The translated Danish version of the Western Ontario Meniscal Evaluation Tool (WOMET) is reliable and responsive

Clementsen, Jon M.; Skou, Søren T.; Hansen, Sascha L.; Eshøj, Henrik Rode; Mølgaard, Carsten M.; Mikkelsen, Lone R.; Thorlund, Jonas B.

Published in:
Knee Surgery, Sports Traumatology, Arthroscopy

DOI (link to publication from Publisher):
10.1007/s00167-021-06551-6

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

The translated Danish version of the Western Ontario Meniscal Evaluation Tool (WOMET) is reliable and responsive

Authors:
Jon M Clementsen (1), Søren T Skou (2, 3), Sascha L Hansen (2), Henrik Rode Eshøj (2, 4), Carsten M Mølgaard (5), Lone R Mikkelsen (1, 6), Jonas B Thorlund (2, 7)

1: Elective Surgery Centre, Silkeborg Regional Hospital, Denmark;

2: Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark;

3: Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Slagelse, Denmark

4: Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense, Denmark

5: Department of Occupational Therapy and Physiotherapy, Aalborg University Hospital, Denmark;

6: Department of Clinical Medicine, Aarhus University, Denmark;

7: Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark;

Correspondence:

Jon M Clementsen

Address: Tom Kristensens Vej 24 st. tv., 2300 København S, Denmark

E-mail: jonclementsen2@gmail.com

MANUSCRIPT WORD COUNT: 2738
ABSTRACT WORD COUNT: 239

Funding: There was no funding source.

Ethical: This study was approved exempted for notification to the Regional Ethics Committee according to Danish law. The study was approved by the Danish Data Protection Agency.

Informed consent: Written informed consent was obtained from all participants after thorough information about the study.

Acknowledgments: We thank the nurses and surgeons at Silkeborg Regional Hospital for their great help in recruiting patients for this study.

Authors' contributions: STS, HE and JBT conceived the study. JMC, SLH, HE, CMPM, LRM and JBT recruited patients, translated the WOMET and collected the data. JBT performed the analyses. JMC and JBT drafted the manuscript. All authors provided important intellectual feedback on the manuscript and approved the final version.

The translated Danish version of the Western Ontario Meniscal Evaluation Tool (WOMET) is reliable and responsive

ABSTRACT

Purpose –The purpose of this study was to translate and cross-culturally adapt the Western Ontario Meniscal Evaluation Tool (WOMET) for use in Denmark and evaluate its test-retest reliability and comparative responsiveness.

Methods – Sixty patients (mean age 50 years (range 19-71 years), females 57%) with meniscal injury scheduled for arthroscopic meniscal surgery at a small Danish hospital in the period from September 2017 to February 2018 were included in this study. The WOMET was translated into Danish using forward and backward translation. The WOMET was completed at baseline (pre-surgery), at 3 and 6 months postoperatively. Additionally, reliability was assessed at 3 months and 3 months plus 1 week, for patients with a stable symptom state (Global Response Question) between test and retest. Comparative responsiveness was assessed between the WOMET and the Knee injury and Osteoarthritis Outcome Score (KOOS4 – aggregate score of 4 of the 5 KOOS subscales).

Results – The Danish version of WOMET showed excellent test-retest reliability, Intraclass Correlation Coefficient of 0.88 (95% CI 0.84-0.92) for the total score. The Standard Error of Measurement was 125 points and the Minimal Detectable Change was 347 points (i.e., 8% and 22% of the total score, respectively). The WOMET was responsive with an effect size (ES) of 1.12 at 6 months after surgery, which was comparable to the KOOS4 (ES 1.10).

Conclusion - The Danish version of the WOMET is a reliable and responsive measure of health-related quality of life in patients with meniscal pathology.

Level of evidence: Level II

INTRODUCTION

Meniscal knee injuries are common in the adult population [15], and are typically associated with pain and functional limitations [16, 24].

A number of patient-reported outcome measures (PROM), such as the Knee injury and Osteoarthritis Outcome Score (KOOS) [18], Lysholm Knee Scale [26], International Knee Documentation Committee Subjective Knee Form (IKDC) [9] and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [3] have been developed to evaluate patient-reported symptoms, functional limitations and health-related quality of life (HRQoL) among patients with knee pathologies. However, there is no consensus as to which PROM is the best instrument to measure patient-reported outcomes for patients with meniscal tears.

The Western Ontario Meniscal Evaluation Tool (WOMET) was developed in 2007 by Kirkley et al., specifically to evaluate meniscal injury symptoms, and therefore might be more responsive to change over time than other generic knee PROMs [10]. For instance, the WOMET was more responsive to measure the impact of meniscal pathology on knee-specific HRQoL compared with the KOOS and WOMAC [21].

The psychometric properties of the WOMET have been investigated several times, including in translational studies into Dutch, German, Chinese, Turkish and Persian, all finding it to be valid, reliable and responsive [5, 6, 19, 28, 29].

No formally translated Danish version of the WOMET is available. Therefore, the scope of this study was to translate and cross-culturally adapt the original English version of the WOMET to Danish and evaluate its reliability and the responsiveness compared with the KOOS.

MATERIALS AND METHODS

Study design

Cross-cultural translation and adaptation of the original English WOMET questionnaire was performed for use in Danish following the standardized and formalized forward and backward translation procedure [2]. The translated Danish version was tested for face validity by semi-structured interviews with patients scheduled for surgery for a meniscal tear. Subsequently, the approved translated Danish version of the WOMET was evaluated for test-retest reliability (3 months and 3 months + 1 week post-surgery). To evaluate comparative responsiveness, the WOMET and KOOS were completed at baseline (pre-surgery) and at 6 months post-surgery. This study was exempted for notification to the Regional Ethics Committee according to Danish law. The study was approved by the Danish Data Protection Agency.

Translation and cross-cultural adaptation of the WOMET

The translation followed the procedure below:

1: Forward translation of the original English version of the WOMET into Danish was independently performed by two native Danish individuals with proficient skills in English. One was a healthcare professional.

2: The two forward translations were reviewed and discrepancies were addressed and handled by consensus between the two forward translators. The two forward translations were then reconciled into one single preliminary Danish version.

2: The preliminary Danish version was then backward translated into English by two individuals proficient in Danish, but with English as first language (one being a healthcare professional).

3: The Danish version of WOMET was evaluated for comprehensibility (linguistic content and understanding) in patients with meniscal injury scheduled for surgery at a Danish university hospital (not otherwise involved in the study) using semi-structured cognitive debriefing interviews. This process was repeated until theoretical data saturation was reached (i.e. no new information was

retrieved by adding further participants).

4: The final Danish version of WOMET was translated into English and sent to the corresponding author of the original English WOMET for approval.

Psychometric testing of the Danish version of the WOMET

Patients

To evaluate reliability and responsiveness, patients with meniscal injuries from a sports surgery facility at a small Danish hospital were recruited by nurses trained in the study procedures at the time they were assigned for knee arthroscopy for MRI-verified meniscal pathology. Written informed consent was obtained after thorough information about the study.

Eligible patients were men and women who met the following inclusion criteria: (1) 18 years or older, (2) knee symptoms compatible with a meniscal tear (i.e. knee pain and joint line tenderness), positive clinical tests (McMurray and Thessaly), (3) MRI-verified meniscal pathology, (4) assigned for knee arthroscopy. Only patients with severe ligamentous instability of the knee were excluded (e.g. ACL rupture).

Patient-reported outcome measures

The Western Ontario Meniscal Evaluation Tool (WOMET)

The WOMET is a disease-specific tool to evaluate HRQoL in patients with meniscal pathology [10, 25]. The WOMET has 16 items, representing three domains. The physical domain has 9 items; sports, recreation, the work and lifestyle domain has 4 items; and the emotion domain has 3 items. Each item is scored on a visual analogue scale from 0 (no symptoms) to 100 (extreme symptoms). The best or least symptomatic overall score is zero and the worst or most symptomatic overall score is 1600. The score can also be converted to a percent of maximum score, to allow comparison to

other PROMs (i.e. scoring on a 0-100 scale), where 0 represents the worst score and 100 the best score. The English WOMET has been tested and found valid in patients with meniscal pathology [1] and has been used as an outcome in several studies on patients with meniscal tears [8, 20, 23].

The Knee injury and Osteoarthritis Outcome Score (KOOS)

For evaluation of comparative responsiveness, patients also completed the KOOS score. The KOOS score is designed to evaluate knee pain, function and symptoms in the continuum from knee injury to knee osteoarthritis (including patients with meniscal tears). The KOOS consists of 5 domains: pain, symptoms, activities of daily living (ADL) function, sports and recreational function (Sport/Rec) and quality of life (QOL). Each domain is scored in a 5-point Likert scale, which is converted to a 0-100 score (0 indicating extreme knee problems and 100 indicating no knee problems) [17, 18]. In previous randomized controlled trials on patients with anterior cruciate ligament (ACL) injuries and meniscal tears, the KOOS4 [7, 11],which is an average of 4 of the 5 KOOS subscales (excluding the ADL subscale due to a ceiling effect in younger populations) has been used as the primary outcome. Therefore, we assessed the comparative responsiveness of the Danish version of the WOMET with the KOOS4 score.

Study procedure

Patients recruited for assessment of test-retest reliability and responsiveness of the Danish version of the WOMET received an online questionnaire, which also included the KOOS questionnaire and questions on baseline characteristics such as age, sex, height, weight, presentation and duration of symptoms and earlier surgery. Surgery was performed maximally 2 weeks after the baseline questionnaire was filled out. The WOMET and KOOS were completed again at 3 months, at 3 months + 1 week (for test-retest reliability) and finally again at 6 months (for responsiveness).To

reduce loss to follow up, patients were contacted by email and text message (SMS) in the case of initial non-response, once weekly for 4 weeks.

To evaluate symptom stability between test and retest assessments (i.e. 3 months and 3 months + 1 week), a Global Response Assessment (GRA) question was used as an external anchor at the 3 months + 1 week assessment time point. The GRA question used was: "Compared with last time you answered this questionnaire, how are your knee symptoms now?", with the response options being "worse", "largely unchanged" and "better". Only patients who responded "largely unchanged" were included in the reliability analysis.

Statistical analysis

Baseline characteristics are summarized as mean with standard deviation (SD) or numbers and percentages as appropriate.

It is generally recommended to include a minimum of 50 subjects in reliability studies [13]. We aimed to include 75 patients to take into account missing data and potential change in symptoms between test and retest. As WOMET scores at test and retest were not normally distributed, we assessed systematic differences between the two time points using Wilcoxon signed rank test. We also performed a paired t-test between the two time points (i.e. test and retest) to facilitate comparison with previous studies. Relative reliability was assessed by calculating the Intraclass Correlation Coefficient (ICC) based on variance components obtained using a linear mixed model approach with time (i.e. test and retest) as fixed effect and ID (person) as random effect. ICC above 0.80 was considered as excellent reliability [12]. Absolute reliability was expressed as the Standard Error of Measurement (SEM) calculated as: SEM = SD of the mean difference (SDdiff) /√2 and the Minimal Detectable Change (MDC) was calculated as SEM x 1.96 x √2. Furthermore, SEM and MDC are presented as absolute values and percentages of maximal scores**.** To provide a visual

145 presentation of the absolute reliability, a Bland-Altman Plot was generated, and inspected for
146 systematic bias and funnel effects [4].
147 Floor and ceiling effects indicate a limited content validity and are considered when more than 15%
148 of patients score either the highest or lowest scores possible. Therefore, the proportion of patients
149 reaching maximum (1600 points) or minimum scores (0 points) was assessed [27].
150 Comparative responsiveness was investigated by comparing the standardized response mean (SRM)
151 as a measure of the effect size (ES) from baseline to 6 months post-surgery between the WOMET
152 and KOOS4 scores. WOMET scores were converted to percentage values as recommended to
153 enable comparison with the KOOS4 [10]. Effect size was calculated as mean change score from
154 baseline to 6 months post-surgery, divided by the SD of change scores. A graphical comparison
155 between WOMET and KOOS4 using all time points (i.e. baseline, 3 months and 6 months) was also
156 performed. The correlation between WOMET and KOOS4 scores was investigated by Pearson's
157 correlation. A correlation higher than 0.80 was considered a strong correlation [27]. No specific
158 sample size calculation was performed for the evaluation of responsiveness, as this was a secondary
159 aim of the study. All statistical analyses were performed using STATA 15.0 software, with a
160 significance level set at $p<0.05$.
161
162

RESULTS

Translation and cross-cultural adaptation

No major language difficulties were seen in the forward and backward translations of the WOMET. Five meniscal injury patients participated in the cognitive debriefing interviews which did not reveal any difficulties in understanding or completing the questionnaire items of the Danish version of WOMET. The backward translated version was accepted without comment by the corresponding author of the original WOMET.

Patients

A total of 81 patients scheduled for arthroscopic meniscal surgery were invited to participate, of whom 76 completed the baseline questionnaire. Ultimately, 55 and 60 patients contributed to the reliability and responsiveness analyses, respectively (Figure 1). (See Table 1 for baseline characteristics of the included patients).

Reliability and Responsiveness

No signs of ceiling or floor effects of the Danish version of the WOMET were observed. Generally, no systematic difference was observed between the total WOMET scores at test and retest, although slightly different scores between test and retest were indicated for the 'emotion' domain (Table 2). Test-retest reliability was excellent, with an ICC of 0.88 (95% CI 0.84-0.92) for the total WOMET score. The ICC scores were similar for all three domains. The SEM and MDC for the total WOMET score was 125 and 347 points, respectively (corresponding to 8% and 22% of the maximum total score, respectively) (Table 2). Visual inspection of the Bland-Altman Plot confirmed the impression of good reliability and did not indicate a systematic bias or funnel effect (Figure 2).

Among the 60 patients with a full dataset, the mean WOMET total score improved from 871 (95% CI 801-941) at baseline to 519 (95% CI 428-611) at 3 months and 481 (95% CI 397-564) at 6 months, and the mean change from baseline to 6 months was 390 (95% CI 301-480).

The mean WOMET percentage score and KOOS4 demonstrated comparable scores at all time points (Figure 3). Similarly, the ES of the change from baseline to 6 months were almost identical, 1.12 for WOMET and 1.10 for KOOS4, and a strong correlation was observed between the WOMET and KOOS4 scores ($r=0.85$, $p<0.0001$) (Supplementary Figure 1).

DISCUSSION

The English version of the WOMET was successfully translated and cross-culturally adapted for use in a Danish patient population with knee meniscal injury. The WOMET was translated into Danish according to international guidelines by bilingual individuals, and with no major difficulties experienced during the translation and adaptation procedure with patients.

The test-retest reliability of the Danish WOMET was excellent, with an ICC of 0.88 for the total score and with all sub-domains showing ICC values exceeding 0.80. This is similar to what has been reported for the original English version of the WOMET total score (ICC=0.83), and what has been reported for the German (ICC=0.90), Chinese (ICC=0.93), Dutch (ICC=0.78), Turkish (ICC=0.87) and Persian (ICC=0.73) versions [5, 6, 10, 19, 28, 29].

The SEM of the total WOMET score was found to be 125 points or 8% of the total score and the MDC was 347 points corresponding to 22% of the total score. These results are similar to that which was found for the Dutch [29] version of the WOMET, reporting an SEM and MDC of 7% and 21% of the total score, respectively. In the assessment of reliability of the Finnish [22] version of the WOMET, the width of the 95% confidence interval for the 'limits of agreement' of the total score was 20%, which corresponds well with the MDC found in this study. Clinicians should be cautious if using the WOMET for individual patients, as a change of 22% or 347 points, corresponding to the MDC, is needed to identify true changes larger than measurement error. For other translated versions of the WOMET (German, Turkish and Persian), much lower SEM values have been reported. However, sufficient data are not reported in these studies to further clarify the reason for the questionably low SEM values. A strength of the current study was the inclusion of a GRA question at the retest, to ensure that patients had not changed their symptom state compared

with the first test session. This led to exclusion of 7 out of 62 patients, highlighting that it cannot be assumed that symptoms are the same, even with a small time period between test and retest. Substantial improvements (i.e. reduced scores) were found in the WOMET total scores from baseline which were assessed prior to surgery and at the 6 months follow up indicating that the Danish version of the WOMET score was able to detect symptom change over time in this population. The large effect size (1.12) observed for the WOMET score from before surgery to the 6 months follow up was comparable to what have been observed for the original English version of the WOMET and other translated versions [5, 10, 19, 28, 29]. The responsiveness of the WOMET and the KOOS4 score was found to be comparable with similar ES and changes in scores over time, which suggest that both questionnaires are equally good at assessing changes in outcome after meniscal surgery.

Some limitations apply to this study. All patients for the test-retest reliability and responsiveness analyses were recruited from a single centre, which may affect generalizability. Nevertheless, patients in this study had a very similar mean age and age range, compared with a larger cohort of Danish meniscal injury patients [24], supporting the use of this tool in the Danish patient population. It was not investigated if the paper version and the electronic version could potentially yield different outcomes in the same group of participants. However, evidence seems to suggest that electronic completion is equivalent to paper completion [14]. Furthermore, using an electronic version resulted in no missing data, nor any incorrectly completed items, which can often be observed when using the paper version of the WOMET. Only five patients participated in the cognitive debriefing interviews to assess their understanding of the translated version of the WOMET. This is substantially lower than recommended, however saturation of information was achieved. Finally, only the MDC was investigated in the current study and thus, the MIC, indicating

241 the smallest change score that is important or meaningful to patients, cannot be extracted from this
242 study.
243
244 **CONCLUSION**
245 The English version of the WOMET was successfully translated and culturally adapted into Danish.
246 The Danish version of the WOMET showed excellent reliability and had comparable
247 responsiveness to the KOOS in detecting change in symptom state following meniscal surgery.
248 Caution should be exercised if using the WOMET for assessing change in individual patients as
249 large changes are needed to rule out measurement variation.
250

References

1. Abram SGF, Middleton R, Beard DJ, Price AJ, Hopewell S (2017) Patient-reported outcome measures for patients with meniscal tears: a systematic review of measurement properties and evaluation with the COSMIN checklist. BMJ Open 7:e017247
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186-3191
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15:1833-1840
4. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8:135-160
5. Celik D, Demirel M, Kus G, Erdil M, Ozdincler AR (2015) Translation, cross-cultural adaptation, reliability and validity of the Turkish version of the Western Ontario Meniscal Evaluation Tool (WOMET). Knee Surg Sports Traumatol Arthrosc 23:816-825
6. Ebrahimi N, Naghdi S, Ansari NN, Jalaie S, Salsabili N (2020) Statistical validity and reliability of the Persian version of the Western Ontario Meniscal Evaluation Tool (WOMET) according to the COSMIN checklist. BMC Musculoskelet Disord 21:183
7. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS (2010) A randomized trial of treatment for acute anterior cruciate ligament tears. N Engl J Med 363:331-342
8. Gelber PE, Torres-Claramunt R, Poggioli F, Perez-Prieto D, Monllau JC (2020) Polyurethane Meniscal Scaffold: Does Preoperative Remnant Meniscal Extrusion Have an Influence on Postoperative Extrusion and Knee Function? J Knee Surg;10.1055/s-0040-1710377

9. Hefti F, Muller W, Jakob RP, Staubli HU (1993) Evaluation of knee ligament injuries with the IKDC form. Knee Surg Sports Traumatol Arthrosc 1:226-234
10. Kirkley A, Griffin S, Whelan D (2007) The development and validation of a quality of life-measurement tool for patients with meniscal pathology: the Western Ontario Meniscal Evaluation Tool (WOMET). Clin J Sport Med 17:349-356
11. Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EM (2016) Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. BMJ 354:i3740
12. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159-174
13. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539-549
14. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. (2015) Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 13:167
15. Peat G, Bergknut C, Frobell R, Joud A, Englund M (2014) Population-wide incidence estimates for soft tissue knee injuries presenting to healthcare in southern Sweden: data from the Skane Healthcare Register. Arthritis Res Ther 16:R162
16. Pihl K, Turkiewicz A, Englund M, Lohmander LS, Jorgensen U, Nissen N, et al. (2018) Change in patient-reported outcomes in patients with and without mechanical symptoms

undergoing arthroscopic meniscal surgery: A prospective cohort study. Osteoarthritis Cartilage 26:1008-1016

17. Roos EM, Roos HP, Ekdahl C, Lohmander LS (1998) Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 8:439-448

18. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD (1998) Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 28:88-96

19. Sgroi M, Daxle M, Kocak S, Reichel H, Kappe T (2018) Translation, validation, and cross-cultural adaption of the Western Ontario Meniscal Evaluation Tool (WOMET) into German. Knee Surg Sports Traumatol Arthrosc 26:2332-2337

20. Sgroi M, Gninka J, Fuchs M, Seitz AM, Reichel H, Kappe T (2020) Chondral lesions at the medial femoral condyle, meniscal degeneration, anterior cruciate ligament insufficiency, and lateral meniscal tears impair the middle-term results after arthroscopic partial meniscectomy. Knee Surg Sports Traumatol Arthrosc 28:3488-3496

21. Sgroi M, Kocak S, Reichel H, Kappe T (2018) Comparison of 3 Knee-Specific Quality-of-Life Instruments for Patients With Meniscal Tears. Orthop J Sports Med 6:2325967117750082

22. Sihvonen R, Jarvela T, Aho H, Jarvinen TL (2012) Validation of the Western Ontario Meniscal Evaluation Tool (WOMET) for patients with a degenerative meniscal tear: a meniscal pathology-specific quality-of-life index. J Bone Joint Surg Am 94:e65

23. Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Kalske J, et al. (2020) Arthroscopic partial meniscectomy for a degenerative meniscus tear: a 5 year follow-up of

the placebo-surgery controlled FIDELITY (Finnish Degenerative Meniscus Lesion Study) trial. Br J Sports Med 54:1332-1339

24. Skou ST, Pihl K, Nissen N, Jorgensen U, Thorlund JB (2018) Patient-reported symptoms and changes up to 1 year after meniscal surgery. Acta Orthop 89:336-344

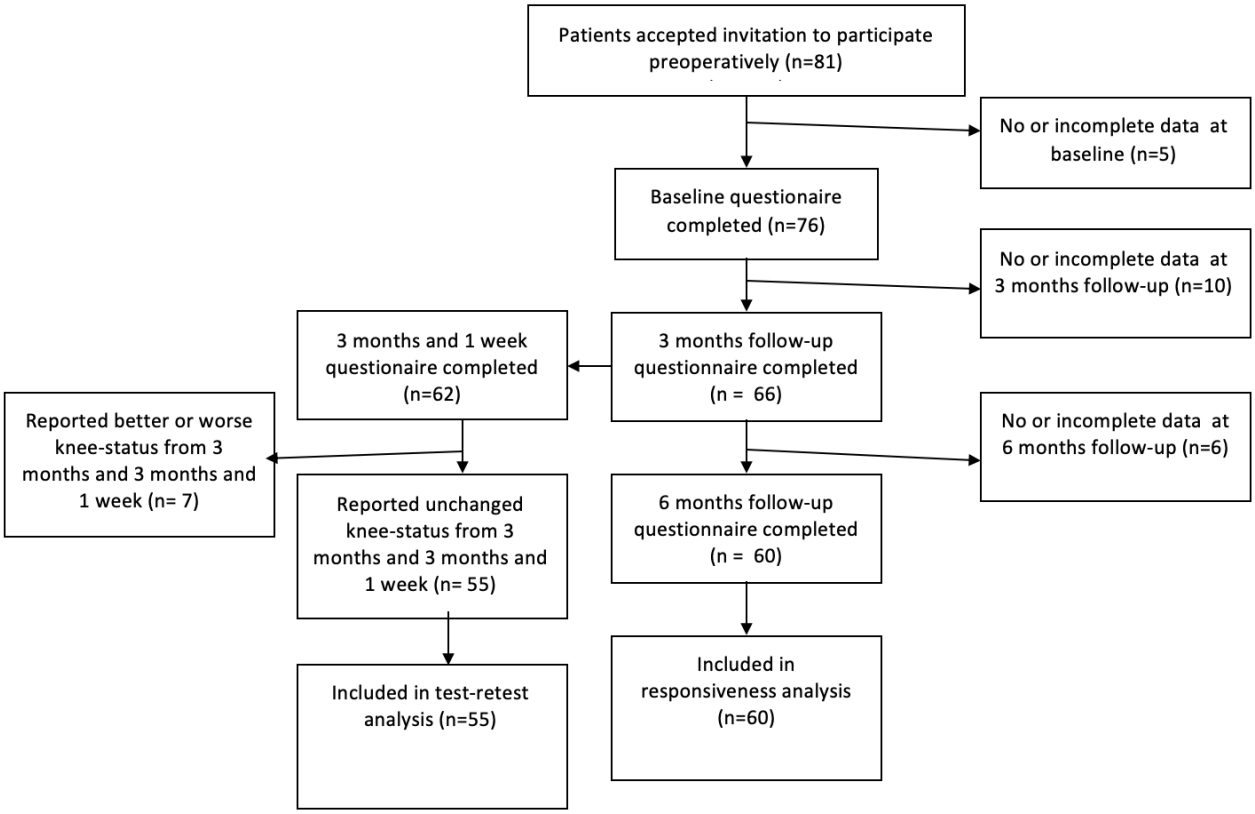
25. Tanner SM, Dainty KN, Marx RG, Kirkley A (2007) Knee-specific quality-of-life instruments: which ones measure symptoms and disabilities most important to patients? Am J Sports Med 35:1450-1458

26. Tegner Y, Lysholm J (1985) Rating systems in the evaluation of knee ligament injuries. Clin Orthop Relat Res 198:43-49

27. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34-42

28. Tong WW, Wang W, Xu WD (2016) Development of a Chinese version of the Western Ontario Meniscal Evaluation Tool: cross-cultural adaptation and psychometric evaluation. J Orthop Surg Res 11:90

29. van der Wal RJP, Heemskerk BTJ, van Arkel ERA, Mokkink LB, Thomassen BJW (2017) Translation and Validation of the Dutch Western Ontario Meniscal Evaluation Tool. J Knee Surg 30:314-322



341

342 **Figure 1:** Study Flowchart for reliability and responsiveness studies

343

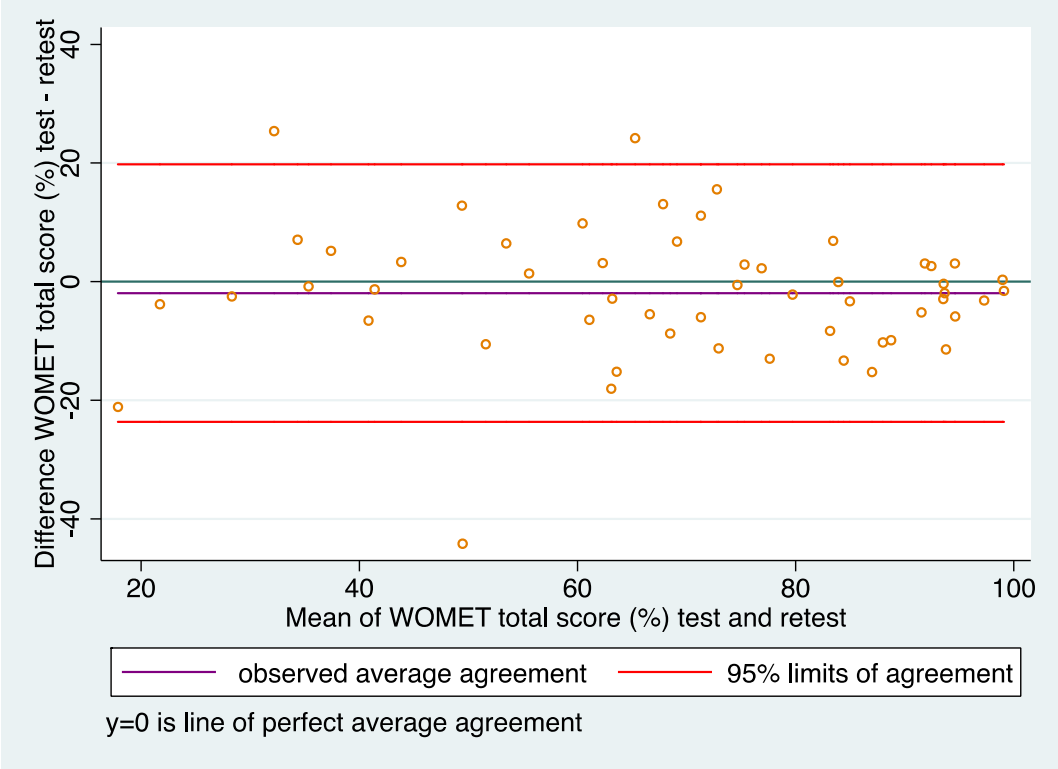


Figure 2: Bland-Altman plot of the reliability of the Danish version of the WOMET score. The green horizontal line intersecting y=0 indicates perfect agreement, whereas the purple line represents the observed mean difference. The closer the purple line is to the green line, the less disagreement between measurements at test and retest. This distance was tested for systematic bias using a paired t-test. Red lines indicate upper and lower limits of agreements (95% LOA), n=55.

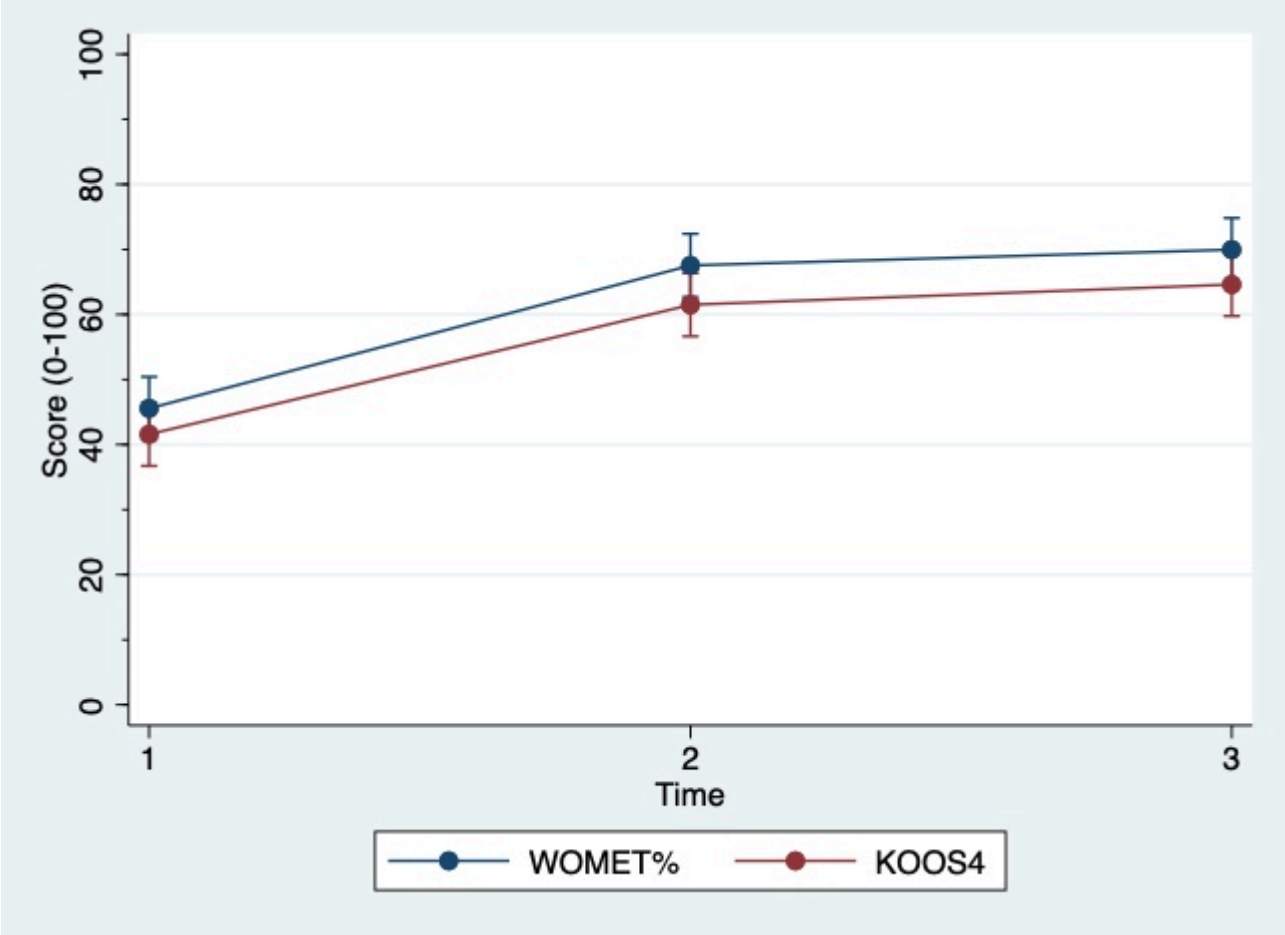
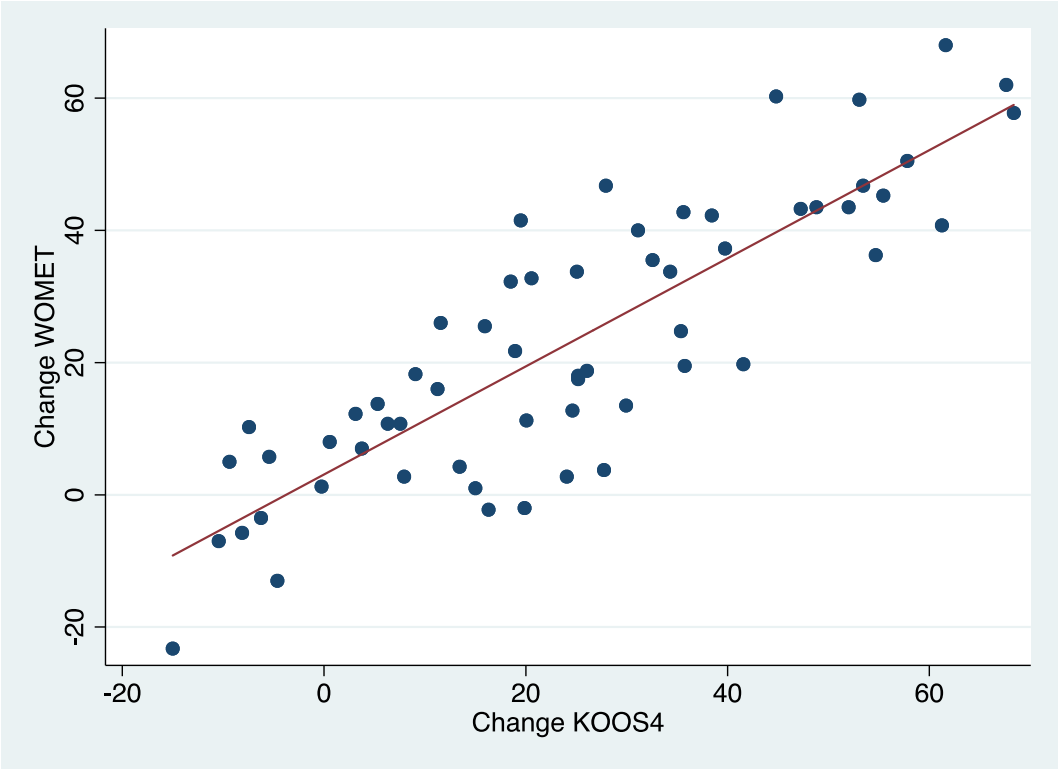


Figure 3: Comparison of WOMET% and KOOS4 scores at 1) baseline (preoperatively), 2) 3 months and 3) 6 months postoperatively. Values are mean scores with 95% CI, n=60.



Supplementary figure 1: Relationship between change in WOMET% score and KOOS4 scores from baseline to 6 months follow up ($r=0.85$, $p<0.0001$).

Table 1: Patient characteristics at baseline.

Variables	Responsiveness analysis (n=60)	Reliability analysis (Test-retest sub-group) (n=55)
Age, years (SD)	50 (11)	50 (11)
Female, no. (%)	34 (57)	28 (51)
BMI, kg/m^2 (SD)	27.6 (5)	27.6 (5)
Duration of symptoms, no. (%)		
0-3 months	6 (10%)	5 (9%)
4-6 months	19 (32%)	17 (31%)
7-12 months	16 (27%)	14 (25%)
13-24 months	11 (18%)	11 (20%)
>24 months	8 (13%)	8 (15%)
Symptom onset, no. (%)		
Slowly evolved over time	26 (43%)	25 (45%)
Semi-traumatic*	22 (37%)	19 (35%)
Traumatic**	12 (20%)	11 (20%)
Mechanical symptoms*, no. (%)**		
Never	31 (52%)	31 (56%)
Monthly	10 (17%)	8 (15%)
Weekly	3 (5%)	2 (4%)
Several times a week	10 (17%)	10 (17%)
Daily	6 (10%)	4 (7%)
Earlier surgery same knee, no. (%)		
Yes	14 (23%)	13 (24%)
No	46 (77%)	42 (76%)
Earlier surgery opposite knee, no. (%)		
Yes	17 (28%)	16 (29%)
No	43 (72%)	39 (71%)

no.: Number, BMI: Body Mass Index, SD: Standard Deviation

*As a result of a specific incident (i.e. kneeling, sliding and/or twisting of the knee)

**As a result of a violent incident (i.e. during sports, a crash, collision or the like)

***The sensation of catching or locking of the knee.

Table 2, Test-retest scores for the reliability analysis sub-group, n=55.

	Test	Retest	Difference Test retest (95% CI)	p-value	ICC (95% CI)	SEM	MDC 95%
WOMET % score (0-100%)							
Mean (SD)	68 (22)	70 (2)	2.0 (-1.0; 4.9)	n.s.*	0.88 (0.84-0.92)	8%	22%
Median (IQR)	74 (56-84)	73 (55-91)		n.s.**			
WOMET total score (0-1600)							
Mean (SD)	512 (355)	480 (364)	31 (-17; 79)	n.s.*	0.88 (0.84-0.92)	125	347
Median (IQR)	410 (258-707)	434 (142-722)		n.s.**			
Physical symptoms (0-900)							
Mean (SD)	203 (191)	190 (187)	12 (-15; 39)	n.s.*	0.86 (0.81-0.91)	71	197
Median (IQR)	151 (60-288)	118 (41-280)		n.s.**			
Sports/recreation/work/lifestyle (0-400)							
Mean (SD)	174 (116)	168 (117)	6 (-10; 22)	n.s.*	0.87 (0.82-0.91)	42	117
Median (IQR)	179 (83-285)	152 (56-289)		n.s.**			
Emotions (0-300)							
Mean (SD)	135 (81)	123 (91)	13 (-1; 26))	n.s.*	0.83 (0.78-0.89)	35	97
Median (IQR)	126 (76-207)	130 (40-202)		0.020**			

*P-value from paired t-test
**P-value from Wilcoxon Signed Rank test