## Appendix W1. Checklist for rating the clinimetric quality of self-assessment questionnaires

<table>
<thead>
<tr>
<th>Clinimetric property</th>
<th>Definition</th>
<th>Criteria used 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 questionnaire.                                                                                                                                                                                                                                   | 1) patients were involved during item selection and/or item reduction  
2) patients were consulted for reading and comprehension.  
Rating: + patients and (investigator or expert) involved  
± patients only  
- no patient involvement  
? no information found on content validity |
| **Readability & comprehension**      | The questionnaire is understandable for all patients                                                                                                                                                                                                                                                                                    | Rating: + readability tested; result was good  
- inadequate readability  
? no information found on readability and comprehension |
| **Internal consistency**             | The extent to which items in a (sub)scale are intercorrelated; a measure of the homogeneity of a (sub)scale                                                                                                                                                                                                                                   | 1) Factor analysis was applied in order to provide empirical support for the dimensionality of the questionnaire.  
2) Cronbach's alpha between 0.70 and 0.90 for every dimension/subscale  
Rating: + adequate design & method; factor analysis; alpha 0.70-0.90  
± doubtful method used  
- inadequate internal consistency  
? no information found on internal consistency |
| **Construct validity**               | The extent to which scores on the questionnaire 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  
3) an adequate measure was used  
Rating: + adequate design, method, and result  
± doubtful method used  
- inadequate construct validity  
? no information found on construct validity |
| **Floor & ceiling effects**          | The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved                                                                                                                                                                                                   | 1) descriptive statistics of the distribution of scores were presented  
2) 15% of respondents achieved the highest or lowest possible score  
Rating: + no floor / ceiling effects  
- more than 15% in extremities  
? no information found on floor and ceiling effects |
| **Test-retest reliability**          | The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred                                                                                                                                                                                      | 1) calculation of an intraclass correlation coefficient (ICC); ICC > 0.70  
2) time interval and confidence intervals were presented  
Rating: + adequate design, method, and ICC > 0.70  
± doubtful method was used  
- inadequate reliability  
? no information found on test-retest reliability |
| **Agreement**                        | The ability to produce exactly the same scores with repeated measurements                                                                                                                                                                                                                                                                     | 1) for evaluative questionnaires reliability agreement should be assessed  
2) limits of agreement, Kappa, or standard error of measurement (SEM) was presented  
Rating: + adequate design, method and result  
± doubtful method used  
- inadequate agreement  
? no information found on agreement |
| **Responsiveness**                   | The ability to detect important change over time in the concept being measured                                                                                                                                                                                                                                                           | 1) for evaluative questionnaires responsiveness should be assessed  
2) hypotheses were formulated and results were in agreement  
3) an adequate measure was used (ES, SRM, comparison with external standard)  
Rating: + adequate design, method and result  
± doubtful method used  
- inadequate responsiveness  
? no information found on responsiveness |
| **Interpretability**                 | The degree to which one can assign qualitative meaning to quantitative scores                                                                                                                                                                                                                                                           | Authors provided information on the interpretation of scores:  
1. presentation of means and SD of scores before and after treatment  
2. comparative data on the distribution of scores in relevant subgroups  
3. information on the relationship of scores to well-known functional measures or clinical diagnosis  
4. information on the association between changes in score and patients' global ratings of the magnitude of change they have experienced  
Rating: + 2 or more of the above types of information was presented  
± doubtful method used or doubtful description  
? no information found on interpretation |
| **Minimal clinically important difference (MCID)** | The smallest difference in score in the domain of interest which patients perceive as beneficial and would mandate a change in patient's management                                                                                                                                                                                               | Information is provided about what (difference in) score would be clinically meaningful.  
Rating: + MCID presented  
- no MCID presented |
| **Time to administer**               | Time needed to complete the questionnaire                                                                                                                                                                                                                                                                                              | Rating: + less than 10 minutes  
- more than 10 minutes  
? no information found on time to complete the questionnaire |
| **Administration burden**            | Ease of the method used to calculate the questionnaire's score                                                                                                                                                                                                                                                                              | Rating: + easy: summing up of the items  
± moderate: visual analogue scale (VAS) or simple formula  
- difficult: VAS in combination with formula, or complex formula  
? no information found on rating method |

** clinimetric property**: quality aspect of a questionnaire

- **Rating**: +: well presented, ±: good presented, -: uncertain method used, –: not presented

- **Definition** refers to the clinimetric property

- **Criteria used to rate the clinimetric quality** refers to the components to be considered for rating the quality of the questionnaire.