An Outcome Measure for Japanese People with Knee Osteoarthritis

MASAMI AKAI, TOKUHIDE DOI, KEIJI FUJINO, TSUTOMU IWAYA, HISASHI KUROSAWA, and TERUO NASU

ABSTRACT. Objective. We describe a new outcome measure for Japanese patients with knee osteoarthritis, the Japanese Knee Osteoarthritis Measure (JKOM).

Methods. The outcome measure was designed to incorporate the concepts of the World Health Organization’s International Classification of Functioning, Disability and Health 2001, and to reflect the specific Japanese cultural lifestyle, which differs from Western countries. To examine the validity and reliability of this measure and to determine the underlying set of measures, psychometric measurements including factor analysis were conducted in comparison with 2 other health related scales, the Western Ontario and McMaster Universities Arthritis Index (WOMAC) and the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36). The following psychometric properties were tested with this new measure: test and retest reliability, internal consistency, content validity, construct validity, and criterion related validity.

Results. The measure was proved to have sufficient reliability and validity by means of statistical evaluation and comparison with other health related scales. Three domains are clearly separated on the loading pattern by factor analysis: pain, limitation in mobility related to daily activity, and restriction of participation in social life and health perception.

Conclusion. The JKOM, a new knee OA measure, has sufficient reliability and validity for studies of the clinical outcomes of Japanese people with knee OA. (J Rheumatol 2005;32:1524–32)

Key Indexing Terms:
KNEE HEALTH RELATED QUALITY OF LIFE OSTEOARTHRITIS PATIENT BASED OUTCOMES

The use of outcome measures has been spreading throughout the world of clinical medicine. The keys to solving many problems in clinical practice depend upon valid and reliable methods to evaluate the outcomes of intervention. It is now recognized by orthopedic surgeons that one key outcome measure for any intervention for osteoarthritis (OA) is the change detected in patient based outcome and health related quality of life (HRQOL). This is because so many patients suffer unavoidable degenerative changes of OA and because of the continuing drastic increase of the aged population in developed countries like Japan.

When we talk about a universal standard of medical outcome assessment, however, it is a question of straining a de facto global standard to apply to individual local circumstances. It is necessary to account for local situations and to fine-tune the global standard to meet countries’ specific conditions. Differences of cross-cultural background are one of the basic problems to be discussed. We need an appropriate measure for patient based HRQOL that also reflects a Japanese social and cultural background. We have evaluated a new instrument, the Japanese Knee Osteoarthritis Measure (JKOM), comparing it with existing measures for the same patient group. The criteria usually used in direct comparisons of this kind are reliability and validity.

In many previous clinical studies dealing with this issue, there are approved de facto standards. The Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36), which is widely accepted for QOL assessment, is one example. The Western Ontario and McMaster Universities Arthritis Index (WOMAC) is also a commonly used disease-specific measures for OA of the hip and knee. We examined the validity and reliability of the JKOM by comparison with these 2 scales. We intended to construct a new measure incorporating the concepts of the World Health Organization’s International Classification of Functioning, Disability and Health (ICF 2001) and reflecting the specificity of the Japanese cultural lifestyle.

MATERIALS AND METHODS

Scales. Patients with OA answered 3 self-completion questionnaires: the WOMAC, the SF-36, and the JKOM.
The WOMAC is a self-administered, disease-specific health measure, developed by Bellamy, et al; it is a 24-item questionnaire grouped into 3 categories, pain, stiffness, and physical function, and it was originally designed for use in clinical trials in patients with hip and knee OA6. (The Japanese translation was obtained with permission of Dr. Bellamy). The SF-36 is a self-administered generic questionnaire consisting of 36 questions divided into 8 domains: physical function, social function, role-physical, role-emotional, bodily pain, general health, mental health, and vitality3. The SF-36 was translated into Japanese and tested by Fukuhara, et al for validity and reliability7. A stratified comparison with the general population in Japan was also done according to sex and age decades. Several reports describe validity of the SF-36 and the WOMAC applied to patients with OA and other conditions8-10.

The JKOM is a self-administered, disease-specific measure, consisting of 23 items in the first draft used in the first inquiry described below, and 25 items in the revised version (used in the second inquiry), which include patient pain in level walking, standing or climbing stairs, physical functions related to the activities of daily living, and social functions including participation. Based on the ICF concept, we assumed that the disabilities associated with knee OA were as follows: health condition: osteoarthritis; impairment: knee pain; activity limitation: difficulty in mobility related to daily activities; and participation restriction: difficulty in participating in social life.

Twenty-three questions in the first draft were constructed to identify disability and impairment. A group of 6 persons who had wide experience in treating knee OA and 2 experts in related fields were involved in the identification, classification, and priority rating of the factors related to therapeutic exercise for knee OA (The Ad Hoc Committee on a New Outcome Measure for Knee OA). The design of the questionnaire partly referred to the Japanese Orthopaedic Association Knee Scoring System11. To check content validity, the expert group assessed the pertinence of each question corresponding to the questionnaire’s aims. The content of each question was determined through repeated discussions among group members. The intention was that items selected for the questionnaire reflect the contemporary lifestyles of Japanese people.

The resulting document was originally written in Japanese and then translated into English. The second version was revised according to the results from the first inquiry in this study. Recruitment of patients. Patients were recruited from outpatient clinic groups attended by the members of the Japanese Orthopaedic Association. Patient selection was conducted consecutively during the planned inquiry period. According to the report from the Subcommittee on Classification Criteria of Osteoarthritis of the American Rheumatism Association, diagnostic entry criteria in our study for “knee osteoarthritis” were defined as (1) knee pain, (2) age over 50 years, and (3) osteophytes confirmed by radiographs12. As well, idiopathic knee OA was defined as (1) knee pain; (2) at least one of the following 3: age over 50 years, stiffness for less than 30 minutes, crepitus; and (3) osteophytes, when using both clinical and radiographic findings with 91% sensitivity and 86% specificity values12. The SF-36 was translated into Japanese and tested by Fukuhara, et al for validity and reliability7. A stratified comparison with the general population in Japan was also done according to sex and age decades. Several reports describe validity of the SF-36 and the WOMAC applied to patients with OA and other conditions8-10.

The JKOM is a self-administered, disease-specific measure, consisting of 23 items in the first draft used in the first inquiry described below, and 25 items in the revised version (used in the second inquiry), which include patient pain in level walking, standing or climbing stairs, physical functions related to the activities of daily living, and social functions including participation. Based on the ICF concept, we assumed that the disabilities associated with knee OA were as follows: health condition: osteoarthritis; impairment: knee pain; activity limitation: difficulty in mobility related to daily activities; and participation restriction: difficulty in participating in social life.

Twenty-three questions in the first draft were constructed to identify disability and impairment. A group of 6 persons who had wide experience in treating knee OA and 2 experts in related fields were involved in the identification, classification, and priority rating of the factors related to therapeutic exercise for knee OA (The Ad Hoc Committee on a New Outcome Measure for Knee OA). The design of the questionnaire partly referred to the Japanese Orthopaedic Association Knee Scoring System11. To check content validity, the expert group assessed the pertinence of each question corresponding to the questionnaire’s aims. The content of each question was determined through repeated discussions among group members. The intention was that items selected for the questionnaire reflect the contemporary lifestyles of Japanese people.

The resulting document was originally written in Japanese and then translated into English. The second version was revised according to the results from the first inquiry in this study. Recruitment of patients. Patients were recruited from outpatient clinic groups attended by the members of the Japanese Orthopaedic Association. Patient selection was conducted consecutively during the planned inquiry period. According to the report from the Subcommittee on Classification Criteria of Osteoarthritis of the American Rheumatism Association, diagnostic entry criteria in our study for “knee osteoarthritis” were defined as (1) knee pain, (2) age over 50 years, and (3) osteophytes confirmed by radiographs12. As well, idiopathic knee OA was defined as (1) knee pain; (2) at least one of the following 3: age over 50 years, stiffness for less than 30 minutes, crepitus; and (3) osteophytes, when using both clinical and radiographic findings with 91% sensitivity and 86% specificity values12. Patients were invited to participate in this study and were given a questionnaire booklet and asked to complete a consent form. All patients were then asked to fill in the WOMAC, the SF-36, and the JKOM questionnaires, which were given in the same order for each patient, at the first inquiry in July 2002. Their attending physicians filled out their patient information sheets to check the diagnostic entry criteria and retained their consent forms. Radiographs of the anteroposterior view of both knee joints in a standing position were taken and used to confirm the diagnosis. A review team reviewed the radiographs after radiological findings were explained to the patients. When all 3 questionnaires were completed, the booklets were collected for analysis. At the second inquiry, in January 2003, only the revised JKOM questionnaire and a pain assessment procedure using a visual analog scale (VAS) were used (Table 1).

Statistical analysis. First, the completion rate for filling out the 3 questionnaires and the distributional properties of the items were checked. Frequency distribution was then calculated for each individual question item and so-called “floor and ceiling” effects, i.e., skewed distributions, were checked among the 3 questionnaires. If many patients showed a score at either extreme of a scale, there was a limitation to detect improvement or deterioration.

The correlations between each item in the 3 questionnaires in the 2 inquiries, which were ordinal scale data, were evaluated using Spearman’s correlation coefficient, Kendall’s tau b, and the Akaia Information Criterion (AIC)13,14. The AIC is a popular method to assess the relationship between items13 and was devised using Syntax language for SPSS software (v. 10.0J; SPSS Inc., Chicago, IL, USA). We checked all items from the 3 questionnaires (WOMAC, SF-36, and JKOM) for test-retest reliability. Internal consistency was tested for the 3 questionnaires and within each one. Cronbach’s alpha was calculated for this internal consistency15. Each patient filled out all 3 questionnaires at the same time, and interobserver (intrarater) reliability was not tested.

Second, for criterion related validity, concurrent validity was examined in terms of a comparison between the similar domains of the SF-36 and WOMAC, which are de facto standards of generic and disease-specific scores regarded as external criteria. Correlations between domains and total score in the SF-36, WOMAC, and the original version of the JKOM were also measured with the AIC to compare the amount of interrelation between items.

Third, as a construct validity factor analysis, a scree test and categorical principal component analysis were carried out for the 83 items from the 3 questionnaires. The component loading plots and categorical combined plots provided a conversion from ordinal scale to interval scale, and made it possible to calculate a weighted score.

Statistical analysis was performed using SPSS.

RESULTS

Patient characteristics and face validity. Distribution of the attending physicians, up to 12 in the first inquiry and 9 in the second inquiry, was nationwide.

First inquiry. A total of 150 patients answered the 3 questionnaires. The mean age of patients was 72.7 ± 9.3 years (range 50–99) and the majority of the sample were women (female 84.9%, male 15.1%). Among this group 35 patients were assessed again within 2 weeks using the same 3 questionnaires.

Second inquiry. Another 84 patients answered the JKOM revised questionnaire and were assessed on a pain VAS. There were no questions unanswered. The mean age of patients in the second inquiry was 73.4 ± 7.7 years (range 51–88) and more than three-quarters were women (female 79.8%, male 20.2%). The frequency distributions, i.e., completion rate of each item in the first inquiry, were 146–149 out of 149 in the JKOM original version, 143–147 out of 149 in WOMAC, and 140–148 out of 149 in SF-36.

The component loading plots and categorical combined plots provided calculated intervals for a weighted score in each item. Some items from the JKOM original version showed uneven distributions on the scatter plot, and were slightly different from those of the WOMAC.

Reliability. The results for test-retest reliability from the first inquiry were indicated by Spearman’s correlation coefficient, Kendall’s tau b, and AIC. Generally speaking, the
relationship between Spearman and Kendall seemed to be relatively high, but the AIC did not provide similar findings. The 5 items revealing the highest correlations were W13: walking on flat (r = 0.858); k15: using walking aids (r = 0.844); k23: travel (r = 0.807); k19: heavy domestic duties (r = 0.791); and k02: pain going up stairs (r = 0.781) from the Spearman correlation coefficient; and W13: walking on flat (r = 0.807); k15: using walking aids (r = 0.797); k23: travel (r = 0.791); k19: heavy domestic duties (r = 0.738); and s03-3: lifting or carrying groceries (r = 0.731)
from Kendall’s tau b. After sorting of Spearman coefficients, Kendall’s tau b matched with the AIC; however, the degree of relative association among the 3 scales showed inconsistency. The 5 highest correlations between test and retest from the AIC were s03-3: lifting or carrying groceries (AIC = −15.89); s03-6: bending, kneeling, or stooping (AIC = −15.74); k14: walking on flat (AIC = −14.04); w13: walking on flat (AIC = −14.96); and s03-8: walking several blocks (AIC = −11.36).

The results shown by the relationships derived from all items using the AIC had high values between the identical items and similar items, i.e., good correlation for test-retest reliability and similarity between the scales.

By contrast, the lowest 5 correlations were s07: amount of bodily pain (AIC = 41.03); s09-8: happy person (AIC = 40.51); s09-5: a lot of energy (AIC = 39.75); s09-7: worn out (AIC = 38.73); and s09-9: tired (AIC = 35.51).

These results showed that bodily pain and physical functioning varied in a relatively narrow range, and general health perception and vitality fluctuated widely.

As for internal consistency, Cronbach alpha coefficients were 0.911 for the JKOM original version, 0.900 for WOMAC, and 0.947 for SF-36. All 3 scored more than 0.9 and the total score for all 3 questionnaires was 0.963.

**Validity.** The analysis for validity focused mainly on the conceptual developing process to build up a new construct incorporating the ICF 2001 concept. We wanted to identify the major dimension of factors in the JKOM data represented by groupings of items, as follows.

**Major dimension of clusters.** When trying to determine the underlying cluster of a set of measures, factor analysis is a powerful method for analyzing data. Factor analysis reveals the patterns of shared variation or interrelationship within a score matrix. The number of dimensions was first assessed to evaluate the dimensional structure of the JKOM original version and the other 2 scales.

The results of the 3 questionnaires were basically qualitative data measured by an ordinal scale with 3 to 5 grades. Factor analysis usually requires data from an interval scale. However, correlation coefficients have proved to be reasonably robust with respect to ordinal scale distortions. Thus, factor analysis represents in a practical way the relationships among the data if the scale categories are not extremely skewed. Therefore, all 83 items from the 3 questionnaires were analyzed using factor analysis, supposing that the data were quantitative.

The result of categorical principal component analysis for the 83 items of the 3 questionnaires is shown in Figure 1. The first group consisted of pain and physical function, activities of daily living (ADL) and instrumental ADL from the WOMAC and JKOM original version, and bodily pain and physical functioning from the SF-36. The second group had activity and participation from the JKOM original version, and role-physical, role-emotional, social functioning, general health perception, vitality, and mental health from the SF-36.

**Dimensional structure.** The scoring distribution of each factor was indicated as the correlation matrix in order of the larger figure in each dimension. Among the 83 items, the first factor had the largest value in 60 items, the second in 10, the third in 4, the fourth in 3, the fifth in 2, and the remainder in 1 item. The items with lower values were those not dependent upon knee joint OA, such as general health perception and role-emotional.

**Varimax rotation.** We used the scree test to identify the point where the decreasing percentage of variance explained by the factors levels off17. After calculating for the accounting ratio of each variance, appropriate dimension numbers were found to be 2 or 3.

From the 3 questionnaires the first 3 dimensions of the JKOM original version are shown in a 3-dimensional graph (Figure 2A). To interpret most factor analysis applications, the factors should be adjusted, i.e., rotated, mathematically to more clearly define their clusters18. After varimax rotation, the 3-dimensional graph and the specific subgroups were converted into a 2-dimensional graph to show the interrelations among the groups (Figure 2B). The first group consisted of various types of pain, the second group consisted of physical function and activities of daily living, and the third group consisted of participation items concerning community events and travel.

**Categorical principal component analysis for JKOM.** After factor analysis was performed on the assumption that the questionnaires provided quantitative data, categorical principal component analysis was also performed to confirm the dimensional structure. The results for the JKOM original version are shown in Figure 2C. The construct structure consisted of 3 clusters of items.

**Reselection of questions.** In consideration of the AIC results to show weighted scoring for each item and that of WOMAC or SF-36, some of the questions were modified, 4 questions were added, and 2 were deleted, as follows: Items related to health condition, general health perception, morning pain, and morning stiffness were added. Items related to ascending and descending stairs were combined into Going up/down stairs. Rising from the floor and Rising from a chair were changed to Getting in/out of the toilet. Pain in standing up was changed to Pain in standing up/sitting down. Leisure activity was changed to Going outside nearby. Travel was changed to Going outside far. Carrying by hand was changed to Difficulties of activities.

Patients responded to this revised version of the JKOM in the second inquiry, with patient pain scoring undertaken using a VAS.

**Relationship with pain.** From results of the second inquiry, the relationship between pain and the JKOM revised version was investigated by Spearman’s correlation coefficient by
ranks. The result is shown in Figure 3 and the correlation coefficient is 0.579 (statistically significant with a level of p < 0.01).

Retest with factor analysis. After repeating 6 times, mathematical rotation was converged, and the varimax-rotated item patterns were extracted to show the final affirmation of the structure of the JKOM construct. Figure 4 and Table 2 display the results of the factor-loading pattern. Three domains are clearly separated on the loading pattern: (1) pain, (2) limitation in mobility related to daily activity, and (3) restriction of participation in social life and health perception.

This result indicated that there was an underlying set of item clusters that reflect the multidimensional and interrelated structure of the revised JKOM measure that was constructed based on the ICF concept and the contemporary Japanese lifestyle and sense of values.

DISCUSSION

Patient based outcome measures. Measuring health status or disease condition has been one of the most important efforts in medical practice. In the past, epidemiological factors, such as disease incidence and infant mortality, were widely used as indicators for health status as well as socioeconomic environment. However, during the past decades a consensus has been developed regarding the patient’s point of view in measuring medical intervention. Patient based outcome measures have recently been established as standardized tools to monitor the functioning and well being of patients.20 Further, the accumulated results from several outcome studies are concentrated into clinical guidelines and other references.21

Consideration of the cultural background of each patient must precede any attempt to apply sociomedical indicators for the assessment of health conditions. The terms “illness/disease” and “health” are compound cross-cultural issues in the system of values, cultures, and historical situations.22

Conceptual development of a construct. The majority of our functions or abilities are not directly measurable like physical or chemical materials. We require a concept that indirectly reflects the status of our functions or abilities. This concept is called a “construct” in a clinical situation. For
example, when measuring gait ability, the measured items range from kinetic parameters to kinematic figures, which never actually indicate gait ability directly.

Recent development of computer systems and rapidly progressing information technology have brought global communication networks and online electronic data sources worldwide. This led to a universal concept of outcome measures, and made it possible to use at least translated scales. On the other hand, we need to know the real situation of our own country, and adjust the global standard to meet the specific conditions of our patients, to maintain our daily clinical activity. We require a new, appropriate measure of knee OA for Japanese that is slightly different from the WOMAC or other widely used scales. In this study we were also trying to develop a new construct based on the ICF 2001 concept that emphasizes participation more than before.

Requirements for a new measure and check of construct validity. New outcome measures should be evaluated relative to existing measures, with direct comparisons performed in the same patient population. The checkpoints used in direct comparisons include reliability and validity.

Figure 2. A. Three-dimensional graph shows the first 3 dimensions and the subgroups of the specific items of the JKOM original version. B. Factor plotting of the JKOM original version after varimax rotation. C. Result of categorical principal component analysis for the JKOM original version.
Several steps must be statistically tested using multivariate analysis: (1) Pretest for measuring the scale. (2) Check for construct, criterion related, and other validities and reliabilities. (3) Definition of measuring concept. (4) Selection of items used in the questionnaire. (5) Completion of the measuring scale.

Our results revealed that the JKOM performed as well as or better than the other 2 measures according to these checkpoints. The JKOM has items from 3 domains, which correspond to (1) pain/stiffness, (2) activities of daily living, and (3) movement/role/participation/health perception. It is necessary to confirm these characteristics in a larger trial, and to reinforce our method with a detailed guideline for assessment; standardized equipment, if necessary; and a training course for researchers.

Validation of psychometric properties. Convergent validity, which is expected to show the strength of interrelations within the same category of a scale, was checked using categorical principal component analysis. Criterion related validity is indicated by comparing the results obtained from a measurement scale with a “gold standard” or validated indicator of the situation. In this study, concurrent validity was investigated because the WOMAC and SF-36 have been shown to have predictive, discriminant validity to evaluate criterion related validity. However, too high a concurrent validity is not useful for a new measure.

Using a generic scoring system or disease-specific scoring system is another important issue to be investigated. Multidimensional assessments of health status are important. To determine the result of measures and to interpret the respective scores, priority should be given to clarifying what is most relevant to the study object and to interpreting the results.

We are planning a randomized controlled trial using this...
outcome measure for patients with knee OA. Accumulating experience with this new measure will provide more information for future research.

ACKNOWLEDGMENT

We thank all orthopedic surgeons, staff, and patients who participated in inquiries for this study: Abe Orthopaedic Clinic, Ousyuku Spa Hospital, Kuboyama Orthopaedics, Kondo Orthopaedics, Sasaki Orthopaedics (inquiry 1 only), Yamashita Orthopaedics, Moriyama Municipal Hospital (inquiry 1 only), Seirei Hamamatsu Hospital, Kusaka Hospital (inquiry 1 only), Nakamura Orthopaedic Clinic, Fujino Orthopaedic Clinic, Nasu Orthopaedic Clinic. We thank Prof. Kunihiko Hayashi and Prof. Eiji Marui for helping to develop this new measure; and Kayoko Nakanishi for preparation of the manuscript.

REFERENCES