Original article
Translation and validation of the PREE (Patient Rated Elbow Evaluation) to a French version

G. Beauchemin\textsuperscript{a}, J.C. MacDermid\textsuperscript{b}, K. Bourduas\textsuperscript{a}, M.-F. Poirier\textsuperscript{c}, C. Gaudelli\textsuperscript{a, c}, D.M. Rouleau\textsuperscript{a, c, *}

\textsuperscript{a} Université de Montréal, Montréal, Canada

\textsuperscript{b} McMaster University, Hamilton, Canada

\textsuperscript{c} Hôpital du Sacré-Cœur de Montréal, Montréal, Canada

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 18 June 2013
Accepted 12 March 2015

\textbf{Keywords:}
Elbow Questionnaire Outcome assessment Reliability Quality of life

\textbf{A B S T R A C T}

\textit{Background:} Only a few outcome measures specific to elbow pathology and the assessment of their impacts on function are valid and reliable when used in French speaking populations. The English version of the Patient Rated Elbow Evaluation (PREE) was determined to be an optimal candidate for translation.

\textit{Hypothesis:} A French version of the PREE (PREE-Fr) will be generated and compared to its original version in terms of reliability and responsiveness.

\textit{Materials and methods:} The PREE was translated following the guidelines of the American Academy of Orthopedic Surgeons. Patients with a variety of elbow pathologies completed the French version of the PREE (PREE-Fr), the Quick Disabilities of the Arm, Shoulder and Hand (QuickDASH) and the Mayo Elbow Performance Score (MEPS) on three different occasions. The test-retest reliability of the PREE-Fr was calculated using questionnaires that were filled out with a one-week interval between them. The responsiveness was assessed using questionnaires filled out six months after treatment.

\textit{Results:} A French version of the PREE was generated. Data gathered from 54 patients yielded an intra-class correlation coefficient for reliability of 0.89 (IC\textsubscript{95}: 0.79–0.94) for the PREE-Fr. For construct validity, using the Pearson correlation coefficient, we obtained excellent correlation between the PREE-Fr and QuickDASH at day one, one week and six months (0.89–0.96) while that between the PREE and MEPS was good to excellent (0.70–0.95). Responsiveness of the PREE-Fr was assessed and yielded a standardized response mean of 1.03, meaning that a large change was recorded between day one and six months.

\textit{Discussion:} The PREE-Fr should be considered in French speaking populations for patients with elbow pathology, whether it is for research or evaluation purposes as it is valid, reliable and responsive to change.

\textit{Level of evidence:} II (questionnaire validation).

© 2015 Published by Elsevier Masson SAS.

1. Introduction

Outcome measures following a medical or surgical treatment are valuable tools for the clinician and researcher. They help to evaluate the quality or efficacy of the intervention and provide a snapshot of the impact it has on the quality of life of the patient. In the past few decades, clinical research has seen a shift from gathering objective measurements using specialized instrumentation towards acknowledging the subjective perspective of the patient [1]. This led to the development of two types of questionnaires: patient-rated and physician-rated questionnaires.

When evaluating elbow pathologies, the clinician is provided with a fair amount of tools [2]. They have traditionally consisted of objective measurements such as range of motion, strength and stability as well as subjective ones, namely pain and function. However, only a few of these have been validated for reliability and responsiveness [3]. The reliability of a questionnaire refers to its ability to give the same results when given to the same patient after a short interval (one week), when no significant change in the medical condition has occurred. The responsiveness refers to the ability of the outcome measure to detect a clinically significant alteration in the underlying condition (such as a surgical intervention or a relapse) [4]. Adding to the difficulties in finding an appropriate outcome measure is that there are very few that are
validated for use in French speaking populations [5,6]. Notably, the Société orthopédique de l’Ouest has published a French language score to evaluate the outcomes in patients with fractures of the radial head [7]. This score is comprised of three different sections: pain, range of motion and strength. It is scored on a scale of eleven which then gives results classified as: “excellent”, “good”, “mediocre” and “poor”. While this scale is convenient and easy to use, it was designed for use in one pathology only and thus cannot be generalized for use in patients with other elbow pathologies.

Many of the questionnaires used to evaluate patients with elbow pathologies found in the literature were designed for the assessment of the entire upper limb [8]. The Patient Rated Elbow Evaluation (PREE), however, stands out as having been specifically designed for all elbow pathologies. It has been shown to have good psychometric properties [9]. The PREE is divided into two brief sections addressing both pain and functional impairment in performing activities of daily living (ADLs) and instrumental ADLs (IADLs). It is composed of twenty questions rating the intensity of symptoms and the extent of functional impact using a Likert scale from zero to ten.

The PREE has been previously compared and validated with the American Shoulder and Elbow Surgeons Elbow questionnaire (ASES-e); the Disabilities of Arm, Shoulder and Hand questionnaire (DASH) and the Short-Form 36 Health Survey questionnaire (SF-36). The test-retest reliability was shown to be very good for the pain sub-scale elements (ICC: 0.74–0.87) and fair to very good for the function subscale (ICC: 0.60–0.84). Construct validity was also determined by calculating the Pearson correlations between the ASES and PREE and was fair to very good depending on subscales used [8]. Thus, these are the reasons the PREE was chosen for translation to the French language. Further supporting this is the fact that it has been successfully translated into German and adapted for different cultures [10,11].

The purpose of this study is therefore to translate and adapt the PREE to the French language and culture while maintaining its good psychometric properties. Our hypothesis is that its reliability and responsiveness will be comparable to its English counterpart.

2. Methods

2.1. Translation and cultural adaptation

The method used for translation is based on the guidelines developed by Beaton and colleagues [12], and recommended by the American Academy of Orthopedic Surgeons (AAOS) (Table 1). The importance of this process lies in the observation that self-administered questionnaires need to be linguistically translated but also culturally adapted if they are to preserve content validity as some activities, while they may commonly practiced by the speakers of the initial language, may be irrelevant to speakers of the new language. This holds true both for speakers of the new language in the same geographical location of the initial language and for those in another country.

In the first stage, two independent native French-speaking translators translated the questionnaire from English to French to produce two versions (T1 and T2). The senior author and a research nurse compared these two different versions. Any discrepancies were resolved in order to produce a common version (T1,2). Following this, two different translators, with English as their first language, translated this version (T1,2) back to English producing BT1 and BT2. These English versions were compared to the initial English version. A review committee formed by an orthopedic surgeon and a registered nurse reviewed all the versions, as well as any notes produced by the translators to synthesize a pre-final version (PREE-Fr-prelim). This pre-final version was distributed to fourteen subjects who agreed to test the questionnaire. They were asked to comment on the items in order to explore their understanding of the meaning of the issues discussed. The final PREE-Fr (Fig. 1) was then validated, as outlined below.

2.2. Questionnaire validation

This prospective study was carried out in a Level 1 trauma. This study was approved by the institutional ethics committee and informed consent was obtained from all the participants. A sample size of fifty patients was chosen. This number was arrived at by reviewing the original AAOS guidelines, a follow-up paper by the same authors as well as the IQLA guideline [13]. Additionally, a literature review was performed looking at previous studies validating upper extremity questionnaires – fifty patients was the median number of subjects recruited.

Patients suffering from elbow pathology admitted between February 2011 and August 2012 were asked to participate. They were included if they were eighteen years or older, had a chronic or sub-acute elbow pathology requiring treatment, French was their first language, and they provided informed consent and accepted to fill out the questionnaires on three different occasions; at the time of initial presentation (Time 1), one week later (Time 2) and six months after treatment (Time 3). Patients were excluded if they were unwilling or unable to answer the questionnaires, were illiterate or were not native French speakers.

2.3. Questionnaires

At the time of initial consultation, demographic information and information pertaining to diagnosis and proposed treatment was gathered. Subjects also completed the first series of questionnaires. These consisted of the PREE-Fr (Fig. 1), the Mayo Elbow Performance Score (MEPS) [12,14] and the F-QuickDASH-D/S [15,16]. The MEPS is a variant of the Mayo Elbow Performance Index (MEPI), one of the most widely used validated elbow evaluation scoring systems. It exists in three different versions. The one we used is made up of eight items organized into four categories and evaluates

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Description of the translation and cross-cultural adaptation process.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td>Process</td>
</tr>
<tr>
<td>1</td>
<td>Translation</td>
</tr>
<tr>
<td>2</td>
<td>Synthesis</td>
</tr>
<tr>
<td>3</td>
<td>Back Translation</td>
</tr>
<tr>
<td>4</td>
<td>Expert Committee Review</td>
</tr>
<tr>
<td>5</td>
<td>Pre-testing</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from the 1998 AAOS recommendations.
Les questions ci-dessous permettent de mesurer les difficultés causées par votre coude au cours de la dernière semaine. Vous devez évaluer les symptômes ressentis au coude en moyenne au cours de la dernière semaine sur une échelle de 0 (aucune douleur) à 10 (douleur maximale). Si vous n'avez pas pratiqué l'activité en question au cours de la dernière semaine, visualisez faire une estimation de la difficulté qu'elle pourrait présenter.

1. DOULEUR
Évaluez le niveau de douleur moyen de votre coude durant la semaine dernière en encerclant la chiffe qui décrit le mieux votre douleur sur une échelle de 0 à 10. Un score (0) veut dire que vous n'avez ressenti aucune douleur et un dix (10) représente la pire douleur que vous n'ayez jamais éprouvée.

EVALUER LE DEGRÉ DE DIFFICULTÉ A

<table>
<thead>
<tr>
<th>Échelle de 0 à 10</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aucune</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Pire douleur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

1. Quand elle est maximale
2. Au repos
3. En soulevant un objet lourd
4. En effectuant une tâche avec mouvements répétitifs du coude

5. À quels fréquence avez-vous de la douleur

6. À quelle partie de votre coude

6. Gaffer mes cheveux
7. Hanger avec une fourchette ou une cuillère
8. Tenir un objet lourd
9. Utiliser mon bras pour me lever d’une chaise
10. Lancer un objet, comme une balle de tennis
11. Utiliser un téléphone
12. Soulager le devant de ma chemise
13. Laver mes assiettes assis
14. Changer mes souliers
15. Tourner la poignée et ouvrir la porte
16. Transporter un objet de 10 lbs (5 kg) avec mon bras le long du corps

7. ACTIVITÉS HABITUELLES
Évaluez le degré de difficulté éprouvé en effectuant vos activités habituelles dans chacun des domaines suivants : l’habileté à effectuer une tâche, la visible de qualité, la capacité à effectuer une tâche et le degré de satisfaction avec la tâche. Un score (0) veut dire que vous n’avez pas d’activité et un dix (10) veut dire que c’était tellement difficile que vous n’avez pas réussi à faire aucune activité habituelle.

EVALUER LE DEGRÉ DE DIFFICULTÉ À

<table>
<thead>
<tr>
<th>Échelle de 0 à 10</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aucune</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Pays d’origine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

1. Sports personnels (chasse, sport
2. Tâches ménagères (nettoyage, manger)
3. Employé (tâche typique de travail d’activité
4. Activités de loisirs

COMMENTAIRES :

To measure the responsiveness of the PREE-Fr, we computed a standardized response mean (SRM), which is calculated as the ratio of the mean change to the standard deviation of the change scores. It is one of the most common indices for measuring change [11] and can be interpreted like any effect size, using Cohen's criteria [13]. Again, the PREE-Fr was compared to the MEPS and F-QuickDASH-D/S questionnaires. Statistical analysis was performed using SPSS 19.0 for Windows (Chicago, Illinois).

3. Results

3.1 Translation

The two translators produced two independent copies (T1 and T2), which were quite similar. A preliminary (T1.2) questionnaire was synthesized by the committee. Two back translations (BT1 and BT2) were produced by two other translators. These were compared to the original, English questionnaire as well as the French translation. A preliminary French translation (PREE-Fr-prelim) was created by the research committee.

Fourteen patients (five men, nine women; age range twenty to sixty) were recruited for the initial review of the questionnaire. The subjects did not have much issues with the questionnaire other than for one question. The initial wording of the French version of item ten was: “Transporter un objet de 10 lbs (5 kg) avec mon bras le long du corps”. This was deemed to be ambiguous by the test subjects and was changed to: “Transporter un objet de 10 lbs (5 kg) avec mon bras le long du corps” which better represents the intended action.

Moreover, the English version showed a sample scale above each section. When translated to French, the translation committee members initially circled the number five on this sample scale to function and pain as experienced by the patient as well as stability and range of motion, requiring an experienced evaluator with a goniometer [2]. The F-QuickDASH-D/S is a validated French translation of the QuickDASH, which is self-administered questionnaire with eleven items evaluating the entirety of the affected upper limb.

At this appointment, patients also received a package containing the same three questionnaires to be filled out a week later (Time 2). Instructions were given and a postage-paid envelope was included for the return of the forms to the research assistant. This second questionnaire was used to determine the reliability of the PREE-Fr.

Patients were then seen six months after receiving treatment for their elbow pathology (Time 3). They underwent a clinical and radiographic assessment and filled out the same three questionnaires. This was undertaken to determine the responsiveness, or the ability to measure the change in symptomatology of the PREE-Fr.

2.4 Statistical analysis

In order to better define the demographics in the sample studied, descriptive statistics were computed. The mean, range and percentsages were obtained for age, sex, hand dominance, diagnosis and treatment. To better assess the content validity of the PREE-Fr, Pearson’s Correlations were used to compare the results between the tools. To compare them at each step in case management, this analysis was performed for the data gathered at Time 1, Time 2 and Time 3. Content validity is the degree to which the items in an instrument adequately reflect the content domain being measured.

The reliability of the PREE-Fr compared to the MEPS and the F-QuickDASH-D/S was determined using Intra-Class Correlations (ICC). The reliability of each individual questions was also assessed using the ICC.
ensure the instructions were clear. However, the subjects felt this was leading. These circles were removed in the final version.

### 3.2. Validation of the PREE-Fr

A total of fifty-four (n = 54) patients were recruited. This sample included 21 women (38.9%) and 33 males (61.1%) with a mean age of 46 (18–75). Of them, 48 were right-handed (88.9%) and 5 were left-handed (9.3%) while only one (1.9%) was ambidextrous. The most common reason for consultation was for sequelae of trauma and most commonly, the treatment was non-surgical. The reasons for consultation and treatments received are shown in Fig. 2.

Fifty patients adequately filled out the Time 1 questionnaire, thirty-seven did so for the Time 2 questionnaire and twenty-nine completed the six-month follow-up.

There are multiple reasons for patients lost to follow-up: some lived far from the tertiary trauma center and chose to get follow-up care closer to home, while others did not attend their follow-up appointments.

### 3.3. Reliability

The reliability of individual items for the PREE-Fr is reported in Table 2. The majority of questions showed either excellent or strong agreement; the remaining two showed moderate agreement. The strongest correlation was observed for the total score (ICC = 0.89, CI95% = 0.79–0.94), which represents and excellent correlation. The ICC of total scores were computed for the other two questionnaires and showed similar results. The QuickDASH and MEPS had excellent overall agreement with respective ICC of 0.91 and 0.87 and CI95% of 0.83–0.95 and 0.76–0.93 (Table 3).

### 3.4. Responsiveness

The SRM between each test filled at day one and six months are shown in Table 3. The most responsive was the PREE-Fr, followed by the QuickDASH and the MEPS (1.06, 0.86 and 0.72). When compared to one another, no test was significantly different from the other two in terms of responsiveness.

### Table 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Worded as the English PREE</th>
<th>Reliability (95% CI)</th>
<th>SRM 1–3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>When it is at its worst</td>
<td>0.82 (0.68–0.90)</td>
<td>1.11</td>
</tr>
<tr>
<td>2</td>
<td>At rest</td>
<td>0.79 (0.63–0.88)</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>When lifting a heavy object</td>
<td>0.83 (0.70–0.91)</td>
<td>0.72</td>
</tr>
<tr>
<td>4</td>
<td>When doing a task with repeated elbow movement</td>
<td>0.69 (0.48–0.83)</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>How often do you have pain?</td>
<td>0.81 (0.66–0.90)</td>
<td>0.73</td>
</tr>
<tr>
<td>Function</td>
<td>A. Specific Activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Comb my hair</td>
<td>0.88 (0.77–0.93)</td>
<td>0.70</td>
</tr>
<tr>
<td>7</td>
<td>Eat with a fork or spoon</td>
<td>0.85 (0.73–0.92)</td>
<td>0.49</td>
</tr>
<tr>
<td>8</td>
<td>Pull a heavy object</td>
<td>0.76 (0.58–0.87)</td>
<td>0.73</td>
</tr>
<tr>
<td>9</td>
<td>Use my arm to rise from a chair</td>
<td>0.65 (0.42–0.80)</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>Throw a small object, such as a tennis ball</td>
<td>0.82 (0.69–0.90)</td>
<td>0.40</td>
</tr>
<tr>
<td>11</td>
<td>Use a telephone</td>
<td>0.75 (0.56–0.86)</td>
<td>0.84</td>
</tr>
<tr>
<td>12</td>
<td>Do up buttons on the front of my shirt</td>
<td>0.80 (0.65–0.89)</td>
<td>0.65</td>
</tr>
<tr>
<td>13</td>
<td>Wash my opposite armpit</td>
<td>0.79 (0.63–0.88)</td>
<td>0.71</td>
</tr>
<tr>
<td>14</td>
<td>Tie my shoe</td>
<td>0.62 (0.38–0.78)</td>
<td>0.69</td>
</tr>
<tr>
<td>15</td>
<td>Turn the doorknob and open a door</td>
<td>0.75 (0.57–0.86)</td>
<td>0.66</td>
</tr>
<tr>
<td>16</td>
<td>Carry a 10 lb object with my arm at my side</td>
<td>0.69 (0.48–0.82)</td>
<td>0.29</td>
</tr>
<tr>
<td>Function</td>
<td>B. Usual Activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Personal care activities (dressing, washing)</td>
<td>0.79 (0.63–0.89)</td>
<td>0.71</td>
</tr>
<tr>
<td>18</td>
<td>Household work (cleaning, maintenance)</td>
<td>0.80 (0.65–0.89)</td>
<td>0.81</td>
</tr>
<tr>
<td>19</td>
<td>Work (your job or everyday work)</td>
<td>0.88 (0.78–0.94)</td>
<td>0.72</td>
</tr>
<tr>
<td>20</td>
<td>Recreational activities</td>
<td>0.82 (0.68–0.90)</td>
<td>0.85</td>
</tr>
</tbody>
</table>

The reliability was assessed using ICC comparing responses to questionnaires filled out at Time 1 and Time 2. The responsiveness was assessed calculating SRMs from questionnaires filled out at Time 1 and Time 3.
Table 3
The Reliability and responsiveness of the PREE-Fr, F-QuickDASH-D/S and MEPS.

<table>
<thead>
<tr>
<th>Functional outcome measure</th>
<th>Reliability (95% CI)</th>
<th>Responsiveness (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREE-Fr</td>
<td>0.89 (0.79–0.94)</td>
<td>1.06</td>
</tr>
<tr>
<td>F-QuickDASH-D/S</td>
<td>0.91 (0.83–0.95)</td>
<td>0.86</td>
</tr>
<tr>
<td>MEPS</td>
<td>0.87 (0.76–0.93)</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The reliability was calculated by determining the Interclass Correlation (ICC) of two sets of questionnaires filled out within a week of each other (Time 1 and Time 2). All three questionnaires showed excellent reliability. The responsiveness was calculated using Standardized Response Means (SRM). This was calculated using the first questionnaire filled out (Time 1) and that filled out six months after treatment (Time 3). All three questionnaires showed good responsiveness.

4. Discussion

This study was successful in creating a French-language adaptation of the PREE. This will be of direct benefit to both clinicians and researchers. Clinicians will be able to use this tool to track their outcomes in French-speaking populations. Researchers and the scientific community will benefit as this allows the same outcome measure to be used across different languages. Because it is validated, it is considered equivalent to the original PREE.

The strength of our study is that, the reliability and responsiveness of the total PREE-Fr score was excellent (ICC: 0.89 for reliability and SRM: 1.05 for responsiveness). When considering individual items, when compared to the original version, the PREE-Fr had ICC: 0.62–0.88, which again is good. Section scores were not computed nor analyzed. We did not expect to observe trends in pain or function ratings since our study included patients with multiple pathologies.

Another strength of this study is that we calculated the external validity of the PREE-Fr with other outcome measures. The correlation between the PREE-Fr and the French language version of the QuickDASH was excellent. This is likely because both outcome measures evaluate pain and function and are self-reported. The correlation between the PREE-Fr and the MEPS was good to excellent; the slightly lower scores are likely explained by the fact that the MEPS has sections on elbow range of motion and stability, as measured by the clinician so is not entirely a patient-reported outcome measure.

One limitation to our study is that the AAOS guidelines suggest using a sample of thirty to forty patients for the pilot study whereas only fourteen were recruited in our study. This is the step where these patients are given the questionnaire and asked about their understanding of the individual questions, the ease of competing the questionnaire and whether the questions are relevant to them (the cultural adaptation portion of the translation). However, this step does not address the construct validity or reliability of the questionnaire, but rather to explore the meaning of individual items. Thus, in order to facilitate the recruitment of sufficient patients for the validation, once the responses to the pilot study questions became redundant, recruitment for this portion was stopped.

In conclusion, our results support the use of the PREE-Fr for the evaluation of elbow function and pain in French speaking patients, extending the list of potential forms to be used both in clinic and in research projects. Furthermore, the fact that it is self reported makes it attractive to the clinician as a means of gathering valuable information about the daily impact of the pathology on patient function in a time efficient manner. The PREE-Fr is valid and reliable, as well as responsive to change, when measuring pain and function in elbow pathologies. We therefore strongly recommend its implementation in rehabilitation and medical clinics as well as research protocols.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

The authors wish to thank Julie Fournier, Stéphanie Boisvert and Jean Paquet for their valuable help with coordination, data gathering and statistical analysis.

References