The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: A systematic review.

Christophe Eechaute, PT, MT, the Physical Therapy Department of the Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium.
Email address: christophe.eechaute@belgacom.net, Corresponding author

Peter Vaes, PHD, PT, MT, the Physical Therapy Department of the Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium.
Email address: pvaes@vub.ac.be

Lieve Van Aerschot, PT, MT, the Physical Therapy Department of the Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium
Email address: lievevanaerschot@hotmail.com

Sara Asman, PT, MT, the Physical Therapy Department of the Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium
Email address: asmansara@hotmail.com

William Duquet, PHD, the Human Biometry and Biomechanics Department of the Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium
Email address: wduquet@vub.ac.be
Abstract.

Background: The aim of the study was to systematically review the clinimetric qualities of patient-assessed instruments designed for patients with chronic ankle instability.

Methods: A computerized literature search of Medline, Embase, Cinahl, Web of Science, Sport Discus and the Cochrane Controlled Trial Register was performed to identify eligible instruments. Two reviewers, using a criterium list, independently evaluated the clinimetric qualities of the selected instruments. To assess the inter-observer reliability of the selection procedure and the clinimetric evaluation of the selected instruments, modified kappa coefficients were calculated.

Results: The inter-observer reliability of the selection procedure was excellent (κ= .86). Three instruments met the eligibility criteria: the Ankle Joint Functional Assessment Tool (AJFAT), the Functional Ankle Outcome Score (FAOS) and the Foot and Ankle Disability Index (FADI). The inter-observer reliability of the quality assessment was substantial to excellent (κ between .64 and .88). Test-retest reliability was demonstrated for the FAOS and the FADI but not for the AJFAT. Content validity, construct validity, internal consistency and the presence of floor-and ceiling effects were only assessed for the FAOS. Cronbach’s alpha coefficients for the four subscales were above .90 and ceiling effects were present for all subscales of the FAOS. Responsiveness was only demonstrated for the FADI. For none of the studied instruments a minimal clinical important difference was presented.

Conclusion: The FADI can be considered as the most appropriate, patient-assessed tool to quantify functional disabilities in patients with chronic ankle instability as it demonstrates reliability and responsiveness. However, content validity, internal consistency and an absence of ceiling effects remain to be demonstrated.
Background.

Lateral ankle sprains are very common sports related ankle injuries. Recurrence rates of ankle sprains have been reported to be up to 70% [1]. Fifty five % to 72% of individuals who sustain a lateral ankle sprain have residual symptoms and develop chronic ankle instability [2-4]. The development of chronic ankle instability has been ascribed to different causes like a delayed muscle reflex of stabilizing lower leg muscles, deficits in lower leg muscle strength, deficits in kinaesthesia or an impaired postural control [5-8]. Most studies, investigating the pathomechanism of chronic ankle instability or evaluating treatments, only use outcome measures that are clinician related and on the level of impaired physiological functions. Research studying patients with chronic ankle instability should also assess this health problem from the patient’s perception and document the presence of disabilities, handicaps and a decreased quality of life.

Patient- assessed instruments, like questionnaires, are therefore appropriate tools. But the clinimetric qualities of these instruments should be documented [9]. To our knowledge, no available systematic review identified and evaluated patient- assessed instruments for chronic ankle instability. Therefore, the purpose of this review was to systematically search the literature for patient- assessed instruments used in populations with chronic ankle instability and to evaluate the clinimetric qualities of the studied instruments.

Methods.

Literature search and selection.

For the identification of patient- assessed instruments for chronic ankle instability, the following databases were screened until May 2006: Medline from 1966, Cinahl from 1982,
Embase from 1994, Sport Discus from 1949, the Cochrane Controlled Trial Register from 1966 and Web of Science from 1972.

To retrieve eligible instruments, Medical Subject Headings (Mesh terms) and key words were combined to describe the patient population and the instrument (figure 1). Mesh terms and key words that were used to identify the patient population were: lateral ligament, ankle; ankle joint; joint instability; sprain*; unstable; chronic*; multiple and repetitive; functional and recurren*. Mesh terms and key words that were used to identify the measure were: questionnaire*; scale*; weights and measures; outcome; outcome assessment, score; self-assessment and self-report. Relevant instruments were identified on the basis of title, abstract and the full text version of the retrieved articles. Additionally, reference lists of relevant articles were screened for eligible instruments.

Instruments were included:

- if they were used in articles studying patients with chronic ankle instability.
- if it was a patient-assessed instrument and contained items referring to disabilities (activities), handicaps (participation) and/or quality of life.
- if one or more clinimetric qualities of the instrument were studied in the retrieved articles.

Instruments were excluded:

- if the instrument was not patient-based.
- if the instrument was not published in its full version.
- if the instrument contained only items related to impaired functions.
- if not published in English, French, Dutch, or German.

Based upon these criteria, two reviewers independently selected eligible instruments. Their inter-observer reliability was assessed using the modified kappa coefficient. When disagreement persisted between the two reviewers concerning eligibility of an instrument, a third person (C.E.) was consulted.
Quality assessment.

The clinimetric qualities of the selected instruments were evaluated by means of a checklist used in the review article of Bot et al. [10] who studied the clinimetric qualities of shoulder disability questionnaires. This checklist was partly based upon the review criteria of the Scientific Committee of the medical outcome trust [11]. It contains the following items: content validity, readability, reliability, internal consistency, construct validity, floor-and ceiling effects, responsiveness, interpretability, minimal clinical important difference, administration burden and time to administer (figure 3). To achieve agreement between the ratings of the reviewers, a pilot testing of the checklist was conducted by evaluating the clinimetric qualities of the Western Ontario Shoulder Instability Index [12] until consensus was reached.

Subsequently, the two reviewers independently evaluated the selected instruments. Items could be rated by “+”, “±”, “-” or “?”’. An item was rated “+” when sufficient information was available and bias was unlikely. An item was rated “±” if the available information was unclear or the used method was doubtful. An item was rated “-” if sufficient information was available but the instrument did not met the criteria. An item was rated “?” if no information was available. Modified kappa coefficients were calculated to assess the inter-observer reliability.

If disagreement persisted about the assignment of a score to an item, a third person (C.E.) was consulted to decide about the final rating.

Results.

Selection.

The inter-observer reliability of the selection of the instruments was excellent (κ= .86). The search strategy revealed 939 articles (figure 2). 17 instruments were identified. Of these 17
instruments, 3 met the inclusion criteria: the Ankle Joint Functional Assessment Tool (AJFAT) [13], the Foot and Ankle Disability Index (FADI) [14] and the Foot and Ankle Outcome Score (FAOS) [15]. 14 instruments were excluded:

- For being a generic health measure (the Short Form Health Survey [16]).
- For containing only “pain related” items (the McGill Pain Questionnaire [17]).
- For not being a patient-based measure (the Kaikkonen scale [18]; the Karlsson score [19]; the Debie Score [20]).
- For containing no distinct disability, handicap or quality of life items (the Good Rating Scale [21]; the Sefton Score [22]; the Subjective Functional Rating Scale [23]; the Subjective Grading Scale [24]; the Keller Score [25]; the Tegner Score [26]).
- Because no full version was available in the studied articles (the Weber Score [27]; the Zwipp Score [28]; the Brunner Score [29]).

The information regarding the clinimetric qualities of the Ankle Joint Functional Assessment Tool [13], the Foot and Ankle Outcome Score [14] and the Foot and Ankle Disability Index [15] was retrieved from the original publications.

Description of the studied instruments.

The Foot and Ankle Outcome Score (FAOS) is a 42-item questionnaire divided into 5 subscales: pain, other symptoms, activities of daily living, sport and recreation function, ankle related quality of life. The subscale “pain” contains 9 items, the subscale “other symptoms” 7 items, the subscale “activities of daily living” 17 items, the subscale “sport and recreation function” 5 items and the subscale “foot- and ankle related quality of life” 4 items. Each question can be scored on a 5-point Likert scale (from zero to four) and each of the five subscale scores is calculated as the sum of the items included. Raw scores are then transformed to a zero to 100, worst to best score.
The Ankle Joint Functional Assessment Tool (AJFAT) contains 5 impairments (pain, stiffness, stability, strength, “rolling over”), 4 activity related items (walking on uneven ground, cutting when running, jogging and descending stairs) and 1 overall quality item. Each item has 5 answer options. The best total score of the AJFAT is 40 points, the worst possible 0 points.

The Foot and Ankle Disability Index (FADI) is a 34-item questionnaire divided into two subscales: the Foot and Ankle Disability Index and the Foot and Ankle Disability Index Sport. The Foot and Ankle Disability Index contains 4 pain related items and 22 activity related items. The Foot and Ankle Disability Index Sport contains 8 activity related items. Each question can be scored on a 5-point Likert scale (from zero to four). The FADI and the FADI Sport are scored separately. The FADI has a total score of 104 points and the FADI Sport 32 points. The scores of the FADI and FADI Sport are then transformed into percentages.

Quality assessment.

The inter-observer reliability for the rating of the items of the checklist was substantial to excellent (κ between .64 and .88). Disagreement between the reviewers existed for the items reliability, construct validity, interpretability and administration burden. The clinimetric qualities are more extensively documented for the FAOS than the FADI and the least for the AJFAT (see additional file 1: table 1 with the final rating and description of the clinimetric qualities of the studied instruments).

Clinimetric qualities.

A survey of the final rating and the description of the clinimetric qualities of the studied instruments are presented in table 1.
Content validity.

For the AJFAT or the FADI no information was available whether patients and experts were involved in the selection and reduction process of items. For the development the FAOS, patients were asked to rate the relevance and importance of the items from one (not relevant, not important) to three (very relevant, very important).

Readability.

For none of the studied instruments information on the clarity of the questions for the patients is available.

Reliability.

Test-retest reliability was demonstrated for the FAOS and the FADI. Intra Class Correlation coefficients (ICCs) for the 5 subscales of the FAOS ranged from .70 to .92. ICCs for the FADI and FADI Sport of the chronic unstable group ranged from .84 to .94. The precision of the measurement (standard error of measurement = SEM) was only estimated for the FADI (SEM= 2.6) and the FADI Sport (SEM= 5.3). For the AJFAT, information on test-retest reliability was lacking.

Internal consistency.

The extent to which items in a subscale are inter-correlated was only assessed for the FAOS. Cronbachs’ alpha coefficients for the 5 subscales of the FAOS ranged from .88 (for the “pain” subscale) to .97 (for the “sport and recreation” subscale). For the AJFAT or the FADI, information on internal consistency was lacking.

Floor- ceiling effects.

Ceiling effects, the failure to demonstrate an increased score in patients who clinically improved, were observed for all 5 subscales of the FAOS. 19% of all patients displayed the best possible score for the “foot and ankle related quality of life” scale, 24% for the
“symptoms” scale, 30% for the “sport and recreation function” scale, 34% for the “pain” scale and 44% for the “activities of daily living” scale. For the AJFAT or the FADI, no information on floor- and ceiling effects was available.

Construct validity.
The FAOS was correlated to the Karlsson Score; a clinician-assessed scoring scale for ankle instability [19]. Moderate correlation coefficients (Spearman Rho) were found (r= .58 to .67). For the AJFAT or the FADI, construct validity was not studied in patients with chronic ankle instability.

Responsiveness.
The ability to detect important change of the health status over time was assessed for the AJFAT and the FADI. In the study of Rozzi et al. [13] a significant improvement in AJFAT score of trained patients could be observed after 4 weeks of wobble board training. However, a standard response mean (i.e. mean change in score divided by standard deviation of the change of score) or an effect size (i.e. mean change of scores divided by standard deviation of baseline scores) was not calculated.

With the FADI, a significant difference between pre- and post training scores could be observed in trained subjects with chronic ankle instability. Effect sizes were estimated for the FADI (ES = .52) and the FADI Sport (ES = .71). For the FAOS no information on the responsiveness was available.

Interpretability.
Interpretability was rated positive for the AJFAT and the FADI. Means and standard deviations of AJFAT scores were presented in the study of Rozzi et al. [13]. Trained patients
who demonstrated significant better AJFAT scores also showed a significantly improved postural balance.

Means and standard deviations of both the FADI and the FADI Sport were presented. Based upon the calculated effect sizes the FADI Sport seems to be more sensitive to change over time than the FADI. Also, results of the FADI and the FADI Sport scores show that both subscales can discriminate between healthy subjects and subjects with chronic ankle instability.

For the FAOS, no detailed information is given on the distribution of the scores of the 213 patients being studied.

For none of the studied instruments, information concerning a minimally clinical important difference was presented.

Time to administer and administration burden.

Only for the FAOS the administration time (7 to 10 minutes) was documented. The final score of the AJFAT is just the result of summing up the different item scores. For the FAOS, the subscale scores are the result of summing up the item scores belonging to that subscale. The raw scores of these subscales are transformed into a 0 to 100 scale.

The scores on the items of the FADI and the FADI Sport are summed up separately and are than transformed into percentages.

Discussion.

There is no gold standard to evaluate the clinimetric qualities of patient- assessed instruments and hence the criterion list that was used can be disputed. This checklist was chosen for its quality of operationalization.
The inter-observer reliability of the quality assessment of the selected measures was substantial to excellent. Disagreement was mostly caused by reading errors. The third reviewer was not consulted for making a final decision about the rating of the items. Many instruments have been used to screen patients with chronic ankle instability but these are not patient-assessed and/or do not contain distinct disability, handicap or quality of life items. Patient-assessed instruments should at least demonstrate validity, reliability and responsiveness before considering them to be useful in clinical practice. With respect to content- and construct validity, only the FAOS has been validated. It is remarkable that the FAOS and the AJFAT are developed to be multidimensional, while the FADI emphasizes on the assessment of functional limitations.

There is no strict cut-off point to decide whether an instrument is reliable or not. It has been stated that the magnitude of the correlation coefficient of a measurement tool should at least be .70 when studying groups of patients and exceed .90 when evaluating individuals [9, 30]. The FAOS and the FADI met this criterion. However, the degree of agreement between the two test sessions for each single item of these measures was not reported.

When considering the internal consistency that was assessed for the FAOS, the Cronbach alpha coefficient for 4 subscales was above .90. This makes it likely that there is some redundancy among items within the subscales [9]. Construct validity was only assessed for the FAOS. The correlation of the five subscales of the FAOS with the total Karlsson score was moderate. Because one could hypothesize the total FAOS score to measure the same theoretical construct as the Karlsson Score, it would have been better to correlate the total scores of these two instruments. Furthermore, correlating the different subscale scores of both instruments would enlighten the construct validity even more.
Floor- and ceiling effects were calculated only for the FAOS. According to the quality list used in this study, with the cut-off point set at 15%, all subscales of the FAOS demonstrated ceiling effects. The choice of cut-off point remains arbitrary. For instance, Barber-Westin et al. [31] studied the presence of floor- or ceiling effects of the Cincinatti knee rating system using a cut-off point set at 33%. The observation of ceiling effects may also be specific for the patient population being studied [9]. The patients that were studied had undergone an anatomical reconstruction of the lateral ankle ligaments on average 12 years prior to the study. It is probable that many of them no longer had ankle problems, which may explain the observation of ceiling effects. Moreover 34% of the same patients also obtained the best possible Karlsson Score. The high percentage of ceiling effects in the FAOS-pain subscale and FAOS- “activities of daily living” subscale may compromise the validity of these subscales.

The subjects with chronic ankle instability that were studied by Hale et al. [15] have at baseline substantially high FADI- and FADI Sport scores. This indicates that these subjects report very few functional limitations. The absence of ceiling effects for the FADI and the FADI Sport have not been established.

Like reliability and validity, responsiveness is another important clinimetric property of an instrument. To establish responsiveness, several estimates (like effect sizes or standardized response means) can be calculated which permits comparison of the sensitivity to change between several instruments. However, only for the FADI an appropriate estimate (effect size) was used. The effect sizes observed for the FADI and the FADI Sport indicate a medium size of change [9].

For none of the studied instruments, a minimal clinical important difference was presented. It would have been interesting if Hale et al. [15] had calculated a minimally detectable change (expressing the magnitude of change that indicates a true improvement). A minimally
detectable change expresses how much change in score should occur before a patient’s condition can be considered as truly improved or deteriorated. In the study of Hale et al. [15], standard errors of measurements (SEM) were presented and minimally detectable changes could have been calculated as these estimates are derived from SEMs [32].

Conclusion.

A systematic computerized literature search of 6 databases revealed 3 patient-assessed instruments for measuring chronic ankle instability: the Ankle Joint Functional Assessment Tool, the Foot and Ankle Disability Index and the Foot and Ankle Outcome Score. The FADI can be considered as the most appropriate, patient-assessed tool to quantify functional disabilities in patients with chronic ankle instability as it demonstrates reliability and responsiveness. However, content validity, internal consistency and an absence of ceiling effects remain to be demonstrated.

Competing interests.

The authors declare that they have no competing interests.

Authors’ contributions.

Christophe Eechaute contributed to the concept and design of the manuscript, to the acquisition, analysis and interpretation of the data and to the draft of the manuscript.

Peter Vaes contributed to the concept and design of the manuscript, to the revision of the manuscript and to the final approval of the manuscript.

Lieve Van Aerschot contributed to the concept and design of the manuscript, the acquisition and interpretation of the data and to the draft of the manuscript.
Sara Asman contributed to the concept and design of the manuscript, the acquisition and interpretation of the data and to the draft of the manuscript.

William Duquet contributed to the acquisition, analysis and interpretation of the data, to the revision of the manuscript and to the final approval of the manuscript.

Acknowledgements.

We want to thank Mrs. Sylvia Van den Heuvel from the Nederlands Paramedisch Instituut in Amersfoort for her substantial contribution to the search actions in databases.

References.


Figure legends.

Figure 1. Survey of the search strategy in the Medline database. *= abbreviation, Mesh= Medical Subject Heading, [TIAB]= in title or abstract.

Figure 2. Flow diagram of the selection procedure of the instruments. AJFAT= the Ankle Joint Functional Assessment Tool; FAOS= the Foot and Ankle Outcome Score; FADI= the Foot and Ankle Disability Index.
Figure 1. Survey of the search strategy in the Medline database.

<table>
<thead>
<tr>
<th></th>
<th>Search Strategy</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>#38</td>
<td>Search #29 AND #38</td>
<td>257</td>
</tr>
<tr>
<td>#37</td>
<td>Search #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36</td>
<td>989004</td>
</tr>
<tr>
<td>#36</td>
<td>Search outcome[TIAB]</td>
<td>245359</td>
</tr>
<tr>
<td>#35</td>
<td>Search score[TIAB]</td>
<td>94893</td>
</tr>
<tr>
<td>#34</td>
<td>Search self-assessment[TIAB]</td>
<td>2896</td>
</tr>
<tr>
<td>#33</td>
<td>Search self-report[TIAB]</td>
<td>11048</td>
</tr>
<tr>
<td>#32</td>
<td>Search measure*[TIAB]</td>
<td>158645</td>
</tr>
<tr>
<td>#31</td>
<td>Search questionnaire[TIAB]</td>
<td>95367</td>
</tr>
<tr>
<td>#30</td>
<td>Search &quot;Questionnaires&quot;[MeSH] OR &quot;Weights and Measures&quot;[MeSH] OR &quot;Outcome Assessment”(Health Care)[MeSH] OR &quot;Treatment Outcome&quot;[MeSH]</td>
<td>466731</td>
</tr>
<tr>
<td>#29</td>
<td>Search #15 AND #21 AND #28</td>
<td>895</td>
</tr>
<tr>
<td>#28</td>
<td>Search #22 OR #23 OR #24 OR #25 OR #26 OR #27</td>
<td>1429234</td>
</tr>
<tr>
<td>#27</td>
<td>Search multiple[TIAB]</td>
<td>357243</td>
</tr>
<tr>
<td>#26</td>
<td>Search repetitive[TIAB]</td>
<td>27785</td>
</tr>
<tr>
<td>#25</td>
<td>Search functional*[TIAB]</td>
<td>425161</td>
</tr>
<tr>
<td>#24</td>
<td>Search recurren*[TIAB]</td>
<td>194753</td>
</tr>
<tr>
<td>#23</td>
<td>Search chronic*[TIAB]</td>
<td>473604</td>
</tr>
<tr>
<td>#22</td>
<td>Search &quot;Recurrence”[MeSH]</td>
<td>100803</td>
</tr>
<tr>
<td>#21</td>
<td>Search #16 OR #17 OR #18 OR #19 OR #20</td>
<td>83178</td>
</tr>
<tr>
<td>#20</td>
<td>Search inversion[TIAB]</td>
<td>13941</td>
</tr>
<tr>
<td>#19</td>
<td>Search instability[TIAB]</td>
<td>34623</td>
</tr>
<tr>
<td>#18</td>
<td>Search sprain*[TIAB]</td>
<td>2071</td>
</tr>
<tr>
<td>#17</td>
<td>Search unstable[TIAB]</td>
<td>31862</td>
</tr>
<tr>
<td>#16</td>
<td>Search &quot;Sprains and Strains”[MeSH:NoExp] OR &quot;Joint Instability”[MeSH]</td>
<td>9501</td>
</tr>
<tr>
<td>#15</td>
<td>Search #12 OR #13 OR #14</td>
<td>19963</td>
</tr>
<tr>
<td>#14</td>
<td>Search ankle*[TIAB]</td>
<td>17290</td>
</tr>
<tr>
<td>#13</td>
<td>Search &quot;Lateral Ligament, Ankle”[MeSH]</td>
<td>136</td>
</tr>
<tr>
<td>#12</td>
<td>Search &quot;Ankle Joint”[MeSH]</td>
<td>5761</td>
</tr>
</tbody>
</table>
Figure 2. Flow diagram of the selection procedure of the instruments.

- Medline: N=332
- Cochrane: N=62
- Web of science: N=213
- Embase: N=109
- Cinahl: N=108
- Sportdiscus: N=115

N=939 articles

Total search

N=17 instruments

Selection on title or abstract

N=3 instruments

- AJFAT (1 article)
- FAOS (1 article)
- FADI (1 article)

Selection on the basis of the full text version
<table>
<thead>
<tr>
<th>Clinimetric quality</th>
<th>Definition</th>
<th>Criteria to rate the clinimetric quality</th>
</tr>
</thead>
</table>
| **Content validity** | The extent to which the domain of interest is comprehensively sampled by the items in the measure | 1) Patients were involved during item selection/reduction  
2) Patients were consulted for reading and comprehension  
Rating: + patients and experts were involved  
± only patients were involved  
- no patient involvement  
? no information found on content validity |
| **Readability** | The questionnaire is understandable for all patients | Rating: + reading was tested and result was good  
- inadequate readability  
? no information about readibility |
| **Reliability** | The extent to which the same results are obtained on repeated administrations of the same measure when no change in physical functioning has occurred (reliability) or the extent to which scores are on repeated measurements (agreement) | 1) Correlation coefficient \((r > .70)\); limits of agreement, kappa or standard error of measurement are presented  
Rating: + adequate design, method and \(r > .70\)  
± doubtful method used  
- inadequate reliability or agreement  
? no information found on reliability or agreement |
| **Internal consistency** | The extent to which items in a subscale are inter-correlated; a measure of the homogeneity of the subscale | 1) Factor analysis was applied in order to provide the dimensionality of the measure  
2) Cronbach’s alpha between .70 and .90 for each subscale  
Rating: + adequate design, factor analysis; alpha: .70–.90  
± doubtful method used  
- inadequate internal consistency  
? no information found on internal consistency |
| **Construct validity** | The extent to which scores relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured | 1) Hypotheses were formulated  
2) Results were acceptable in accordance with the hypotheses  
Rating: + adequate design, results in accordance with the hypotheses  
± doubtful method used  
- inadequate construct validity  
? no information found on construct validity |
| **Floor-ceiling effects** | The measure fails to demonstrate a worse score in patients who were clinically deteriorated and/or an improved score in patients who clinically improved | 1) Descriptive statistics of the distribution of scores were presented  
2) 15% of the respondents achieved the highest or lowest possible score  
Rating: + no floor- and ceiling effects  
- > 15% in extremities  
? no information found on floor-ceiling effects |
| **Responsiveness** | The ability to detect important change over time in the concept being measured | 1) Hypotheses were formulated and results were in agreement  
2) An adequate measure was used (effect size, standard response mean or comparison with external standard)  
Rating: + adequate design, method and result  
± doubtful method used  
- inadequate responsiveness  
? no information found on responsiveness |
<table>
<thead>
<tr>
<th>Clinimetric quality</th>
<th>Definition</th>
<th>Criteria to rate the clinimetric quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretability</td>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Authors provided information on the interpretation of scores: 1) Presentation of means and standard deviations of scores 2) Comparative data in relevant subgroups 3) Information on the relationship of scores to well-known functional measures or clinical diagnosis 4) Information on the association between change in scores and patients' global ratings of the magnitude of change they have experienced Rating: + 2 or more types of information was presented ± doubtful method used or doubtful description ? no information found on interpretability</td>
</tr>
<tr>
<td>Minimally clinical important difference (MCID)</td>
<td>The smallest difference in scores in the domain of interest which patients perceive as beneficial and would mandate a change in patients' management</td>
<td>Information is provided about what difference in score would be clinically meaningful Rating: + minimally clinical important difference presented - no minimally clinical important difference presented</td>
</tr>
<tr>
<td>Time to administer</td>
<td>Time needed to complete the measure</td>
<td>Rating: + less than 10 minutes - more than 10 minutes ? no information</td>
</tr>
<tr>
<td>Administration burden</td>
<td>Ease of method used to calculate the questionnaire's score</td>
<td>Rating: + easy: summing up the items ± moderate: visual analogue score or simple formula - difficult: complex formula ? no information found on rating method</td>
</tr>
</tbody>
</table>
Additional files provided with this submission:

Additional file 1: additional file table 1 BMC.doc: 38Kb
http://www.biomedcentral.com/imedia/6404425581114682/sup1.DOC