro razumljivega. PEM-Slo ima odlično notranjo skladnost (Cronbachova α 0,932) in dobro do odlično zanesljivost ponovljene meritve (znotrajrazredni količnik povezanosti s 95-% območjem zaupanja (ICC) je znašal 0,874). Povezanost med dosežki PEM-Slo, DASH ter med nekaterimi postavkami vprašalnika EQ-5D-5L (skrb zase, vsakdanje aktivnosti, bolečina/neugodje) ter z močjo stiska roke je bila zmerne do močne stopnje, kar kaže na konvergentno veljavnost PEM-Slo. Ni pa bilo statistično značilne povezanosti med dosežkom PEM-Slo in postavkama pokretnost, tesnoba/potrtost EQ-5D-5L, kar potrjuje divergentno veljavnost. Tudi odzivnost vprašalnika PEM-Slo je bila visoka. Najmanjša zaznavna sprememba PEM-Slo je znašala 18,01, najmanjša klinično pomembna sprememba pa 17,31. Učinka tal in stropa nismo ugotovili.*

Zaključek: *Slovenska različica vprašalnika PEM-Slo je zanesljiva, veljavna in odzivna samoocenjevalna lestvica za slovensko populacijo bolnikov z okvarami zapestja in roke in je prosto dostopna za uporabo v klinične in raziskovalne namene.*

*Correspondence: bojnec.vida@gmail.com

1 INTRODUCTION

The hand is of paramount importance for humans to interact with their environment. Any impairment of hand function can have an enormous impact on one's wellbeing. When assessing patients with hand disorders in the clinical and research environment, objective measures such as range of motion, joint stability and grip strength are correlated with subjective patient reported outcome measures (1, 2). The latter give us information about functioning and quality of life from the patient's perspective in their own specific living and working environment, which cannot be obtained with mere objective measurements (2).

The selection of the patient reported outcome measure for clinical and research work should be based on its simple use regarding clarity and time to answer the questions (3). The results obtained should be comparable to the results of other studies and not influenced by differences in the socio-cultural environment (4). When the patient reported measure is used in an environment where a language other than its original version is spoken, the translation and cross-cultural adaptation must follow the uniform principle for translation and cross-cultural adaptation (4). According to the COnsensus based Standards for the selection of health Measurement INstruments (COSMIN) group, the patient reported outcome measure should have good psychometric properties in terms of being reliable, valid and responsive (5). It is reliable when it is free of measurement errors, valid when it measures what it is intended to measure, and responsive to detect change over time in the construct it measures (5).

The Patient Evaluation Measure (PEM) was first introduced in 1995 to be used in patients with hand surgery (2), followed by its application in patients with diverse hand disorders, i.e. after carpal tunnel surgery (6, 7), distal radius fracture (8), scaphoid fracture (9) and in patients with miscellaneous hand pathology (10). It is a self-administered, region-specific measure for patients with wrist and hand disorders and consists of 18 items divided into three sections. The first section (Part One: Treatment) has five items evaluating the treatment process and is not included in the final score. Section two (Part Two: How your hand is now) is focused on hand and wrist impairments and contains ten items, of which three are about the symptoms (e.g. feeling and pain), five about hand function (e.g. movement and grip), and two items about the impact the hand impairment has on the patient (e.g. daily activities and work). Dias et al. have added one item to the second section, asking about the duration of the pain in the hand (9). The third section (Part Three: Overall assessment) has only three items asking about general satisfaction with the treatment and hand status (2). The PEM is easy to understand, requires little time for completion and can easily be analysed, which makes it more appropriate for use in daily clinical practice (6, 8,

10) compared to e.g. the Michigan Hand Questionnaire. PEM unanimously assesses only one hand which is advantageous in comparison to DASH and QuickDASH.

The aim of this study was to translate and cross-culturally adapt PEM into Slovenian and further evaluate its psychometric properties.

2 METHODS

2.1 Study design

The study was a single-centre observational prospective study conducted from July 2020 to March 2021. Ethical approval was obtained from the UMC Maribor Ethics Committee on May 11, 2020 (UKC-MB-KME-20/20). Written informed consent was obtained from all the participants prior to inclusion in the study.

The study participants were outpatients with hand and wrist disorders aged 18 years and above, all native Slovenian speakers. The exclusion criteria were 1) not being able to answer the questionnaire due to cognitive impairment, 2) bilateral disorders of wrists or hands, 3) rheumatologic or neurologic disease affecting the upper extremities.

2.2 Outcome measures

PEM is a self-administered, region-specific patient-reported outcome measure for patients with wrist and hand disorders and consists of 18 items divided into three sections. Part One is not included in the final score (2). The items of PEM are scored on a scale from 1 to 7, where 1 represents the best and 7 the worst score (2, 11). The previous validation study of PEM calculated the PEM score by summing the values of each item in Part Two and Part Three, and expressed the sum of the two as a percentage of the maximum score (9). Since no specific instructions for the calculation were given, we developed a formula to express a score on a scale as percentage points from 0–100, a higher score representing a worse functional outcome. The formula used to calculate the PEM score is as follows: ((calculated sum–lowest possible sum)/range) ×100 i.e. ((calculated sum–13)/78) ×100.

The Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) is a region-specific patient-reported outcome measure for patients with upper extremity impairments (12). It consists of 30 items scored on a five-point scale where 1 represents the best and 5 the worst score. The official Slovenian translation of DASH was used (13).

The EQ-5D-5L is a generic patient-reported outcome measure assessing general health related quality of life (HRQoL) (14). It consists of five dimensions rated on a five-level descriptive system. Each combination of answers has an index-based value assigned ranging from -0.495 to 1 in a given crosswalk value-set for Slovenia (15). An official Slovenian translation of EQ-5D-5L was used in the study.

The Global Rating of Change (GROC) scale used in our study is an 11-point scale allowing patients to rate their improvement. It was used as the external anchor to dichotomise the patients into improvers and non-improvers. The question for the GROC scale was formulated as follows: "How is your hand today compared to the day when you first completed the questionnaires?" The possible answers were 0 for no change, -1 to -5 as deterioration and +1 to +5 as improvement. The minimal clinically important change for an 11-point GROC scale is 2 points (16). The arbitrary value for a meaningful improvement was set at 3 points and above to point out the meaningful improvement surpassing the 2 points that represent the minimal clinically important difference (MCID). According to der Roer et al. a change in scores bellow the MCID value should be considered as irrelevant to patient improvement, whereas changes that fall within the range of MCID can be influenced by other factors (such as patient satisfaction with therapy or return to work) and may be false since they may lie within the measurement error (17).

Grip strength was measured with a hydraulic Jamar® hand dynamometer (Sammons Preston Rolyan Inc., USA) following a standardised procedure (18). The measurements were taken three times with each hand and the mean value for each hand was calculated. The values given are expressed as the percentage of the grip strength value of the contralateral unaffected side.

2.3 Stages of the study

The first stage of the study was the translation of PEM following the guidelines proposed by Beaton et al. (4).
The second part of the study was cognitive interviews with twelve patients with hand or wrist disorders.

The third stage was the assessment of the psychometric properties of PEM-Slo. According to COSMIN guidelines, a sample size of 50 participants is the objective for a study to have good quality status when assessing psychometric properties (19). Fifty-one native Slovenian-speaking patients with unilateral wrist and hand disorders treated at our rehabilitation institute were included in the study.
At their first outpatient appointment (T1), the participants were asked to complete the three patient-reported outcome measures (PEM, DASH and EQ-5D-5L). They completed the PEM for the second time (T2) after two to four days. All three patient-reported outcome measures were completed again at the end of the outpatient rehabilitation programme (T3). Grip strength was measured twice, at T1 and T3 time points. At T3, participants also assessed their level of improvement on the GROC scale.

2.4 Translation and cross-cultural adaptation of PEM-Slo

The original English version of PEM was independently translated into Slovenian by two native speakers: a physical and rehabilitation medicine (PRM) specialist and a professional translator. Two bilingual native English speakers without medical background and without being familiar with the original English version performed the back-translation. The two back-translated versions were compared with each other and with the original version, where no alteration in any question connotation was observed. At the expert group meeting consisting of the professional translator, linguistic expert for Slovenian and two PRM specialists, the final PEM-Slo version was reached unanimously.

Cognitive interviews were performed on a group of 12 native Slovenian-speaking patients (6 men) with an average age of 52 years (range 34–84) with hand and wrist disorders treated at our rehabilitation institute. The interviews were conducted by one of the authors (ZKR) who provided guidance when answering the PEM-Slo question by question, ensuring that any uncertainties of understanding the questions was reported and discussed. Only some minor uncertainties were raised. Those were regarding the questions in the first part, where it was not clear to the participants which doctor the questions were about since they were treated by a trauma surgeon before visiting a PRM specialist. Another ambiguity was the time point in Part Two, when participants were not certain what the word "now" signifies. They raised questions such as: "Do you mean now, at this moment, or today, or in the last few days?" Since the described uncertainties are not related to possible translational flaws or cross-cultural differences no change to the questionnaire was made.

2.5 Reliability

The reliability is calculated by Cronbach's α, where values above 0.7 should be reached, indicating that items are adequately correlated and measure the same construct (5,20). The PEM completed at T1, T2 and T3 was used in this analysis. Another reliability measure is expressed with ICC measuring the stability level of the patient-reported outcome measure over time at repeated measurements when no true health status change is expected (5). ICC was used for test-retest reliability purposes: two-way mixed effects, consistency, single rater/measurement model selection (21).

The PEM completed at T1 and T2 in the interval of 2 to 4 days when no true health status change is expected were used in this analysis, as well as in calculating the measurement error expressed with standard error of measurement (SEM), defined as a change in the patient's score occurring by chance (5). The SEM was used to calculate the minimal detectable change (MDC) (22).

2.6 Validity

Construct validity of PEM was assessed by calculating the correlations among PEM, DASH, EQ-5D-5L and grip strength. We hypothesised the correlations to be high between PEM and DASH, PEM and some items of EQ-5D-5L (self-care, usual activities, pain), and PEM and grip strength since they measure similar constructs. We expected to find a low or no correlation between PEM and some EQ-5D-5L items, which measure different constructs (i.e., mobility, anxiety/depression). Values obtained at T1 were used in the analysis. The correlations are considered high when the correlation coefficient is >0.6, moderate when r=0.3–0.6 and low when r<0.3 (23).

2.7 Responsiveness

To evaluate the ability of the PEM, DASH and EQ-5D-5L to detect clinical change, the standardised response mean (SRM) was calculated from the mean score improvement divided by standard deviation (SD) of score improvement. The SRM is considered large when above 0.8, moderate when the value is between 0.5–0.8, and small when less than 0.5 (24). The T1 and T3 values were used in this analysis. Effect size (ES) with Cohen's d was calculated for all the patient-reported outcome measures in the study using T1 and T3 values. The ES is interpreted as small when d=0.2, moderate when d=0.5, and strong when d=0.8 (25).

2.8 Interpretability – assessment of minimal clinically important difference (MCID)

The anchor-based approach was used for the MCID evaluation with the receiver operating characteristic (ROC) method using the GROC scale as the external anchor to dichotomise patients. The discriminative power of PEM to differentiate improvers from non-improvers was estimated with the area under curve (AUC), where values less than 0.5, 0.5 to 0.7, 0.7 to 0.8, 0.8 to 0.9, and above 0.9 denote no, poor, acceptable, excellent and outstanding discrimination, respectively (26).

2.9 Floor and ceiling effect

Floor or ceiling effects are present if more than 15% of the population achieves the lowest or highest possible score (22). Values of PEM at T1 and T3 were used in this analysis.

2.10 Statistical methods

Simple descriptive statistics was used for the analysis of demographic and clinical patients' data. The IBM SPSS 26 was used for data analysis (IBM Corp., Armonk, NY). The data distribution was analysed using the Shapiro-Wilk test. The Wilcoxon signed-rank test was used to compare the initial and final values (level of significance, p<0.05). Internal consistency was calculated with Cronbach's α. ICC

with 95%CI was used for calculating test-retest reliability. SEM was calculated from the square root of variance between measurements and the error variance of the ICC. MDC with a confidence level of 95% was calculated using the following formula: $MDC=SEM \times 1.96 \times \sqrt{2}$ (22). Spearman's correlation coefficient was used to assess correlations among variables. ROC curve was plotted for PEM and its cut-off point was determined by the maximum value of Youden's index: sensitivity + specificity –1 (27). The AUC was estimated for the accuracy of PEM to distinguish improvers from non-improvers. ES and SRM were calculated for PEM, DASH and EQ-5D-5L by comparing values at T1 and T3.

3 RESULTS

The total number of patients in the study was 51 (25 men), and the mean age was 53 years. The most common type of injury was distal radius fracture. The average timespan between T1 and T2, and T1 and T3 was 2.3 and 22.2 days, respectively. The patients received on average 10 rehabilitation sessions. The demographic and clinical data is presented in Table 1.

Table 1. Demographic and clinical data of patients (n=51), mean number of rehabilitation sessions, and timespan between assessments.

Patients Rehabilitation sessions Timespan between assessments	Number Number Days
Sample size (n)	51
Mean age (years(range))	53.3 (18-78)
Men (n(%))	25 (49)
Distal forearm fracture (n(%))	25 (48.1)
Carpal bones fracture (n(%))	2 (3.8)
Wrist soft tissue injury (n(%))	3 (5.8)
Metacarpal fracture (n(%))	3 (5.8)
Finger fracture (n(%))	7 (13.5)
Interphalangeal joint luxation (n(%))	5 (9.6)
Tendon injury (n(%))	2 (3.8)
Hand soft tissue injury (n(%))	4 (7.7)
Dominant side affected (n(%))	29 (56.9)
Treated surgically (n(%))	16 (31.4)
Number of rehabilitation sessions (mean(range))	10.2 (3-23)
Timespan from T1 to T2 (days) (mean(range))	2.3 (2-4)
Timespan from T1 to T3 (days) (mean(range))	22.2 (10-80)

Legend: T1 - initial assessment at the beginning of the rehabilitation programme; T2 - second assessment; T3 - final assessment at the conclusion of the rehabilitation programme.

3.1 Reliability

Internal consistency for PEM expressed with Cronbach's α at T1 was 0.932, at T2 0.938, and at T3 0.953. The ICC values with 95%CI for individual items and the overall PEM score are given in Table 2. The SEM was 6.497 and used to determine the MDC95, which was 18.01.

3.2 Validity

A high correlation with r=0.757 was proved only between PEM and DASH, moderate correlations with r=0.415-0.6 between PEM and grip strength, and EQ-5D-5L-Self-care, EQ-5D-5L-Usual activities, EQ-5D-5L-Pain, EQ-5D-5L-Anxiety/Depression and EQ-5D-5L¬-total. No correlation was found between PEM and EQ-5D-5L-Mobility (r=0.185 was not statistically significant) (Table 3).

3.3 Responsiveness

All the observed changes between baseline and follow-up measurements were statistically significant (Table 4). PEM and DASH were highly responsive with ES and SRM 1.25 and 1.42, and 0.85 and 1.08, respectively. EQ-5D-5L showed only moderate responsiveness with ES 0.66 and SRM 0.78.

Table 2. Test-retest reliability of PEM-Part Two, PEM-Part Three, and PEM total expressed with intraclass correlation coefficient (ICC) with 95% confidence interval (95%CI).

PEM item/question (Q)	ICC (95%CI)
Part Two	
Q1	0.599 (0.228-0.774)
Q2	0.791 (0.630-0.882)
Q3	0.824 (0.687-0.901)
Q4	0.789 (0.626-0.881)
Q5	0.798 (0.642-0.886)
Q6	0.684 (0.440-0.882)
Q7	0.848 (0.730-0.914)
Q8	0.871 (0.770-0.927)
Q9	0.714 (0.493-0.839)
Q10	0.825 (0.690-0.901)
Part Three	
Q1	0.702 (0.471-0.832)
Q2	0.801 (0.647-0.888)
Q3	0.740 (0.539-0.853)
PEM total (Part Two and Part Three)	0.874 (0.777-0.929)

Legend: Q - question, item; PEM - Patient Evaluation Measure; ICC - intraclass correlation coefficient; 95%CI - 95% confidence interval.

Table 3. Correlations of different outcome measures with PEM at baseline (T1, *n*=51).

	PEM (r)	*p value*
Grip strength	0.415	0.003
DASH	0.757	<0.001
EQ-5D-5L	0.600	<0.001
EQ-5D-5L-Mobility	0.185	0.193
EQ-5D-5L-Self-care	0.547	<0.001
EQ-5D-5L-Usual activities	0.505	<0.001
EQ-5D-5L-Pain	0.575	<0.001
EQ-5D-5L-Anxiety/Depression	0.417	0.002

Legend: PEM - Patient Evaluation Measure; DASH - Disability of Arm, Shoulder and Hand Questionnaire; EQ-5D-5L - EuroQol questionnaire; r - Spearman's correlation coefficient.

Table 4. Total scores of outcome measures at baseline (T1), second assessment (T2), and follow-up (T3) for all patients (n=51), observed change (T3)-(T1), standardised response mean (SRM) and effect size (ES).

Outcome measure	Baseline (T1) mean (SD) (range)	Two to four days after T1 (T2) mean (SD) (range)	Follow-up (T3) mean (SD) (range)	Observed change (T3) – (T1) * mean (SD) (range)	ES	SRM
PEM	44.1 (18.7) (3.9-82.1)	37 (17.9) (3.9-69.2)	22 (16.6) (0-73.1)	- 21 (14.8) (-3.8– -68)	1.25	1.42
DASH	41.6 (23.5) (5-91)	/	23.4 (18.9) (0-71.6)	-18.8 (17.4) (-55.2– -21.6)	0.85	1.08
EQ-5D-5L	0.7 (0.15) (0.195-1)	/	0.799 (0.15) (0.522-1)	0.097 (0.125) (-0.193-0.506)	0.66	0.78
Grip strength (% of the unaffected hand)	40.6 (26.4) (7-99)	/	61.2 (26.2) (9-108)	19.7 (19.9) (-10-70)	0.81	0.99

Legend: PEM - Patient Evaluation Measure; DASH - Disability of Arm, Shoulder and Hand questionnaire; EQ-5D-5L - EuroQol questionnaire; SD - standard deviation; SRM - standardised response mean; ES - effect size; * Wilcoxon signed-rank test with level of significance $p<0.001$.

3.4 Interpretability

Twenty-six patients with GROC score 3 or above represented a cohort with a meaningful improvement whereas 25 of them with a GROC score of 2 or less fell into the cohort without meaningful improvement. The two cohorts were used to plot the ROC curve, which is shown in Figure 1. The cut-off point was determined with Youden's index representing the MCID of PEM-Slo in our study, which was 17.31 with a 76.2% sensitivity and 80.8% specificity. AUC was 0.869 (95%CI 0.769–0.969).

3.5 Floor and ceiling effect

No maximal or minimal values of PEM were attained at T1, whereas only one minimal and no maximal score was observed at T3. This represents 1.9% of the total sample reaching minimal (best) value at T3.

4 DISCUSSION

The psychometric analysis of this study shows the PEM-Slo to be a reliable, valid, and responsive patient-reported outcome measure for patients with unilateral hand and wrist disorders. The reliability assessment with Cronbach's α above 0.9 at all time points shows the PEM-Slo to have an excellent internal consistency with highly interrelated items of PEM (20). Hobby et al. obtained the value of Cronbach's α of 0.94 in patients after surgical decompression of the carpal tunnel (CT) (6). This is similar to the findings in the non-specified cohort of patients comparing three patient-reported outcome measures (0.91 for Part Two and 0.88 for PEM Part Two and Part Three) (11), whereas the same value of Cronbach's α as in our study (0.932) for PEM total (Part Two and Part Three) was reported in patients with scaphoid fracture

Figure 1. Receiver operating characteristics (ROC) curve for PEM with an 80.8% specificity and 76.2% sensitivity, area under curve (AUC) 0.869 (95%CI 0.769-0.969), cut-off value representing minimal clinically important difference (MCID) of 17.31.

(9). Cronbach's α in patients with distal radius fracture using only Part Two of the questionnaire was 0.94 (8) and in patients after surgical CT decompression was above 0.9 at all time points of the study (7).

Test-retest reliability of PEM-Slo was in the range of good stability level of the patient-reported outcome measure over a time span of 2.3 days on average (range 2-4 days) which is too short for a true health status change in hand and wrist disorders to occur with the ICC value of PEM being 0.874. Our results are in line with the study where the ¤ value of 0.83 for PEM total was obtained (11). The low ICC for item 1 in the PEM score may indicate the uncertainty of participants in interpreting the term "feeling in the hand". A similar finding was described by Dias et al., where the word "feeling" in the item "The feeling in my hand is normal…completely absent" may connote different meanings besides feeling of touch, such as "feeling of stiffness, feeling swollen or feeling not right" (9). They have proposed a more precise explanation of the feeling in item 1, such as "feeling of touch", which could clarify the focus of the question on the assessment of sensation. However, item 1 in the original version of PEM as well as in PEM-Slo remained unchanged.

The construct validity of PEM in our study was established using hypothesis testing with other outcome measures for hand function. The convergent validity was proved with a high correlation of PEM and DASH, and moderate correlation of PEM and grip strength, EQ-5D-5L and parts of EQ-5D-5L, which can be influenced by impaired hand function (self-care, usual activities, pain, anxiety/ depression). The divergent validity was demonstrated with no correlation of PEM with mobility dimension of EQ-5D-5L, the latter assessing problems in walking about where the hand function in our cohort of patients is insignificant. Correlation between PEM and DASH was stronger than the correlation of both DASH and PEM with grip strength, since both are region-specific instruments assessing a similar construct. Similar findings were reported in other studies with strong correlation of PEM and DASH where the values of correlation coefficients were 0.73 (8) and 0.82 (10). Moderate correlations were also found among wrist and finger range of motion and PEM (0.57 and 0.41, respectively) (10). The range of motion was not used in the analysis in our study since it could not be directly compared among participants due to the miscellaneous cohort of patients.

The PEM was established as the most responsive questionnaire to detect change among the three patient-reported outcome measures in our study. This emphasizes the importance of using a region-specific or disease-specific questionnaire. EQ-5D-5L as a general HRQol measure showed only moderate responsiveness for patients with hand and wrist impairments, whereas PEM and DASH were highly responsive, with PEM's SRM and ES

values surpassing those of DASH. A better responsiveness of PEM in comparison to DASH was also proved in CT syndrome patients (ES for PEM 0.97 and ES for DASH 0.49) (6) and in patients with scaphoid fracture, where the values of ES and SRM for PEM were similar to ours (SRM -1.46 and -1.47, and ES -1.12 and -1.1 at two different time points) (9).

When interpreting the scores of the outcome measures it is important to know the change in score that may occur by chance due to measurement error. The MDC95 for PEM-Slo in our study was 18.01. This represents a lower limit above which the observed change in a scale can be interpreted as a true change. The MCID of a scale corresponds to the value where the observed change represents the meaningful improvement in a patient. The MCID in our study calculated with the ROC curve method was 17.31 with the AUC value of 0.869 indicative of excellent discriminative power. The MDC95 and MCID values for PEM-Slo in our cohort of patients are very close to each other, with MDC95 even slightly surpassing the MCID value. The MDC95 was calculated with the distribution-based method using SEM and the calculations were made on an entire sample (n=51) without using the external anchor to dichotomise patients into improvers and non-improvers. The values obtained with anchor-based methods are more reliable and should be used to establish the MCID in outcome measures (28). However, since both values in our study are very close (MDC95=18.01, MCID=17.31) we propose the change in PEM score above 18 to be considered as a clinically important improvement. To the best of our knowledge there is only one conference paper about the interpretability of PEM reporting the MCID=3 in patients with Dupuytren's contracture using the ROC (AUC) method (29). However, no direct comparison to this study can be made since it is important to bear in mind that the MCID can vary among different cohorts of patients (30) and methods used to estimate its value (28).

The advantage of PEM over other region-specific patient-reported outcome measures such as DASH or Patient Rated Wrist Evaluation (PRWE) is its unambiguous focus of questions on the impaired hand. DASH evaluates the capacity to carry out a task irrespective of which hand is used to perform the task (12). The PRWE is clearly focused on the affected hand (31), yet some questions are assessing the function of the dominant hand (e. g. "cutting meat using my affected hand" or "using bathroom tissue with the affected hand") where patients can have difficulties in choosing the appropriate answer if they never use the affected hand for such tasks. Items of PEM in Part Two from 1-6 and 9-10, and items 2 and 3 in Part Three are focused on the assessment of one hand whereas items 7 and 8 about function in Part Two and item 1 in Part Three could assess both hands simultaneously. It may be reasonable to assess separately each side with bilateral

disorders, however the items that assess function would probably have lower values with bilateral hand and wrist problems. It is therefore reasonable to limit the use of PEM in unilateral wrist and hand disorders. We believe that none of the items in PEM is biased by the handedness of a patient.

The use of PEM in literature is, in our opinion, inconsistent. There are only numbers reported without any clear explanation whether this is the calculated sum of scores, or the number represents the score on a scale from 0-100 (32). Sometimes only Part Two of PEM is used in the analysis instead of Part Two and Part Three as suggested by the developers (8, 33). Our study clarifies the use of the scale as suggested by the developers. It appeals to a more consistent use of PEM using Part Two and Part Three in the calculation of the final score with the formula that transforms the result on a scale from 0-100, which represents a more clear-cut presentation of the results.

The major shortcoming of our study is a relatively small sample size. The total number of 51 patients in our study falls in the lower limit of necessary subjects according to Terwee et al. (19) when assessing the construct validity of the instrument in relation to other measures. A larger sample size would also be needed to check for the unidimensional structure of the instrument with confirmatory factor analysis, where a sample size of 150 to 1,000 subjects is recommended in the literature (34). In the previous validation studies of PEM the sample sizes were in the range from 32 to 200 patients (number of participants were as follows: 32 (6), 35 (11), 50 (7), 80 (9), 100 (10), 200 (8)). According to Bacchetti, who recommends choosing a sample size that was used "in the past for similar or analogous studies" (35) we aimed for a sample size of 100 patients. The study was conducted in the time of COVID where the number of patients in our rehabilitation ward was reduced to a minimum (36), and the study time was limited which resulted in a smaller sample size than we aimed for, with the total of 51 patients. Another shortcoming of our study related again to the small sample size is calculation of the internal consistency expressed with Cronbach's alpha coefficient based only on the assumption of unidimensionality of the scale, since the confirmatory factor analysis to check the structure of the scale could not be performed due to the small sample size. The other shortcoming is the heterogeneous cohort of patients with hand disorders, especially when determining the MCID where the values depend on the characteristics of the examined cohort.

5 CONCLUSION

The PEM-Slo is a reliable, valid and highly responsive patient-reported outcome measure for patients with hand and wrist disorders. It has excellent internal consistency and good test-retest reliability. Its construct validity was demonstrated in correlation to DASH, EQ-5D-5L and grip strength. It has no floor or ceiling effect. Our study adds new knowledge about the interpretability of PEM with values of the MDC95 and MCID. It also precisely describes the calculation method for the PEM score, using Part Two and Part Three and expressing it as a value from 0-100. We believe the PEM-Slo is ready to be used in the clinical and research environment. Further research is needed about the interpretability of PEM on more homogenous cohorts where the MDC and MCID values could differ from the ones obtained in our study.

ETHICAL APPROVAL

Ethical approval was obtained from the Ethical Committee of University Medical Centre Maribor on 11 May 2020 (document number UKC-MB-KME-20/20).

AVAILABILITY OF DATA AND MATERIALS

Data sharing is not applicable to this article as no datasets were generated during the current study.

REFERENCES

1. Alderman AK, Chung KC. Measuring outcomes in hand surgery. Clin Plast Surg. 2008;35(2):239-250. doi: 10.1016/j.cps.2007.10.001.

2. Macey AC, Burke FD, Abbott K, Barton NJ, Bradbury E, Bradley A, et al. Outcomes of hand surgery. J Hand Surg Br. 1995 Dec 29;20(6):841-855. doi: 10.1016/s0266-7681(95)80059-x.

3. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000 Dec;81:15-20. doi: 10.1053/apmr.2000.20619.

4. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000 Dec 15;25(24):3186-3191. doi: 10.1097/00007632-200012150-00014.

5. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(7):737-745. doi: 10.1016/j.jclinepi.2010.02.006.

6. Hobby JL, Watts C, Elliot D. Validity and responsiveness of the Patient Evaluation Measure as an outcome measure for carpal tunnel syndrome. J Hand Surg Br. 2005 Aug;30(4):350-354. doi: 10.1016/j.jhsb.2005.03.009.

7. Zyluk A, Piotuch B. A comparison of DASH, PEM and levine questionnaires in outcome measurement of carpal tunnel release. Handchir Mikrochir Plast Chir. 2011;43(3):162-166. doi: 10.1055/s-0031-1273686.

8. Forward DP, Sithole JS, Davis TRC. The internal consistency and validity of the patient evaluation measure for outcomes assessment in distal radius fractures. J Hand Surg Eur. 2007;32(3):262-267. doi: 10.1016/J.JHSB.2007.01.010.

9. Dias JJ, Bhowal B, Wildin CJ, Thompson JR. Assessing the outcome of disorders of the hand. Is the patient evaluation measure reliable, valid, responsive and without bias? J Bone Joint Surg Br. 2001;83(2):235-240. doi: 10.1302/0301-620x.83b2.10838.

10. Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability, validity and ease of use of the patient evaluation measure, the disabilities of the arm, shoulder and hand and the Michigan hand outcome measure. J Hand Surg Eur. 2008;33(1):9-17. doi: 10.1177/1753193407087121.

11. Sharma R, Dias JJ. Validity and reliability of three generic outcome measures for hand disorders. J Hand Surg Br. 2000;25 B(6):593-600. doi: 10.1054/jhsb.2000.0398.

12. Hudak PL, Amadio PC, Bombardier C, Beaton D, Cole D, Davis A, et al. Development of an upper extremity outcome measure: The DASH (disabilities of the arm, shoulder, and head). Am J Ind Med. 1996 Jun;29(6):602-608. doi: 10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L.

13. Novak E, Lavrić M, Semprimožnik K. Institute for Work and Health [Internet]. 2006 [cited 2021 Jul 4]. Slovene DASH. Available from: https://dash.iwh.on.ca/sites/dash/public/translations/DASH_Slovene.pdf

14. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, present and future. Appl Health Econ Health Policy. 2017 Apr 13;15(2):127-137. doi: 10.1007/s40258-017-0310-5.

15. Prevolnik Rupel V, Ogorevc M. Crosswalk EQ-5D-5L value set for Slovenia. Zdr Varst. 2020 Sep;59(3):189-194. doi: 10.2478/sjph-2020-0024.

16. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: A review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-170. doi: 10.1179/jmt.2009.17.3.163.

17. Van Der Roer N, Ostelo RWJG, Bekkering GE, Van Tulder MW, De Vet HCW. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine. 2006 Mar 1;31(5):578-582. doi: 10.1097/01.brs.0000201293.57439.47.

18. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S. Grip and pinch strength: Normative data for adults. Arch Phys Med Rehabil. 1985;66(2):69-74.

19. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Qual Life Res. 2012 May;21(4):651-657. doi: 10.1007/s11136-011-9960-1.

20. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. doi: 10.1016/j.jclinepi.2006.03.012.

21. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016 Jun;15(2):155-163. doi: 10.1016/j.jcm.2016.02.012.

22. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge: Cambridge University Press; 2011. 91, 243 p.

23. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. 2nd Ed. Boston: Houghton Mifflin; 1988.

24. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol. 1997 Jan;50(1):79-93. doi: 10.1016/s0895-4356(96)00296-x.

25. Cohen J. Statistical power analysis for the behavioral sciences. 2nd Ed. New York: Lawrence Erlbaum Associates; 1988. 24-26 p.

26. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr. 2011 Apr;48(4):277-287. doi: 10.1007/s13312-011-0055-4.

27. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978 Oct;8(4):283-298. doi: 10.1016/s0001-2998(78)80014-2.

28. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol. 2010 Jan;63(1):28-36. doi: 10.1016/j.jclinepi.2009.01.024.

29. MCID for the patient evaluation measure as a patient rated outcome measure for Dupuytren's contracture. In: Dias J, Sayeeed L, Bhowal B, editors. The British Society for Surgery of the Hand, Autumn Scientific Meeting [Internet]. 2015 [cited 2021 Jul 4]. Available from: https://www.bssh.ac.uk/_userfiles/pages/files/professionals/A%202015%20Programme%20Final.pdf

30. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: Large variation among populations and methods. J Clin Epidemiol. 2010 May;63(5):524-534. doi: 10.1016/j.jclinepi.2009.08.010.

31. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: A reliable and valid measurement tool. J Orthop Trauma. 1998 Nov;12(8):577-586. doi: 10.1097/00005131-199811000-00009.

32. Pinder EM, Chee KG, Hayton M, Murali SR, Talwalkar SC, Trail IA. Survivorship of revision wrist replacement. J Wrist Surg. 2018 Feb;7(1):18-23. doi: 10.1055/s-0037-1603320.

33. Lane JCE, Rodrigues JN, Furniss D, Burn E, Poulter R, Gardiner MD. Basal thumb osteoarthritis surgery improves health state utility irrespective of technique: A study of UK Hand Registry data. J Hand Surg Eur. 2020 Jun;45(5):436-442. doi: 10.1177/1753193420909753.

34. Anthoine E, Moret L, Regnault A, Sbille V, Hardouin JB. Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. 2014 Dec 9;12:176. doi: 10.1186/s12955-014-0176-2.

35. Bacchetti P. Current sample size conventions: Flaws, harms, and alternatives. BMC Med. 2010 Mar 22;8:17. doi: 10.1186/1741-7015-8-17.

36. Jesenšek Papež B, Šošić L, Bojnec V. The consequences of COVID-19 outbreak on outpatient rehabilitation services: A single-center experience in Slovenia. Eur J Phys Rehabil Med. 2021 Jun;57(3):451-457. doi: 10.23736/S1973-9087.21.06678-8.