Master's Thesis of Medicine

Translation and Validation of the Korean Version of the Forgotten Joint Score

한국어판 슬관절 설문지 Forgotten Joint Score(FJS)의 타당성 검사

August 2021

Graduate School of Medicine
Seoul National University
Orthopaedics Major

Jangyun Lee

Translation and Validation of the Korean Version of the Forgotten Joint Score

- 한국어판 슬관절 설문지 Forgotten Joint Score의 타당성 검사 -

지도교수 이 명 철

이 논문을 의학석사 학위논문으로 제출함
2021년 4월

서울대학교 대학원
정형외과 전공
이 장 연

이장연 의 석사 학위논문을 인준함
2021년 7월

위 원 장 정 선 근

부위원장 이 명 철

위 원 한 혁 수

Abstract

Jangyun Lee
Orthopaedics
The Graduate School
Seoul National University

Background: Forgotten Joint Score (FJS) is a newly developed patient reported outcome measure (PROM) designed to evaluate clinical outcome after total knee arthroplasty (TKA). FJS is known as a sensitive test with low ceiling effect. It has been recently translated into many languages. However, no study has reported the validity or reliability of the Korean Version of FJS. Thus, the purpose of this study was to address this issue.

Method: According to guidelines for cross-cultural adaptation, translation of the English version of the FJS was performed. After obtaining license from the original developer, 150 patients who had undergone total knee arthroplasty at more than one year ago to less than five years completed the Korean version of FJS (K-FJS), visual analogue scale (VAS), Western Ontario McMaster Universities Osteoarthritis index (WOMAC), and the 36-Item Short Form Health Survey (SF-36). To measure test-retest reliability, K-FJS was completed twice by telephone survey for an additional 100 patients. Responsiveness was retrospectively calculated based on a survey of 50 patients at three months and one year after surgery.

Results: The Korean version FJS exhibited an excellent reliability [Cronbach's α = 0.967, Intraclass correlation coefficient = 0.958,

95% CI: 0.930-0.974]. The ceiling effect of K-FJS was 8.7% (N = 13). There was no floor effect. This was lower than that of WOMAC's ceiling effect (10%). Its correlation coefficients with WOMAC and SF-36 (Physical Function) were 0.708 and 0.682, respectively, indicating good construct validity. However, its correlation with mental health subscale of SF-36 was low (r = 0.143). At 3 to 12 months after TKA, standardized response mean (SRM) was 0.67, which was lower than the SRM of WOMAC (1.03) obtained in the same period. The Korean version of Forgotten Joint Score demonstrated strong measurement properties in terms of good construct validity and reliability.

Conclusions: This study suggests that the Korean version of Forgotten Joint Score is an excellent instrument that can be used to monitor clinical outcomes after total knee arthroplasty. Using this standardized version of K-FJS, it would be possible for institutions to share clinical results more accurately.

Keyword: Knee; Osteoarthritis; Total knee arthroplasty; Forgotten joint score; Patient reported outcome measure

Student Number : 2018-20443

Table of Contents

Chapter 1. Background

Total knee arthroplasty (TKA) is widely accepted as standard therapy for end-stage degenerative osteoarthritis as it can provide pain relief and functional improvements.[1] Postoperative outcome of TKA is often assessed based on objective surgeons' ratings rather than patients' subjective satisfaction. A gap between the two has led to the concept of patient-reported outcome measures (PROMs).[2] Various tools such as Health-related quality of life (HRQoL) questionnaires have been developed to evaluate patient-centered clinical results.[2] Western Ontario McMaster Universities Osteoarthritis index (WOMAC) was the first PROM developed in the early 1980's to assess patients with hip and knee osteoarthritis.[3] After rigorous validation, WOMAC has been translated into more than 60 languages and used in many clinical studies related to the knee.[3] With improvement of surgical tools, implants materials, and techniques, average scores by the WOMAC and commonly used questionnaire are increasing. Many patients are now receiving the maximum scores on WOMAC and Oxford knee score, indicating the presence of ceiling effect.[4] In an attempt to reduce this ceiling effect, Behrend et al.[5] proposed a new disease-specific PROM known as the Forgotten Joint Score (FJS) in 2012. FJS can be sued to assess post-arthroplasty joint awareness using 12 equally weighted questions consider patients' ability to forget the artificial joint in everyday life as an ultimate outcome of arthroplasty. Recent studies have demonstrated that the FJS has higher reliability, better validity, and lower ceiling effect than the WOMAC score.[5] First developed in Switzerland, the FJS has been translated into many languages (including English, German, Japanese, and Danish) and successfully

validated.[6) 7) 8) 9) 10)] Recently, Adriani et al.[11)] have performed a systemic review focusing on the utility of the FJS and demonstrated that it has good construct validity and test-retest reliability. In Korea, many studies have individually analyzed the FJS score along with other PROMs. For instance, Kim et al.[12)] have reported that FJS score is higher in a unicompartmental knee arthroplasty (UKA) group than in a TKA group. However, there is no consensus form for the Korean version of FJS. A validation study or a cross-cultural adaptation study of FJS in Korea has not been reported. Creation and validation of a unified Korean version of FJS hold a promise for its widespread application as a PROM tool for TKA patients. We believe that cross-cultural adaptation and conceptual equivalence are crucial to develop a Korean version of FJS. Thus, the aim of this study was to develop a Korean version of FJS (K-FJS) that would be equally natural and acceptable as the original version. We further investigated the reliability, validity, and responsiveness of the K-FJS questionnaire to be used as a PROM tool in Korea.

Table 1. Original version of Forgotten Joint Score-12 score questions [5)]

Questions
Are you aware of your artificial joint …
1. … in bed at night?
2. … when you are sitting on a chair for more than 1 hour?
3. ... when you are walking for more than 15 minutes?
4. … when you are taking a bath/shower?
5. … when you are traveling in a car?
6. … when you are climbing stairs?
7. … when you are walking on uneven ground?
8. … when you are standing up from a low-sitting position?
9. … when you are standing for long periods of time?
10. … when you are doing housework or gardening?
11. … when you are taking a walk/hiking?
12. … when you are doing your favorite sport?

Chapter 2. Methods

(1) Translation and Cross-cultural adaptation

Translation and cross-cultural adaptation were proceeded in five steps according to guidelines of Guillemin et al. and Wild et al. [13) 14)]

1) Forward translation

The English version of FJS was translated by two independent bilingual translators fluent in both English and Korean.

2) Synthesis of the translation

The two translated FJS questionnaires were merged into a single survey form. The merging process was focused on conceptual translation rather than literal translation. The final survey process was revised to convey acceptable language for the broadest audience.

3) Backward translation

The questionnaire was translated back to English by two independent health professionals with English as their mother tongue. The two translators are familiar with terminologies of the area covered by this study.

4) Expert committee review including the Licensor

A bilingual expert panel including the inventor of the questionnaire, original translators, and researchers of this study was convened to identify and resolve any inadequate expression or concept of the translation. The goal was to minimize any discrepancies between the final translation and the original version of questions. Both the licensor and the licensee agreed to the revised version of K-FJS.

5) Confirmation and Pretesting

The pre-final version of K-FJS was tested on 20 patients who underwent TKA. Respondent debriefing questions included what they thought about those questions and whether they could repeat those

questions in their own words.

(2) Validation study

<Reliability>

Reliability refers to the degree to which a measurement tool is repeatable regardless of time and that the tool can achieve consistent results. When measurement error is decreased, reliability is increased. We tested reliability by measuring whether the test was consistent across time (test-retest reliability) and across items (internal consistency). Internal consistency was measured using Cronbach's alpha as an index of whether items in one measurement tool were closely related to each other. Generally, a Cronbach's alpha of 0.80 and above indicates good internal consistency and a Cronbach's alpha of 0.90 and above indicates excellent internal consistency. [15] To measure test-retest reliability, a telephone questionnaire was conducted once again at two weeks to a month after the first survey. The time period was selected to be not too long so that the postoperative status was not changed. In the meantime, the time period was selected to be not too short so that patients could not recall previous questionnaire. Intraclass correlation coefficient (ICC) was calculated to estimate test-retest reliability.[15]

<Validity>

Validity refers to whether a measurement tool can accurately measure what the researcher intends to measure. Two general forms of validity, construct validity and content validity, were assessed for the K-FJS. Construct validity is the degree to which an instrument measures the trait or theoretical construct that it is intended to

measure. To assessing the construct validity, it was hypothesized that the K-FJS score would have a moderate to strong positive correlation with other PROM scoring systems (i.e., WOMAC and SF-36). Pearson's correlations coefficient was also calculated. Values greater than 0.6 indicated a strong correlation.[16] Content validity expresses how well the questionnaire covers all symptoms experienced by patients. The content validity was assessed for floor and ceiling effects. Floor effect includes the proportion of patients scoring the lowest possible, whereas ceiling effect expresses patients scoring the highest possible. Floor and ceiling effects of less than 15% of patients were considered to be adequate.[17]

<Responsiveness>

Responsiveness measures sensitivity to changes within patients over time. High responsiveness means that the measurement tool is more sensitive in detecting changes within patients over time. Responsiveness to change was assessed using the standardized response mean (SRM) for the change between the 3-month follow-up time point and the 12-month follow-up time point. SRM was calculated as the average difference between two measurements divided by the standard deviation of differences between paired measurements, with higher SRM indicating greater responsiveness. According to the Cohen criteria, SRM of greater than 0.8, SRM of 0.5 to 0.8, and SRM of 0.2 or less indicate large, moderate, and small changes, respectively. [18] [19] [20]

(3) Patient selection

We retrospectively reviewed clinical databases at Seoul National University Hospital for degenerative osteoarthritis patients who had

undergone primary total knee arthroplasty between January 2013 and December 2018. We identified 150 Korean-speaking patients whose average follow-up period was between 1 year and 5 years. We conducted a retrospective survey using the Korean version of FJS-12, Visual Analog Scale (VAS), WOMAC, SF-36, and Knee society score. For 100 patients, telephone questionnaire using FJS-12 was conducted once again at 3 weeks to a month after the first survey. In addition, among 150 patients, 50 patients with postoperative records of 3-month and 1-year follow-ups were compared. Score changes of VAS, WOMAC, and Korean version FJS-12 were compared to investigate responsiveness over time. Patients with rheumatoid arthritis, post traumatic osteoarthritis, history of previous knee surgery, history of severe trauma or uncontrolled systemic disorders were excluded.

To minimize any difference in the level of understanding according to each patient's education level, a specialized orthopedic physician assistant was assigned to help patients understand exact meanings of questionnaires before filling out. Both the interview-based survey and telephone survey were conducted by the same physician assistant to minimize error variance. Nonetheless, all questions were read by patients themselves. Intervention was minimized once the filling out process began. During this process, patients who could not read or understand Korean fluently were excluded from the analysis. General demographics of the patient population are summarized in Table 2.

Table 2. Demographic and clinical characteristic of patients

Demographics	Value
Age (years)	71.3 ± 6.5
Range	57-84
Gender	
Female	124 (82.7%)
Male	26 (17.3%)
Side	
Right	74 (49.3%)
Left	76 (50.7%)
Time after surgery (months)	20 ± 12
Range	12 - 60

Data are shown as mean ± standard deviation or number (%).

(4) Statistical analysis

Study population was determined based on the standards proposed by Terwee et al.[17] A minimum of 100 patients were required for internal consistency analysis and a minimum of 50 patients were needed for analyzing floor or ceiling effects, reliability, and validity. All statistical analyses were conducted using SPSS for Windows version 23.0 (SPSS Inc., Chicago, IL, USA). Pearson's correlation coefficient for individual scoring system was calculated to study construct validity. Cronbach's alpha value, ICC, and SRM were used to determine internal consistency, test-retest reliability, and responsiveness, respectively. 95% confidence intervals (CIs) were provided. Statistical significance was considered at p-value < 0.05.

Chapter 3. Results

슬관절 질문지 (Forgotten Joint Score – 12)

환자 : _____ 날짜 : ___.___._____

귀하는 일상생활에서 건강한 관절을 인지하지 않고 지내게 됩니다. 하지만 아주 작은 문제라도 생기게 되면 관절을 인지하게 됩니다. 이것은 귀하가 관절을 생각하거나 그것에 주의를 기울인다는 것을 의미합니다. 다음 질문은 일**상생활에서 문제가 있는 슬관절을 얼마나 자주 인지하는지**에 관한 것입니다.

각 질문에 가장 알맞은 답을 선택하십시오.

	귀하는 다음 상황에서 귀하의 슬관절에 대해 인지하고 있습니까?	**전혀 인지하고 있지 않다**	**거의 인지하고 있지 않다**	**드물게 인지하고 있다**	**종종 인지하고 있다**	**거의 인지하고 있다**
1.	밤에 침대에 누워있을 때	O	O	O	O	O
2.	한 시간 이상 의자에 앉아있을 때	O	O	O	O	O
3.	15분을 넘게 걸을 때	O	O	O	O	O
4.	샤워하거나 목욕할 때	O	O	O	O	O
5.	차를 타고 여행할 때	O	O	O	O	O
6.	계단을 오를 때	O	O	O	O	O
7.	울퉁불퉁한 길을 걸을 때	O	O	O	O	O
8.	낮게 앉은 자세에서 일어날 때	O	O	O	O	O
9.	오랫동안 서 있을 때	O	O	O	O	O
10.	집안일을 하거나 정원일을 할 때	O	O	O	O	O
11.	산책하거나 가벼운 등산을 할 때	O	O	O	O	O
12.	좋아하는 운동을 할 때	O	O	O	O	O

Fig. 1. Korean version of Forgotten Joint Score

Scoring: for scoring the FJS-12, all responses are summed (never, 0 points; almost never, 1 point; seldom, 2 points; sometimes, 3 points; mostly, 4 points) and then divided by the number of completed items. This mean value is subsequently

multiplied by 25 to obtain a total score range of 0 to 100. Finally, the score is subtracted from 100, to change the direction of the final score in a way that high scores indicate a high degree of "forgetting" the artificial joint, that is, a low degree of awareness. If more than 4 responses are missing, the total score should not be used. [5)]

Table 3. Measurement properties of the Korean version FJS

Psychometric properties	Value (*p*-value)
Validity	
Construct validity†	
WOMAC score	0.708 (< 0.001)*
Knee Society Score	
Knee score	0.258 (< 0.001)*
Function score	0.889 (< 0.001)*
SF-36	
General health	0.149 (0.074)
Physical function	0.682 (< 0.001)*
Role physical	0.373 (< 0.001)*
Emotion physical	0.390 (< 0.001)*
Bodily pain	0.579 (< 0.001)*
Vitality	0.073 (0.385)
Social function	0.597 (< 0.001)*
Mental health	0.143 (0.085)
Content validity	
Ceiling effect	8.7% (N=13)
Floor effect	No
Reliability	
Internal consistency	
Cronbach's alpha	0.967
Test-retest reliability	
Intraclass correlation coefficient	0.958 (0.930-0.974) ‡
Responsiveness	

Standardized response mean (SRM)	0.67

*Asterisks indicate statistically significant associations (p<0.05)

†Pearson's correlation coefficient

‡95% CI, $p < 0.001$

Table 4. Ceiling and floor effects of K-FJS and WOMAC

	Ceiling effect	Floor effect
K-FJS	13 (8.7%)	No
WOMAC	15 (10%)	No

Table 5. Responsiveness of VAS, K-FJS, and WOMAC scores

	Mean of change	SD	SRM
VAS (0-10)	0.70	2.54	0.28
K-FJS	-10.64	15.98	0.67
WOMAC Score	11.44	11.11	1.03

SD, standard deviation; SRMs = (mean postoperative score–mean preoperative score)/standard deviation of the change in score; VAS, visual analog scale

Table 6. Pearson's correlation coefficient between K-FJS and WOMAC subscale

	Ceiling effect	Floor effect	Correlation Coefficients (r)
WOMAC (total)	15 (10%)	No	0.708*
WOMAC Pain	50 (33%)	No	0.561*
WOMAC Stiffness	50 (33%)	No	0.420*
WOMAC Physical Function	21 (14%)	No	0.649*

*$p < 0.001$

The Korean version FJS exhibited an excellent reliability [Cronbach's α = 0.967, ICC = 0.958, 95% CI: 0.930-0.974] (Table 3). The ceiling effect of K-FJS was 8.7% (N = 13), which was lower than that of WOMAC's ceiling effect (10%). There was no floor effect

(Table 4). Its correlation coefficients with WOMAC and SF-36 (Physical Function) were 0.708 and 0.682, respectively, indicating good construct validity (Fig. 2 and 4). At 3 to 12 months after TKA, SRM was 0.67, which was lower than the SRM of WOMAC (1.03) obtained in the same period (Table 5). Compared to WOMAC subset scores, K-FJS had a high correlation with the pain (r = 0.561) and the physical function (r = 0.649) subscales and moderate correlation with the stiffness (r = 0.420) subscale (Table 6). However, its correlation with the mental health subscale of SF-36 was low (r = 0.143). Besides, although not PROMs, function score of Knee society score showed a strong correlation with K-FJS (r = 0.889) (Fig. 4).

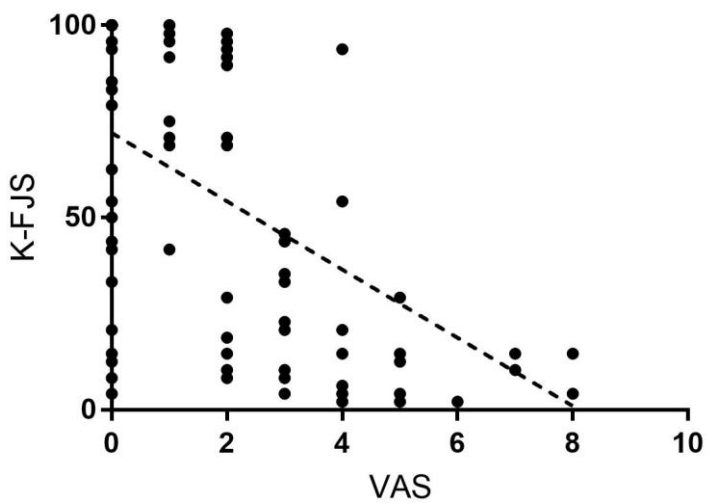

Fig. 2. Correlation of Korean version of Forgotten Joint Score and Visual Analog Scale

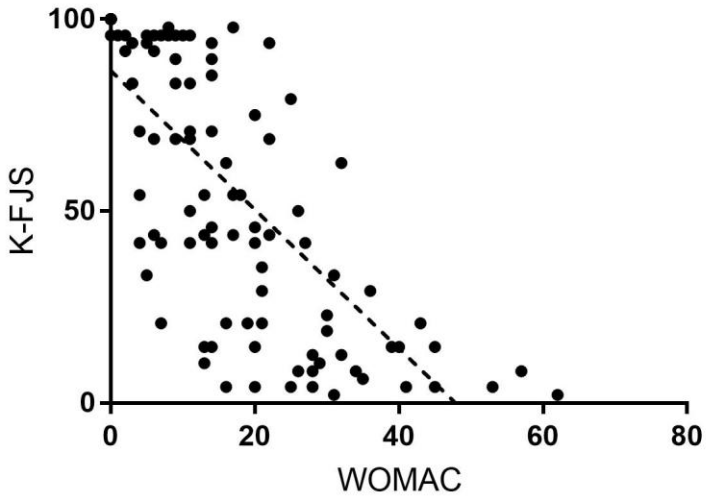

Fig. 3. Correlation of Korean version of Forgotten Joint Score and Western Ontario McMaster Universities Osteoarthritis index

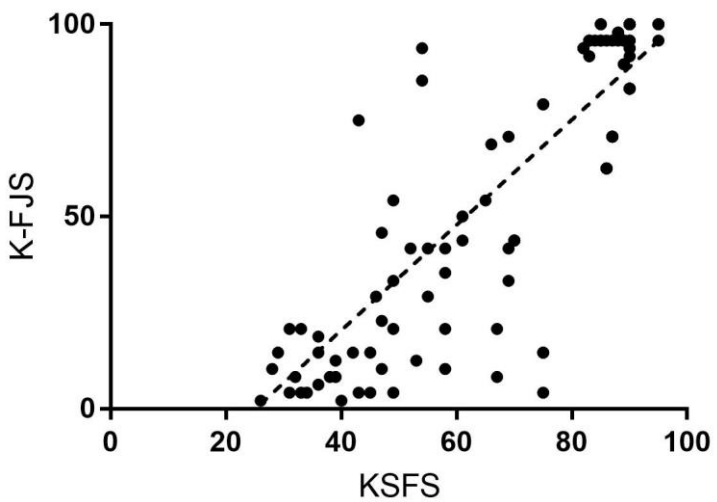

Fig. 4. Correlation of Korean version of Forgotten Joint Score and Function Score of Knee Society Score

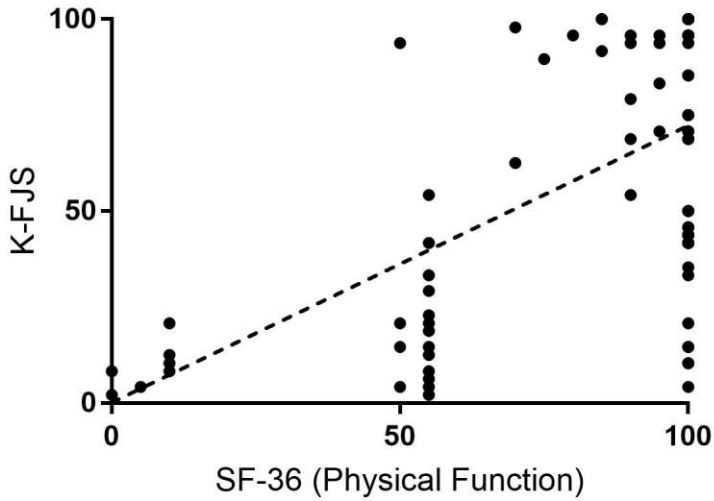

Fig. 5. Correlation of Korean version of Forgotten Joint Score and Physical Function Subscale of 36-Item Short Form Health Survey

Chapter 4. Discussion

To minimize any difference in the level of understanding according to each patient's education level, researchers were assigned to help patients understand exact meanings of questionnaires before filling out. We believe this process is a good way to avoid half-hearted consistent answers or missing answers due to the lack of understanding. It also helps researchers to figure out what to improve during the early stage of the pilot study.

During the pilot study, many patients asked about the exact meaning of "Are you aware of your artificial joint". A brief explanation was added to make sure that patients understood it as "how much are you aware of". Afterwards, questions about confusing sentence were largely reduced. Also, as mentioned in the study of Cao et al.[21], many patients mistakenly considered that items in prefinal K-FJS were asking the frequency they were able to finish a corresponding activity after TKA. Thus, we added the term "aware of" to each item in bold to minimize misunderstanding of questionnaires.

A systemic review on FJS using all articles that reported missing response percentage, question number 12 ("Are you aware of your artificial knee when doing your favorite sport?") had a significant high missing response (> 10%). Likewise, we found a significant high response saying that they had no favorite sport or it had been too long since they quitted exercise. [7) 9) 21)] However, because we excluded all cases with any missing item, the exact missing response rate was not assessed. We believe this requires further supplement.

The validity study revealed a good correlation between the K-FJS and the WOMAC total score ($r = 0.708$, $p < 0.001$) (Fig. 3). This was

similar to earlier cross-cultural studies reporting a high correlation between the FJS and the WOMAC score (Behrend et al.[5], r = 0.79; Shadid et al.[9] r = 0.75). Some studies have also shown a less significant result. The study conducted by Matsumoto et al.[8] showed a moderate correlation (r = 0.52) between Japanese version of FJS and WOMAC score.

Compared to WOMAC subset scores, K-FJS had a relatively weak correlation with the stiffness subscale (r = 0.420) (Table 6). Similar results have been previously reported in a Japanese study by Matsumoto et al. (r = 0.4) and the original version study by Thompson et al. (r = 0.52). The WOMAC is composed of 24 items with 3 subscales (17 for physical function, 5 for pain, and 2 for stiffness). Two stiffness subscale questionnaires are "Stiffness after waking up in the morning" and "Stiffness after sitting/lying or resting during the day". However, there was no question in the original version of FJS that specifically evaluated stiffness. The difference in questionnaire contents might have resulted in this weak correlation. Moreover, of a total of 96 points, the stiffness subscale was assigned with a maximum of 8 points. Due to this small range, we believe that the stiffness subscale would not have affected much on the correlation between K-FJS and the total WOMAC.

In addition, K-FJS showed good correlation with another PROM, KSFS (r = 0.886). Unlike KSFS or WOMAC physical function score which asks about whether the actual knee function is good, K-FJS questionnaire focuses on how well each patient is aware of his or her own knee. Such difference resulted in outliers presented in Figure 3 and 4. Patients with relatively sensitive personality feel uncomfortable even with the slightest sound normally generated from an artificial joint, whereas patients with blunt personality forget

about the artificial joint even when it shows moderate instability.

FJS could be affected by differences in individual personality. However, there is lack of study comparing FJS with central sensitization score which evaluate hyperalgesia or allodynia derived from the central nervous system. In addition, it would be meaningful to study whether the FJS score could be improved by using drugs such as pregabalin or duloxetine, which are known to reduce neuropathic pain.

Besides, the K-FJS had moderate correlations with SF-36 subscales of pain, physical function, and social function. This is comparable to the validation result of a Chinese version FJS presented by Cao et al.[21] Concerning these results and its correlations with WOMAC pain and WOMAC physical function subscales, the K-FJS showed a good convergent validity. In addition, the K-FJS showed low correlation with SF-36 mental health subscale ($r = 0.143$). This reflects good discriminant validity, as highlighted in the English and Chinese version of FJS ($r = 0.23$ and $r = 0.086$, respectively)

No floor effect was observed in the total score of K-FJS and WOMAC. We found a ceiling effect of 8.7% for the K-FJS as compared to 10% for the WOMAC (Table 4). Because the WOMAC questionnaire is composed of twice as many items as the FJS, it is harder for the WOMAC to have a ceiling effect. Considering that K-FJS has less question numbers, we can conclude that the K-FJS has a lower ceiling effect. Earlier studies have presented that floor and ceiling effects lower than 15% are required for a study to have a reasonable content validity and that effects lower than 10% are considered ideal.[17] The ceiling effect of K-FJS was 8.7%, meaning an ideal content validity. This result is better than findings presented

in the first FJS study by Behrend et al.[5] (ceiling effect for the FJS and the WOMAC were 9.2% and 16.7%, respectively).

This study showed an excellent test-retest reliability (ICC = 0.958). High reliability is crucial to demonstrate the stability of questionnaire over time. However, most other FJS validation studies have shown an ICC score range of 0.9 to 0.92.[11] Unlike other studies, our test-retest was conducted over the telephone to relive patients' discomfort and burdens. This procedure might have affected the ICC. However, previous studies have reported that there is no statistical difference between in-person and telephone test-retest reliability. [22] [23]

Internal consistency was measured using Cronbach's alpha value. The K-FJS received an excellent value of 0.967, which was comparable to the original version's value of 0.95.[5] High score of internal consistency might mean homogeneity and reliability, indicating that items in the questionnaire were closely related to each other. However, too high Cronbach's alpha value (over 0.95) might reflect excessive duplication of contents. In such cases, researchers should be cautious when interpreting results.[24] In the present study, we eliminated each item and re-evaluated Cronbach's alpha. Calculated Cronbach's alpha values ranged from 0.962 to 0.969. Internal consistency was not increased after removing certain items and therefore, we could not find redundant items. According to a systemic review published by Adriani et al., the mean Cronbach's alpha value of 10 reviewed articles was 0.95 (range, 0.91-0.98). Of these 10 articles, there published in English-speaking culture had a greater mean value of 0.97 (range, 0.95-0.98).[11] Therefore, we can conclude that regardless of cross-cultural adaptation, high internal consistency is a characteristic of the FJS.

The responsiveness was evaluated at 3 months and 1 year after surgery, showing a moderate SRM (0.67). Moderate and high SRMs indicate sufficient internal responsiveness. Since our study design did not include preoperative K-FJS, we compared the change over time from 3 months to 1-year post-surgery. Many patients did not show up at postoperative 6-month follow-up. According to a study by Hamilton et al., the effect size (Cohen's d) that compared change from six months to 12 months post-surgery was 0.12. However, effect size that compared change from preoperative data to six months post-surgery was 2.6.[7] Our study results showed that the K-FJS was relatively sensitive in discriminating changes in clinical outcomes between 3 months and 1 year after surgery. We believe it is a suitable tool for monitoring clinical outcomes after surgery.

However, because clinical results might vary depending on how the follow up period is selected, it is meaningful to compare the absolute SRM value with VAS or the WOMAC test result. In this study, SRM of the K-FJS was higher than that of VAS (0.28), but lower than that of the WOMAC (1.03). This result is in accordance with a previous study by Bellamy et al.[3] showing that the WOMAC score tends to show a higher responsiveness than other evaluation tools.

This study has several limitations. First, since pre-operative FJS was not evaluated, we could not identify the floor effect for K-FJS. According to the study conducted by Hamilton et al., the floor effect of pre-operative FJS in patients with total knee arthroplasty was 15%.[7] Unlike Oxford knee score and WOMAC score which initially targeted osteoarthritis patients, FJS-12 was developed to assess post-operative population. Such difference would have resulted in higher floor effect for FJS.

Second, the mode of questionnaire administration was different for

the two surveys. All the surveys were performed by a single orthopedic physician assistant who fully understood this study. However, the first survey was performed based on interviewer-administered mode and the second by telephone survey. According to previous research, interview format of survey or telephone survey is more effective for the elderly or patients with low social economic status.[25] However, according to Lyons et al., interview format could systematically exaggerate health status compared with self-assessment.[26] Likewise, patients may respond differently to K-FJS measures depending on mode of questionnaire administration. Therefore, our study has a major limitation of not being able to confirm how the questionnaire results differ depending on the presence or absence of a researcher helping the questionnaire.

Chapter 5. Conclusion

The Korean version of Forgotten Joint Score (K-FJS) demonstrated strong measurement properties in terms of good construct validity and reliability. Our results suggest that it is an excellent instrument that can be used to monitor clinical outcomes after total knee arthroplasty. Using this standardized version of K-FJS, it would be possible for institutions to share clinical results more accurately.

References

1. Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, et al. Health-related quality of life after knee replacement. Results of the knee replacement patient outcomes research team study. Jbjs. 1998;80(2):163-73.
2. Kon E, Altadonna G, Filardo G, Di Matteo B, Marcacci M. Knee scoring systems. European Surgical Orthopaedics and Traumatology The EFORT Textbook: Springer Berlin Heidelberg; 2014. p. 3371-88.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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. The Journal of rheumatology. 1988;15(12):1833-40.
4. Marx RG, Jones EC, Atwan NC, Closkey RF, Salvati EA, Sculco TP. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. JBJS. 2005;87(9):1999-2005.
5. Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. J Arthroplasty. 2012;27(3):430-6.e1.
6. Baumann F, Ernstberger T, Loibl M, Zeman F, Nerlich M, Tibesku C. Validation of the German Forgotten Joint Score (G-FJS) according to the COSMIN checklist: does a reduction in joint awareness indicate clinical improvement after arthroplasty of the knee? Arch Orthop Trauma Surg. 2016;136(2):257-64.
7. Hamilton DF, Loth FL, Giesinger JM, Giesinger K, MacDonald DJ, Patton JT, et al. Validation of the English language Forgotten Joint Score-12 as an outcome measure for total hip and knee arthroplasty in a British population. Bone Joint J. 2017;99-b(2):218-24.
8. Matsumoto M, Baba T, Homma Y, Kobayashi H, Ochi H, Yuasa T, et al. Validation study of the Forgotten Joint Score-12 as a universal patient-reported outcome measure. Eur J Orthop Surg Traumatol. 2015;25(7):1141-5.
9. Shadid MB, Vinken NS, Marting LN, Wolterbeek N. The Dutch version of the Forgotten Joint Score: test-retesting reliability and validation. Acta Orthop Belg. 2016;82(1):112-8.
10. Thomsen MG, Latifi R, Kallemose T, Barfod KW, Husted H, Troelsen A. Good validity and reliability of the forgotten joint score

in evaluating the outcome of total knee arthroplasty. Acta Orthop. 2016;87(3):280-5.
11. Adriani M, Malahias MA, Gu A, Kahlenberg CA, Ast MP, Sculco PK. Determining the Validity, Reliability, and Utility of the Forgotten Joint Score: A Systematic Review. J Arthroplasty. 2020;35(4):1137-44.
12. Kim MS, Koh IJ, Choi YJ, Lee JY, In Y. Differences in Patient-Reported Outcomes Between Unicompartmental and Total Knee Arthroplasties: A Propensity Score-Matched Analysis. J Arthroplasty. 2017;32(5):1453-9.
13. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993;46(12):1417-32.
14. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104.
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. 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. 2010;19(4):539-49.
16. Evans JD. Straightforward statistics for the behavioral sciences: Thomson Brooks/Cole Publishing Co; 1996.
17. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
18. Kim JG, Ha JK, Lee JY, Seo SS, Choi CH, Lee MC. Translation and validation of the korean version of the international knee documentation committee subjective knee form. Knee Surg Relat Res. 2013;25(3):106-11.
19. Cohen J. Set correlation and multivariate methods. Statistical Power Analysis for the Behavioral Sciences 2nd ed Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 1988:467-530.
20. Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Horak FB, Boonsinsukh R. Responsiveness of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. Physical therapy. 2016;96(10):1638-47.
21. Cao S, Liu N, Han W, Zi Y, Peng F, Li L, et al. Simplified Chinese version of the Forgotten Joint Score (FJS) for patients who

underwent joint arthroplasty: cross-cultural adaptation and validation. J Orthop Surg Res. 2017;12(1):6.
22. García M, Rohlfs I, Vila J, Sala J, Pena A, Masiá R, et al. Comparison between telephone and self-administration of Short Form Health Survey Questionnaire (SF-36). Gac Sanit. 2005;19(6):433-9.
23. Braunsberger K, Wybenga H, Gates R. A comparison of reliability between telephone and web-based surveys. Journal of Business Research. 2007;60(7):758-64.
24. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53-5.
25. Sullivan LM, Dukes KA, Harris L, Dittus RS, Greenfield S, Kaplan SH. A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients. Medical care. 1995:AS183-AS94.
26. Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. Journal of Public Health. 1999;21(1):41-5.

초 록

배경: Forgotten Joint Score (FJS)는 인공관절 이후 환자의 주관적인 임상결과를 측정하는 patient reported outcome measure (PROM)으로 2012년 새롭게 개발된 설문도구이다. 이는 환자가 인공관절 치환술 이후 수술한 무릎을 자연스럽게 느껴서 마치 수술한 사실조차 잊고 사는 상태가 인공관절의 궁극적인 목표라는 점에 착안하여 만들어졌다. FJS는 WOMAC score에 비해 천장효과가 낮은 민감한 측정도구로 알려져 널리 사용되고 있다. 최근 세계 각국에서 여러 언어로 번역이 및 타당성에 대한 연구가 발표되었지만 국내에서는 아직 이에 대한 연구가 이루어진 바가 없다.

방법: 먼저 Cross-cultural adaptation을 위해 스위스의 개발자와 함께 guideline에 맞추어 5단계에 걸쳐 번역을 시행하였다. 이후 슬관절 인공관절 치환술을 받은 150명의 환자를 대상으로 한국어판 FJS와 함께 visual analogue scale (VAS), WOMAC, SF-36 설문을 작성하였다. Test-retest reliability를 평가하기 위해서 추가로 100명의 환자를 대상으로 전화 설문을 통하여 간격을 두고 한국어판 FJS를 2회 반복 작성하였다. Responsiveness는 50명의 환자를 대상으로 수술 후 3개월째와 1년째를 비교하여 후향적으로 평가하였다.

결과: 한국어판 FJS는 좋은 reliability(신뢰도)를 보여주었다. [Cronbach's α = 0.967, Intraclass correlation coefficient = 0.958, 95% CI: 0.930-0.974]. WOMAC score에 비해서 낮은 천장효과(Ceiling effect)를 보였으며(8.7%) 바닥효과(Floor effect)는 관찰되지 않았다. WOMAC, SF-36 score와의 상관계수는 각각 0.708, 0.682로 훌륭한 construct validity(구성타당도)를 보였다. 한국어판 FJS의 standardized response mean (SRM)값은 0.67로 이는 WOMAC score의 SRM(1.03)에 비해 낮았다.

결론: 한국어판 FJS는 좋은 reliability와 construct validity를 가지고

있으며 천장효과가 낮은 민감한 측정도구이다. 따라서 인공관절 수술 후 임상결과를 측정하기에 적합한 설문도구이다.