



The Dutch version of the Forgotten Joint Score : test-retesting reliability and validation

Marwan B. SHADID, Nick S. VINKEN, Louis N. MARTING, Nienke WOLTERBEEK

From the department of Orthopaedic Surgery St. Antonius Hospital Nieuwegein, The Netherlands

The aim of this study was to translate the Forgotten Joint Score (FJS) into the Dutch language. This questionnaire was tested for internal consistency (Cronbach's alpha) and test-retest reliability (intra-class correlation coefficients (ICC)). 159 patients were included in this study ; 74 with a total hip arthroplasty (THA) and 85 with a total knee arthroplasty (TKA). The FJS showed a high internal consistency (Cronbach's alpha = 0.957 ; ICC = 0.943). The FJS showed a significant correlation ($r = 0.751$) with the WOMAC and low ceiling effects (3.1%). This study proved the Dutch FJS to be highly discriminative in patients treated with a THA or TKA. This makes the FJS a reliable patient related outcome measurement, measuring a new dimension in arthroplasty : the ability to forget an artificial joint in everyday life.

Keywords : Forgotten Joint Score ; reliability ; validity ; total hip arthroplasty ; total knee arthroplasty.

INTRODUCTION

Joint arthroplasty has been proven as an effective treatment for several decades (14,11,18). Although there are many "surgeon dependent scoring systems" there are several patient dependent questionnaires which mainly focus on pain, stiffness and quality of life. With the evolution of joint arthroplasty, postoperative patient outcome are getting better and better. On the other hand, patients' demands and expectations are increasing. With

improving patient outcome, new patient reported outcome measures with increased discriminatory power especially in well-performing patients are required (16).

Recently, the Forgotten Joint Score (FJS) has been introduced, a questionnaire which focuses on the awareness of having a joint prosthesis (2). The rationale of this questionnaire is that the ultimate goal in joint arthroplasty resulting in the greatest possible patient satisfaction, is the ability to forget the artificial joint. The Forgotten Joint Score consists of 12 questions, each measuring the awareness of the artificial joint in several daily activities.

The purpose of this study is to cross-culturally adapt the Forgotten Joint Score for use in the Dutch population and to test the validity, reliability, agreement and floor and ceiling effects in a group of patients having undergone a total hip arthroplasty (THA) or a total knee arthroplasty (TKA).

■ Marwan B. Shadid, MD.

■ Nick S. Vinken, MD.

■ Louis N. Marting, MD.

■ Nienke Wolterbeek, PhD.

Department of Orthopaedic Surgery St. Antonius Hospital Nieuwegein, The Netherlands.

Correspondence : Nienke Wolterbeek Department of Orthopaedic Surgery St. Antonius Hospital P.O. box 2500 3430 EM Nieuwegein, The Netherlands.

E-mail : n.wolterbeek@antoniusziekenhuis.nl

© 2016, Acta Orthopædica Belgica.

PATIENTS AND METHODS

In this prospective study the FJS was first translated into Dutch and after that evaluated in a population of patients who received a primary THA or TKA. The protocol was approved by our institutional review board and all patients gave informed consent.

The FJS is developed by Behrend *et al* (2) in German for patients with osteoarthritis treated with an artificial joint, specifically of the hip or knee. It is a questionnaire consisting of 12 items, measuring the awareness of having a joint prosthesis during various activities of daily living. For each item there is a choice of five answers: never = 0; almost never = 1; seldom = 2; sometimes = 3; mostly = 4. All scores are summed and divided by the number of completed items. This value is then multiplied by 25 to obtain a total score range of 0 to 100. The total score is then subtracted from 100 to change the direction of the score and making a score of 100 the highest possible score, indicating the lowest level of awareness of having a joint prosthesis. A total score of 0 indicates the highest level of awareness of having a joint prosthesis. If more than 4 responses are missing, the total score should not be used (Appendix I).

The Western Ontario and McMaster Universities (WOMAC) Arthritis Index is a widely used self-assessed disease specific survey instrument for patients with lower limb osteoarthritis and has been thoroughly validated (3,7,15). It comprises 24 items in three dimensions: pain (5 items), function (2 items) and stiffness (17 items) (3). For each item there is a choice of 5 answers: never = 0; mild = 1; moderate = 2; severe = 3; extreme = 4. All scores are summed making a minimal score of 0 and a maximal score of 96 (pain 0-20; function 0-8; stiffness 0-68). The total score is subtracted from 96 and multiplied with 100 and divided by 96. Total final WOMAC scores can, therefore, vary from 0%, the lowest functional status level, to 100%, the highest functional status level.

The translation of the questionnaire was done according to the MAPI method (1). First the German FJS was translated into Dutch by two independent translators, both native Dutch speakers with a medical background. The two forward translations were merged into one forward translation. This forward translation was translated back into German by a native German speaker who lived in the Netherlands for more than 25 years. This back-translation was compared to the original questionnaire and there were no major differences, ensuring content validity. Clarity, understandability and acceptability of the FJS were tested on 5 people with THA or TKA, fol-

lowed by face to face interviews. These results were discussed, and no major changes were considered to be needed.

All patients who received a primary THA or TKA between November 2010 and May 2012 were considered for enrolment in this study. Patients were asked to complete the questionnaires if they met the following inclusion criteria: primary arthroplasty of the hip with the cemented Muller straight stem and low profile Muller cup (Zimmer) or cemented primary knee arthroplasty with a Scorpio Total knee prosthesis (Stryker) or Triathlon Total knee prosthesis (Stryker), no previous joint arthroplasty at a different location, sufficient knowledge of the Dutch language, written informed consent. If the inclusion criteria were met, the following data were collected: age (at time of surgery), gender, location of joint implant, side of surgery.

Patients were sent two FJS and two WOMAC questionnaires with an accompanying letter and were asked to give written informed consent. If willing to take part in the study they were asked to complete the first FJS and WOMAC and to complete the second FJS and WOMAC after one to two weeks. The one to two weeks test-retest interval was chosen because it was unlikely that the patient's condition would substantially change. However the time span would be large enough for patients to forget their initial responses to the questions. If no response was received, patients were reminded by telephone after 2 weeks. If no response was received after another 4 weeks, patients were excluded.

This term encompasses the internal consistency of a scale, usually measured as Cronbach's alpha, which assesses the degree of correlation among items (8). An alpha < 0.6 indicates a poor internal consistency, 0.7-0.8 acceptable, 0.8-0.9 good and > 0.9 excellent internal consistency. High internal consistency indicates a strong correlation between the items, which supports summarizing the items (21). Intraclass correlation coefficients (ICC) (two-way random effects model, single measurements and absolute agreement) were calculated for test-retest reliability (20). In general, 0.7 is recommended as a minimum standard for test-retest reliability; a correlation less than 0.5 is described as weak, whereas a correlation greater than 0.8 is described as strong (19). The 95% confidence interval (CI) for the ICC, the range of values that contains with a 95% confidence the 'true' correlation coefficient, were also calculated. Agreement was measured using the Standard Error of Measurement (SEM) calculated as $SEM_{\text{agreement}}$, which estimates the reliability of the FJS (23). The smaller the SEM, the higher the reliability and precision of the instrument. The smallest detect-

able change (SDC), based on the measurement error, was defined as $1.96 * \sqrt{2} * SEM$ (23).

Validity is the degree to which a questionnaire measures the construct it intends to measure (17). As no gold standard exists, construct validity of the FJS was assessed by determining the correlation (Spearman correlation coefficients) with an already accepted method of measurement; the WOMAC. The WOMAC is a widely used questionnaire to evaluate patients with osteoarthritis of the hip or knee. Correlation coefficients < 0.3 are considered low, between 0.3 and 0.6 moderate and > 0.6 high.

The floor and ceiling effects were determined. These were defined as the percentage of respondents who scored the minimum or maximum levels. Ideally no more than 10% should be at the bottom or top of the scale (5).

Descriptive statistics were compiled for demographic and clinical characteristics of the study population. Differences in FJS for gender or THA and TKA were tested with the independent Student t-test. Spearman correlation coefficients were used to determine correlations between continuous variables. Based on the general recommendations for comparing measurement properties, at least 50 patients need to be included (21). The aim of this study was to include 100 patients in both (THA and TKA) groups, at different postoperative moments (range 6-24 months postoperatively). All data was analyzed with SPSS statistical software (SPSS, Inc., version 21). For statistical analyses, the level of significance was set at 5%.

RESULTS

Of the 201 patients who were sent an invitation to participate, 159 complete responses were received

(response rate 79%). Reasons for not participating were: not willing to participate in the study, 1 patient died, 2 patients moved and could not be traced. From 28 patients we did not receive an answer at all. Seven patients returned the two questionnaires at the same time, and it was not clear whether they waited a week between assessments. Four patients returned only 1 set of questionnaires and didn't return the second assessment. Demographics and clinical characteristics of the study population are presented in table I. The average age was 68.6 yrs; 64% was female. 47% were treated with a total hip arthroplasty and 53% with a total knee arthroplasty. The right joint was involved in 60% of the patients. Mean postoperative time was 15 months and the mean time-interval between the first and second assessment was 10 days (Table I).

The translated FJS showed a high internal consistency with a Cronbach's alpha of 0.957 (Table II). Test-retest reliability were recorded at a mean interval of 10 days (range 1 to 49 days). Test-retest reliability was measured and showed a very good result with a ICC of 0.943 (95% CI 0.922-0.958) (Table II). Seven patients returned the two questionnaires at the same time, and it was not clear whether they waited a week between assessments. Therefore, these patients were excluded for the test-retest analysis. Four patients returned only 1 pair of questionnaires and didn't return the second assessment. These patients were also excluded for test-retest analysis. SEM and SDC of the FJS for the total group and the THA and TKA separately varied

Table I. — Patient characteristics at the first administration of the questionnaires for the total group and for the total hip arthroplasty group (THA) and total knee arthroplasty group (TKA) separately. Mean data is presented as well as standard deviation and range

	THA (N = 74)	TKA (N = 85)	Total group (N = 159)
Female / male	49 / 25	53 / 32	102 / 57
Operated side (right/left)	52 / 22	43 / 42	95 / 64
Age (years)	68.6 (8.5; 48-85)	68.5 (9.2; 39-91)	68.6 (8.9; 39-91)
Time since surgery (months)	15.6 (5.5; 7-24)	14.5 (6.2; 6-24)	15.0 (5.9; 6-24)
Interval T0 – T1 (days)	10.0 (4.9; 1-35)	10.0 (6.4; 4-49)	10.0 (5.7; 1-49)
Forgotten Joint Score	56.1 (27.1; 4-100)	38.1 (30.3; 0-100)	46.5 (30.2; 0-100)
WOMAC – Total	81.2 (20.4; 6-100)	71.3 (21.0; 23-100)	75.8 (21.2; 6-100)
WOMAC – Pain	85.4 (21.3; 0-100)	75.5 (24.4; 25-100)	80.1 (23.4; 0-100)
WOMAC – Stiffness	76.2 (21.1; 25-100)	66.0 (23.6; 13-100)	70.7 (23.0; 13-100)
WOMAC – Function	80.6 (21.2; 6-100)	70.5 (21.5; 19-100)	75.1 (21.9; 6-100)

Table II. — Cronbach's Alpha, Intraclass Correlation Coefficients (ICC), Standard error of Measurement (SEM) and Smallest Detectable Change (SDC) of the Dutch version of the FJS to evaluate reliability and agreement in a group of patients having a total hip arthroplasty (THA) or total knee arthroplasty (TKA). Cronbach's Alpha is calculated for first administration. SEM and SDC are given as points on a scale of 0-100

	Cronbach's Alpha	ICC (95% CI)	SEM	SDC
Total group	0.957	0.943 (0.922 – 0.958)	7.2	20.0
THA	0.942	0.924 (0.881 – 0.952)	7.6	21.1
TKA	0.961	0.947 (0.918 – 0.966)	6.9	19.0

between respectively 6.9-7.2 and 19.0 – 21.1 points (Table II).

There was a significant positive correlation between the FJS and the WOMAC total score ($r = 0.751$; $p < 0.001$). Correlation with the WOMAC subscale pain was the highest and the lowest correlation was found with the WOMAC subscale stiffness (Table III).

The FJS showed lower ceiling effects than the WOMAC. The FJS showed a ceiling effect of 3.1% versus 5.7% for the WOMAC (Table IV). We found some missing items especially question 20 of the WOMAC (8.8%), concerning getting in or out of the bathtub. We also found that quite a lot (13.8%) of the patients didn't fill in question 12 of the FJS concerning sports activity.

Patients who received a THA showed higher scores ($p < 0.001$) on the FJS than patients who have undergone TKA in the whole postoperative period (respectively 56.1 versus 38.1 points). No difference was found in FJS score between men and women ($p = 0.244$). Also no association was found between age and scores on the FJS ($r = 0.143$; $p = 0.073$). For the THA we observed that the longer the postoperative period, the higher the scores on the FJS ($r = 0.377$; $p = 0.001$). For TKA, no cor-

relation was found between FJS and time since surgery ($r = 0.098$; $p = 0.374$).

DISCUSSION

A wide range of questionnaires regarding surgeon related outcomes, stiffness and pain are commonly used. With improving patient outcome, new patient reported outcome measures with increased discriminatory power especially in well-performing patients are required (16). Furthermore, short questionnaires are known to improve patient compliance and response rates (6,12,13). The FJS is a short, relatively new questionnaire which measures the patient's ability to forget having an artificial joint. Translations and validation studies are needed for international use and comparison of results. This paper shows the step-by-step creation of a conceptually equivalent version of the Dutch FJS by following the MAPI methodology.

For internal consistency the Dutch version of the FJS showed a high Cronbach's alpha (0.957), which was comparable to the Cronbach's alpha of 0.95 found by the original authors (2). The high Cronbach's alpha implies that the FJS has a very high level of interrelatedness among the items and there is no redundancy among the items. The Dutch FJS is showing strong test-retest reliability with a ICC 0.943, indicating low variability in measurements.

Although the WOMAC has shown high ceiling effects, the WOMAC has been a widely used patient related outcome measurement for years in patients with a THA or TKA. Behrend *et al* (2) shows that the FJS has much lower ceiling effects (9.2%) than the WOMAC (ranging from 16.7% till 46.7% for the total score and sub scales). Our data confirms this with a ceiling effect of 3.1% for the FJS,

Table III. — Spearman's Correlations Coefficient between FJS and WOMAC (N = 159)

	Correlation coefficients (r)
WOMAC – Total	0.751*
WOMAC – Pain	0.739*
WOMAC – Stiffness	0.670*
WOMAC – Function	0.717*

Correlation is significant at the 0.001 level (2-tailed).

Table IV. — Ceiling and floor effect and missing items for FJS and WOMAC

	Ceiling effect	Floor effects	Missing items
FJS	5 (3.1%)	6 (3.8%)	1.7%
WOMAC – Total	9 (5.7%)	0 (0%)	1.7%
WOMAC – Pain	49 (30.8%)	1 (0.6%)	1.5%
WOMAC – Stiffness	32 (20.1%)	0 (0%)	0.9%
WOMAC – Function	11 (6.9%)	0 (0%)	1.8%

although ceiling effects of the WOMAC in our study are much lower (5.7%). This indicates that the FJS has a higher power to discriminate patient with a high score.

With a Spearman correlation coefficient of 0.751 the FJS showed a high correlation with the WOMAC, which means that they measure the same construct. Patients having undergone a THA showed higher scores (mean 56.1) on the FJS than patients having undergone TKA (mean 38.1) in the whole postoperative period. This was previously shown by Behrend *et al* (2), although both THA and TKA groups in our study scored lower. Behrend *et al* (2) found in their healthy control group a lower score for the hip than the knee as well, suggesting that the difference in scores between hip and knee may not be contributed to the arthroplasty effect. This may indicate that people are more aware of their knee joint than of their hip joint whether or not having undergone an arthroplasty.

For THA we see a increasing score related to time since surgery till 24 months. In TKA we see increasing scores till 18 months. After that period the scores seems to decline. An important note hereby is that patients were measured only once in time and a longitudinal study is needed to determine real changes over time. Behrend *et al* (2) found no differences in scores related in time with patients between 15 en 58 months postoperative. Fitzgerald *et al* (9) showed that improvement after THA improved till it reached an end at 12 months.

We found 13.8% of the patients not replying to question 12 of the FJS concerning awareness of the joint during sports activity. This is the same as found by Thienpont *et al* (22), who showed that especially question 11 and 12 of the FJS were not answered by the elderly and female patients. This assumes that this question is not appropriate for a

group of patients condition after having undergone an arthroplasty. Geisinger *et al* (10) developed a computer-adaptive test version of the FJS, using individually tailored sets of questions to reduce the number of questions administered to each patient. This showed that for some patients the FJS could be reduced by half at comparable measurement precision.

Limitations of this study are that we sent all questionnaires at once and asked the patients to assess the questionnaires with one tot two weeks in between. For the test-retest analysis, 7 patients who replied all questionnaires at the same day needed to be excluded. Four patients returned only 1 pair of questionnaires and didn't return the second assessment. Also there was a wide range (1-49 days) between assessing the test and retest questionnaires. We did not have a preoperative measurement for both the FJS and the WOMAC and didn't have a healthy control group. That makes it difficult to make an interpretation of our scores which were lower than Behrend *et al* (2) and Thienpont *et al* (22) found in their groups. By picking different patients at different periods postoperative we observed higher scores for patients longer postoperative then for patients shorter after arthroplasty. A longitudinal study is needed to determine real changes over time. Hopefully the questionnaire will be translated into several languages, creating the possibility to compare results internationally.

The results of this study show that the Forgotten Joint Score is a reliable and valid questionnaire and thereby a valuable questionnaire to evaluate how a joint prosthesis is integrated in a patient life and measure the awareness of an artificial joint. The low ceiling effects found in this study indicates high discriminative power between patients performing well and patients performing very well.

REFERENCES

1. **Acquadro C, Jambon B, Ellis D et al.** *Language and translation issues*. 2nd ed, Lippincott-Raven, Philadelphia ; 1996.
2. **Behrend H, Giesinger K, Giesinger JM et al.** The "Forgotten Joint" as the ultimate goal in joint arthroplasty. *J Arthroplasty* 2012 ; 3 : 27.
3. **Bellamy N, Buchanan WW.** A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol* 1986 ; 5 : 231.
4. **Bellamy N, Buchanan WW, Goldsmith CH et al.** 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* 1988 ; 15 : 1833.
5. **Bridwell KH, Cats-Baril W, Harrast J et al.** The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12 : a study of response distribution, concurrent validity, internal consistency, and reliability. *Spine* 2005 ; 30 : 455.
6. **Burchell B, Marsh C.** The effect of questionnaire length on survey response. *Quality and Quantity* 1992 ; 26 : 233.
7. **Chesworth BM, Mahomed NN, Bourne RB et al.** Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol* 2008 ; 61 : 907.
8. **Cronbach LJ, Warrington WG.** Time-limit tests : estimating their reliability and degree of speeding. *Psychometrika* 1951 ; 16 : 167.
9. **Fitzgerald JD, Orav EJ, Lee TH et al.** Patient quality of life during the 12 months following joint replacement surgery. *Arthritis Rheum* 2004 ; 51 : 100.
10. **Geisinger JM, Kuster MS, Holzner B et al.** Development Of A Computer-Adaptive Version Of The Forgotten Joint Score. *J Arthroplasty* 2013 ; 3 : 28.
11. **Hawker G, Wright J, Coyte P et al.** Health-related quality of life after knee replacement. *J Bone Joint Surg Am* 1998 ; 80 : 163.
12. **Herzog AR, Bachman JG.** Effects of questionnaire length on response quality. *The public opinion quarterly* 1981 ; 45 : 549.
13. **Kalantar JS, Talley NJ.** The effects of lottery incentive and length of questionnaire on health survey response rates : a randomised study. *J Clin Epidemiol* 1999 ; 52 : 1117.
14. **Keener JD, Callaghan JJ, Goetz DD et al.** Long-term function after Charnley total hip arthroplasty. *Clin Orthop Relat Res* 2003 ; 417 : 148.
15. **Martin DP, Engelberg R, Agel J et al.** Comparison of the Musculoskeletal Function Assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. *J Bone Joint Surg Am* 1997 ; 79 : 1323.
16. **Marx RG, Jones EC, Atwan NC et al.** Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. *J Bone Joint Surg Am* 2005 ; 87 : 1999.
17. **Mokkink LB, Terwee CB, Patrick DL et al.** The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010 ; 63 : 737.
18. **Noble PC, Conditt MA, Cook KF et al.** The John Insall Award : patient expectations affect satisfaction with total knee arthroplasty. *Clin Orthop Relat Res* 2006 ; 452 : 35.
19. **Nunnally JC, Bernstein IH.** *Psychometric theory*, 3rd edn. New York : McGraw-Hill ; 1994.
20. **Shrout PE, Fleiss JL.** Intraclass correlations : uses in assessing rater reliability. *Psychol Bull* 1979 ; 86 : 420.
21. **Terwee CB, Bot SD, de Boer MR et al.** Quality criteria were proposed for measurement properties of health status questionnaires. *J Clinical Epidemiology* 2007 ; 60 : 34.
22. **Thienpont E, Opsomer G, Koninckx A et al.** Joint Awareness in Different Types of Knee Arthroplasty Evaluated With the Forgotten Joint Score. *J Arthroplasty* 2013 ; 29 : 48.
23. **de Vet HC, Terwee CB, Knol DL et al.** When to use agreement versus reliability measures. *J Clinical Epidemiology* 2006 ; 59 : 1033.

Appendix I : **FORGOTTEN JOINT SCORE**

De volgende 12 vragen hebben betrekking op hoe bewust u bent van uw kunstheup / knie in het dagelijks leven. Graag 1 antwoord per vraag aankruizen.

Bent u zich bewust van uw kunstgewricht....

1. ... 's nachts in bed ?

nooit bijna nooit zelden soms meestal

2. ... wanneer u langer dan 1 uur op een stoel zit ?

nooit bijna nooit zelden soms meestal

3. ... wanneer u langer dan 15 minuten loopt ?

nooit bijna nooit zelden soms meestal

4. ... wanneer u een douche / bad neemt ?

nooit bijna nooit zelden soms meestal

5. ... wanneer u per auto reist ?

nooit bijna nooit zelden soms meestal

6. ... wanneer u trap loopt ?

nooit bijna nooit zelden soms meestal

7. ... wanneer u op oneffen terrein loopt ?

nooit bijna nooit zelden soms meestal

8. ... wanneer u omhoog komt vanuit een lage positie ?

nooit bijna nooit zelden soms meestal

9. ... wanneer u gedurende langere tijd staat ?

nooit bijna nooit zelden soms meestal

10. ... wanneer u huishoudelijk werk verricht / tuiniert ?

nooit bijna nooit zelden soms meestal

11. ... wanneer u een langere wandeling maakt ?

nooit bijna nooit zelden soms meestal

12. ... wanneer u uw favoriete sport uitoefent ?

nooit bijna nooit zelden soms meestal

Scoring :

voor het scoren van de FJS-12 worden alle antwoorden opgeteld (nooit, 0 punten ; bijna nooit, 1 punt ; zelden, 2 punten ; soms, 3 punten ; meestal, 4 punten) en vervolgens gedeeld door het aantal beantwoorde vragen. Deze gemiddelde waarde wordt vervolgens vermenigvuldigd met 25 om tot een score tussen 0 en 100 te komen. Tenslotte wordt de score afgetrokken van 100 om de richting van de eindstand zodanig te veranderen, dat een hoge score een hoge mate van "vergeten" weergeeft, dat wil zeggen een lager bewustzijn van het kunstgewricht.

Indien meer dan 4 antwoorden ontbreken, dan kan de totaal score niet gebruikt worden.