Progress towards understanding anterior knee pain after total knee arthroplasty
Breugem, S.J.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

Translation and Validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty

D. Haverkamp, S.J.M. Breugem, I.N. Sierevelt, L. Blankevoort, C.N. van Dijk

Abstract

Background
In 1998, the Oxford 12-item knee questionnaire was developed by Dawson et al as a self-administered disease and site-specific questionnaire, specifically developed for knee arthroplasty patients. Since then it has proven to be an effective outcome questionnaire, and is widely used. Despite the positive psychometric properties for the Total Knee Arthroplasty (TKA) population, the 12-item knee questionnaire has only been translated in a few languages. We thus translated and validated the Oxford 12-item knee questionnaire for the Dutch population.

Methods and results
After translation according to a forward/backward protocol, 174 knee arthroplasty patients were asked to answer the questionnaire together with an SF-36, an AKSS and a VAS. The reliability, validity, content validity and the sensitivity to change were all tested. Our Dutch version of the Oxford 12-item knee questionnaire achieved excellent scores in all of these properties.

Interpretation
The Dutch Oxford 12-item knee questionnaire proved to be an excellent evaluation instrument for the Dutch Orthopaedic Surgeon and can be used for all total knee arthroplasty patients.
Introduction

Orthopaedic surgeons are increasingly interested in determining the outcome of their surgical interventions, which is reflected by the variety of outcome measures \[^{[2,7-8,10-11,14]}\]. In 1998, the Oxford 12-item knee questionnaire was developed by Dawson et al as a self-administered disease and site specific questionnaire, specifically developed for knee arthroplasty patients \[^{[8]}\]. Since then, it has proven to be an effective outcome questionnaire, and is widely used \[^{[11,15-16]}\]. Despite the positive psychometric properties for the Total Knee Arthroplasty (TKA) population, the 12-item knee questionnaire has only been translated and validated in a few languages and subsequently validated in only two cases: Swedish and Italian \[^{[11,16]}\]. In both languages the 12-item knee questionnaire has good psychometric properties and is recommended for the use of evaluating TKA results \[^{[11,16]}\].

In a large survey of 3600 knee arthroplasty patients Dunbar et al compared several site/disease specific questionnaires and showed that the 12-item knee questionnaire scored the best for burden, feasibility, content validity and reliability \[^{[11,9]}\]. Davies recently performed a review of all knee-specific questionnaires; concluding that the 12-item knee questionnaire, 36 item Short-Form health survey (SF-36) and the WOMAC were the most appropriate for the assessment of outcome after total knee arthroplasty \[^{[7]}\]. Liow et al recently compared studies of different functional rating systems for TKA and the 12-item knee questionnaire emerged as the most reliable, since it eliminates interobserver errors \[^{[15]}\].

The only knee specific questionnaire that is validated in Dutch is the WOMAC \[^{[18]}\]. The WOMAC is a commonly used disease and site specific questionnaire, but it was not developed specifically to evaluate functional outcomes of TKA \[^{[2]}\]. The 12-item knee questionnaire has been shown to be reliable and the most appropriate for the assessment of outcome after total knee arthroplasty \[^{[7,11,9,15]}\].

A validated Dutch version of the 12-item knee questionnaire therefore would be of great value. The purpose of this study was to translate the 12-item knee questionnaire into the Dutch language and to validate it.
Patients and methods

Translation procedure
A Dutch translation was made using a forward/backward translation protocol according to the guidelines of cross-cultural adaptation from Guillemin et al [12]. Since no major cultural differences in lifestyle exist between the Dutch and English population, cultural adaptation of the questionnaire was assumed not to be required.

Patients
The study population consisted of 174 patients (mean age 69 (31-92) years), 65% of whom were female. A preoperative 12-item knee questionnaire in Dutch was obtained from 119 patients who were scheduled for a primary TKA for osteoarthritis of the knee joint. From 111 patients, a 12-item knee questionnaire was obtained 1 year after TKA. Within these patient groups, 56 patients took part in both preoperative and postoperative sampling. All patients were examined clinically and radiographs were taken both preoperatively and at follow up.

Questionnaires
Besides the 12-item knee questionnaire, the outcome measures that we used were the SF36, the American Knee Society (AKS) score, and a 100 mm Visual Analog Scale (VAS) for pain.

The 12-item knee questionnaire is a self-administered disease and site specific questionnaire and consists of 12 questions to be answered by the patient concerning the knee, and is therefore entirely subjective [8]. Each question consists of a 5-point likert scale, leading to a total score ranging from a best functional score of 12 to the worst functional outcome of 60 [8]. By incorporating questions about pain and disability, these items have more weight in the score. For this analysis, we inverted (made reciprocal) the values of the 12-item knee questionnaire in order to simplify the interpretation of the results, since all the other scores have values that are reciprocal to those of the 12-item questionnaire. The SF36 is a general health questionnaire and consists of 36 Likert box questions [1,5,19-20]. The questionnaire contains 8 health concepts: physical functioning (PF), role limitation due to physical problems (RP), bodily pain (BP), perception of general health (GH), energy and vitality (VT), social
functioning (SF), role limitation due to emotional problems (RE), and mental health (MH). The results of the SF36 range from 0 (the worst outcome) to 100 (the best outcome) \[5,19\].

The American Knee Society (AKS) score is a clinical rating system and is divided into a knee score and a function score \[13\]. The knee score is designed to assess pain, stability and range of motion. This score can be influenced by subtracting points for flexion contractures, extension lag and malalignment \[13\]. The function score is designed to assess walking distance and the ability to climb stairs. Again, points are subtracted for the use of walking aids. Both scores range from 0 to 100, with a high score signifying a good result \[13\]. The VAS is a 10-point visual analog scale (100mm) and is used to determine the seriousness of the general pain in the affected knee over the preceding month \[17\].

**Testing**

Psychometrics can be defined as “the scientific measurement of mental capacities and processes of personality” \[9\]. Psychometrics is therefore the “process that allows researchers to apply scientific methodology to the measurement of subjective outcomes” \[9\]. The psychometric properties of a questionnaire define how well a questionnaire measures what it is supposed to measure. The aspects that were tested were reliability, validity and responsiveness. The patient burden imposed by administering the 12-item knee questionnaire was considered to be identical to the original version of Dawson and these authors found it to be minimal \[8\].

**Reliability**

Reliability is defined as the ability of a test to “yield the same results on repeated trials under the same conditions” \[15\]. To determine the test-retest reliability, a second questionnaire was given to a random sample of 26 patients who were scheduled for total knee arthroplasty. They were asked to complete the questionnaires with one week interval and return them under similar circumstances and return them immediately after completion. Test-retest reliability was assessed by the intra class coefficient (ICC) \[4\], the Pearson correlation coefficient and the Coefficient of reliability of Bland and Altman \[3\]. T-tests were performed to determine the systematic difference between the first and second test.
Internal consistency was assessed to determine whether all 12 items cover the same construct. It was calculated from the data from the preoperative (n=119), and postoperative (n=111) populations separately, and data from the total population (n=174). Internal consistency was assessed by Cronbach’s alpha [6]. Cronbach’s alpha addresses the homogeneity of the questions included in a questionnaire and is complimentary to the ICC as a measure of reliability [11,6]. An alpha of 0.7 is considered to represent a fair degree internal consistency, 0.8 is considered to represent good internal consistency, and 0.9 would represent excellent internal consistency [11].

**Validity**

Validity relates to the ability of a questionnaire to measure the outcome parameter of interest [9]. Criterion validity refers to comparison of the new test with the “gold standard”. Unfortunately, there is no “gold” standard for knee arthroplasty. Consequently, questionnaires for knee arthroplasty are usually validated against a postulated effect that should result from the intervention [9]. Such a postulation is referred to as a construct [9]. Construct validity was tested by comparing the 12-item knee questionnaire to a 100-mm VAS scale for pain, the AKS knee and function score and the relevant domains of the SF-36 scores. The scores of the 8 domains of the SF-36 were used to assess convergent and divergent validity of the 12-item knee questionnaire. Construct validity was evaluated using Pearson correlation coefficients between the 12-item knee questionnaire and the 100 mm VAS, the AKS knee and function score and the SF-36.

We evaluated convergent and divergent validity by hypothesising that correlation coefficients between the study questionnaire and SF-36 domains bodily pain (BP), role of physical limitations (RP) and physical functioning (PF), were higher than correlations with the other domains.

Content validity addresses whether a questionnaire has enough items and adequately covers the domain of interest [11]. Content validity was evaluated by assessing the distribution and floor and ceiling effects of the 12-item knee questionnaire. The floor effect occurs when the patient scores the lowest possible score (12), and therefore the patient appears to be very satisfied [11]. The ceiling effect is the highest possible score and therefore the opposite of the floor effect [11]. Construct validity and content validity were tested on the scores of the preoperative and postoperative populations separately.
Responsiveness
Responsiveness is a measure of the ability of a questionnaire to determine sensitivity to change, before and after an intervention. A sample of 56 patients with preoperative and postoperative scores were used to determine sensitivity to change. Effect sizes of all the questionnaires were calculated and compared. Effect size is defined as the ratio of the mean change in pre- and post operative score; this value is then divided by the standard deviation of the pre-operative score. It represents a standardized measure of change in a group, and can be used to compare different clinical measures [11]. Effect sizes of 0.2, 0.5 and 0.8 are considered as small, moderate and large respectively [8,11].

Statistics
The statistical analysis was done using SPSS 12 (version 12.01; SPSS Inc., IL), and a p-value of less than 0.05 was considered to be statistically significant.

Results

Reliability
The ICC of the questionnaire was high at 0.97 (p < 0.01) (95% CI: 0.94 to 0.99), and we found a mean difference of 2.6 between the 2 assessments (p < 0.01). The reliability coefficient of Bland and Altman was 4.6 (95% CI: -7.2 to 2.0). Crohnbach’s alpha showed that the questionnaire had a strong internal consistency, with values of 0.87 and 0.90 preoperatively and at 1-year follow-up, respectively. The value of Cronbach alpha for the total population was 0.94.

Validity
The 12-item knee questionnaire correlated very well with the VAS before the TKA operation and at 1-year follow-up. However, the correlation between the 12-item knee questionnaire and the 2 AKS scores was low before the operation and moderate one year after TKA. The highest correlation coefficients with the domains of the SF36 were seen for the PF, RP and BP. However, the correlation coefficient of the RP before the operation was moderate (table 1).
Table 1: Pearson Correlation Coefficients of the Oxford 12-item knee score, VAS, AKS and SF36 scores.

<table>
<thead>
<tr>
<th></th>
<th>Preoperatively</th>
<th>1 year follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain VAS</td>
<td>0.62*</td>
<td>0.81*</td>
</tr>
<tr>
<td>AKSS knee score</td>
<td>.33*</td>
<td>.62*</td>
</tr>
<tr>
<td>AKSS function</td>
<td>.31*</td>
<td>.48*</td>
</tr>
<tr>
<td>SF36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Functioning (PF)</td>
<td>.64*</td>
<td>.64*</td>
</tr>
<tr>
<td>Role limitation due to Physical problems (RP)</td>
<td>.40*</td>
<td>.63*</td>
</tr>
<tr>
<td>Role limitation due to Emotional problems (RE)</td>
<td>.26*</td>
<td>.50*</td>
</tr>
<tr>
<td>Social Functioning (SF)</td>
<td>.46*</td>
<td>.57*</td>
</tr>
<tr>
<td>Mental Health (MH)</td>
<td>.27*</td>
<td>.40*</td>
</tr>
<tr>
<td>Energy Vitality (VT)</td>
<td>.14*</td>
<td>.48*</td>
</tr>
<tr>
<td>Bodily pain (BP)</td>
<td>.68*</td>
<td>.67*</td>
</tr>
<tr>
<td>General Health Perception (HP)</td>
<td>.31*</td>
<td>.39*</td>
</tr>
</tbody>
</table>

All correlations are significant with p<0.01

Content validity

The scores of the questionnaire before operation and at 1-year follow up were normally distributed. Ceiling and floor effects were not observed preoperatively. At 1-year follow-up a slight floor effect was observed, 10 patients (9%) scored the lowest possible score. A ceiling effect was not present postoperatively.

Sensitivity to change

A statistically significant improvement of -16 (95% CI -18; -13) points for the 12-item knee questionnaire was observed one year after total knee arthroplasty (p < 0.01). The scores of the other outcome measures improved significantly too (p < 0.01), except Mental Health of the SF-36 (p = 0.5). The effect size of the 12-item knee questionnaire one year postoperatively was -2.03, which was larger than the effect sizes of the other outcome measures (table 2).
Table 2: Effect size of the Oxford 12-item knee, VAS, AKS and SF36 scores.

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Preoperative mean (sd)</th>
<th>1 year follow-up mean (sd)</th>
<th>Difference</th>
<th>95% CI</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford-12 item (n=56)</td>
<td>34.0 (7.5)</td>
<td>49.4 (9.0)</td>
<td>-15.4</td>
<td>-17.6; -13.1</td>
<td>-2.03</td>
</tr>
<tr>
<td>VAS</td>
<td>47.3 (26.7)</td>
<td>84.1 (24.6)</td>
<td>-36.8</td>
<td>-44.6; -29.0</td>
<td>-1.38</td>
</tr>
<tr>
<td>AKSS knee score</td>
<td>59.2 (17.4)</td>
<td>87.1 (14.6)</td>
<td>-27.9</td>
<td>-35.4; -20.4</td>
<td>-1.63</td>
</tr>
<tr>
<td>AKSS function score</td>
<td>54.0 (19.6)</td>
<td>76.3 (21.6)</td>
<td>-22.3</td>
<td>-29.7; -14.9</td>
<td>-1.13</td>
</tr>
<tr>
<td>SF36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Functioning (PF)</td>
<td>33.1 (18.6)</td>
<td>58.0 (25.6)</td>
<td>-24.9</td>
<td>-31.9; -17.9</td>
<td>-1.34</td>
</tr>
<tr>
<td>Role limitation due to Physical problems (RP)</td>
<td>16.7 (27.9)</td>
<td>57.4 (41.7)</td>
<td>-40.7</td>
<td>-53.1; -28.3</td>
<td>-1.46</td>
</tr>
<tr>
<td>Role limitation due to Emotional problems (RE)</td>
<td>57.3 (43.2)</td>
<td>73.8 (41.9)</td>
<td>-16.5</td>
<td>-27.6; -5.4</td>
<td>-0.38</td>
</tr>
<tr>
<td>Social Functioning (SF)</td>
<td>59.3 (21.6)</td>
<td>72.5 (20.6)</td>
<td>-13.1</td>
<td>-20.1; -6.1</td>
<td>-0.61</td>
</tr>
<tr>
<td>Mental Health (MH)</td>
<td>77.6 (19.1)</td>
<td>79.5 (20.4)</td>
<td>-1.9</td>
<td>-6.8;3.0</td>
<td>-0.10</td>
</tr>
<tr>
<td>Energy Vitality (VT)</td>
<td>61.0 (22.3)</td>
<td>70.2 (22.6)</td>
<td>-9.2</td>
<td>-13.9; -4.5</td>
<td>-0.41</td>
</tr>
<tr>
<td>Pain (BP)</td>
<td>38.8 (19.6)</td>
<td>70.7 (28.9)</td>
<td>-31.9</td>
<td>-39.5; -24.3</td>
<td>-1.63</td>
</tr>
<tr>
<td>General Health (GH)</td>
<td>72.0 (19.0)</td>
<td>64.5 (23.4)</td>
<td>7.5</td>
<td>3.1;11.9</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Discussion

The aim of this study was to translate the original disease- and site-specific 12-item knee questionnaire, on the subjective input of patients with total knee arthroplasties, into the Dutch language and to validate it. A mere translation of the questionnaire into another language - without proper validation of the translated version – would not suffice [12]. The translation procedure did not create any problems, since the items in this questionnaire are universal for all knee arthroplasty patients and there is little or no cultural difference between Dutch and British patients.

According to the ICC and correlation coefficient, the test-retest reliability was good. However a statistical significant difference of 2.85 was observed between the two assessments. Although this difference is small and not relevant clinically, it can be explained by the fact that it was not specifically asked whether the complaints of the patients had changed at the second assessment. After checking the patient’s files, it appeared that 29 patients were treated with physiotherapy or analgesics. Taking this into account, the results seem to indicate good test-retest reliability.
The study questionnaire has a very high internal consistency (Cronbach’s alpha 0.94) for the total population. This could indicate redundancy, but since the Cronbach’s alpha did not exceed 0.9 when assessed for the preoperative and postoperative populations separately, no questions need to be excluded from the list. Dunbar and Dawson obtained comparable results \cite{8,11}. Pearson correlation coefficients between the study questionnaire and VAS were good, low preoperatively and moderate postoperatively with AKSS. These moderate preoperative correlations are comparable to the results of Dawson, who did not mention the postoperative values \cite{8}. However, convergent and divergent validity is confirmed, since the study questionnaire correlated better with the domains BP, PF and RP than with the other domains of the SF36. The low correlation coefficient with the PF domain preoperatively can be explained by a considerable floor effect of the PF; 73 patients (64%) had the minimum score. Considering these results, we conclude that the questionnaire has good construct validity. The 12-item knee questionnaire has proven to be highly sensitive, since the effect size of the study questionnaire one year post operatively was -2.0, which was higher than for any of the individual parts of the SF-36 (table2).

Conclusion

The Dutch Oxford 12-item knee questionnaire proved to be an excellent evaluation instrument for the Dutch Orthopaedic Surgeon to be used for all total knee arthroplasty patients.
References


FUNCTIEBEPERKINGEN Vragenlijst (12-questionnaire score)

In dit onderdeel wordt u gevraagd om aan te geven in hoeverre uw aangedane knie uw dagelijkse leven beïnvloedt. Wilt u bij ieder onderdeel één mogelijkheid aankruisen welke het meest op u van toepassing is?

1. Hoe zou u de pijn die u gewoonlijk in u knie heeft omschrijven?
   ☐ Geen pijn
   ☐ Hele lichte pijn
   ☐ Milde pijn
   ☐ Matige pijn
   ☐ Ernstige pijn

2. Heeft u vanwege uw knie moeite met uzelf te wassen en af te drogen?
   ☐ Geen moeite
   ☐ Weinig moeite
   ☐ Moeite
   ☐ Veel moeite
   ☐ Onmogelijk

3. Heeft u problemen om in/uit de auto te komen en/of problemen om met het openbaar vervoer te reizen vanwege u knie?
   ☐ Geen problemen
   ☐ Weinig problemen
   ☐ Problemen
   ☐ Veel problemen
   ☐ Onmogelijk

4. Hoe lang kunt u lopen voordat de pijn in uw knie te hevig wordt?
   (met of zonder stok/rollator)
   ☐ Geen pijn / meer dan 30 minuten
   ☐ 16 tot 30 minuten
   ☐ 5 tot 15 minuten
   ☐ Rondom het huis
   ☐ Niet mogelijk
5. Hoe pijnlijk is het om na het eten van tafel op te staan?
- Niet pijnlijk
- Licht pijnlijk
- Pijnlijk
- Zeer pijnlijk
- Ondraaglijk

6. Loopt u wel eens mank vanwege de knie?
- Zelden/nooit
- Af en toe of sinds kort voor het eerst
- Regulier
- Vrijwel altijd
- Altijd

7. Kunt u door de hurken gaan en weer omhoog komen?
- Ja, gemakkelijk
- Met wat moeite
- Met moeite
- Met zeer veel moeite
- Nee, onmogelijk

8. Heeft u ‘s-nachts wel eens pijn in de knie?
- Geen enkele nacht
- 1 of 2 nachten per week
- 3 of 4 nachten per week
- 5 of 6 nachten per week
- alle nachten

9. Hoeveel invloed heeft de pijn in uw knie op de dagelijkse bezigheden?
- In het geheel niet
- Weinig
- Matig
- Veel
- Volledig
10. Heeft u weleens het gevoel dat u door de knie heen zakt?
- Zelden/nooit
- Af en toe of sinds kort voor het eerst
- Regelmatig
- Vrijwel altijd
- Altijd

11. Kunt u zelf uw boodschappen doen?
- Ja, makkelijk
- Met wat moeite
- Met moeite
- Met zeer veel moeite
- Nee, onmogelijk

12. Kunt u de trap aflopen?
- Ja, makkelijk
- Met weinig moeite
- Met moeite
- Met zeer veel moeite
- Nee, onmogelijk