

# Intra- and inter-rater reliability of the Knee Society Knee Score when used by two physiotherapists in patients post total knee arthroplasty

**ABSTRACT: Background and Purpose:** It has yet to be shown whether routine physiotherapy plays a role in the rehabilitation of patients post total knee arthroplasty (Rajan et al 2004). Physiotherapists should be using valid outcome measures to provide evidence of the benefit of their intervention.

The aim of this study was to establish the intra and inter-rater reliability of the Knee Society Knee Score, a scoring system developed by Insall et al (1989). The Knee Society Knee Score can be used to assess the integrity of the knee joint of patients undergoing total knee arthroplasty. Since the score involves clinical testing, the intra-rater reliability of the clinician should be established prior to using the scores as data in clinical research. Where multiple clinicians are involved, inter-rater reliability should also be established.

**Design:** This was a correlation study.

**Subjects:** A sample of thirty patients post total knee arthroplasty attending the arthroplasty clinic at Johannesburg Hospital between six weeks and twelve months postoperatively.

**Method:** Recruited patients were evaluated twice with a time interval of one hour between each assessment.

**Statistical Analysis:** The intra- and inter-rater reliability were estimated using Intraclass Correlation Coefficient (ICC).

**Results:** The intra-rater reliability showed excellent reliability ( $h=0.95$ ) for Examiner A and good reliability ( $h=0.71$ ) for Examiner B. The inter-rater reliability showed moderate reliability ( $h=0.67$  during test one and  $h=0.66$  during test two).

**Conclusion:** The KSKS has good intra-rater reliability when tested within a period of one hour. The KSKS demonstrated moderate agreement for inter rater reliability.

**KEYWORDS:** TOTAL KNEE ARTHROPLASTY, KNEE SOCIETY KNEE SCORE, REHABILITATION OUTCOME MEASURES.

## INTRODUCTION

Total knee arthroplasty is used in the treatment of osteoarthritis of the knee to bring about a decrease in pain (Hawker et al 1998; McAuley et al 2002) and improvement in function (Hawker et al 1998; Walsh et al 2001). Impairments acquired post total knee arthroplasty may include, knee flexion contracture, limited range of motion, quadriceps weakness, instability and malalignment (Bhave et al 2005). Physiotherapy aims to prevent these impairments through appropriate treatment techniques which include continuous passive mobilisation, stretching, strengthening, active and passive mobilisations, functional electrical stimulation, gait training and patient education.

It has yet to be shown whether routine physiotherapy plays a role in the rehabilitation of patients post total knee arthroplasty (Rajan et al 2004). If patients are not routinely referred for physiotherapy, it becomes essential to continuously assess patients postoperatively to monitor for the development of such impairments. If patients are being routinely referred for outpatient physiotherapy, as is common practice in many facilities, then physiotherapists should be using valid outcome measures to provide evidence of the benefit of their intervention.

Whether patients are being referred for outpatient physiotherapy or not, the outcome measures used should be valid, reliable, responsive and standardized to

facilitate the communication of results in the medical (between healthcare professionals) and scientific community (Kreibich et al 1996). An outcome measure must provide the user with an objective measure (Davies 2002) of the subject's impairment which can be compared with other similar subjects and should be applicable before and after

## Correspondence to:

Wendy-Ann Wood  
Department of Physiotherapy,  
Medical School  
University of the Witwatersrand  
7 York Road Parktown 2193  
Johannesburg, South Africa  
Email: bradwend@hotmail.com

Gopal S, MSc<sup>1</sup>;  
Wood W, MSc<sup>1</sup>;  
Myezwa H, PhD<sup>1</sup>;  
Stewart A, PhD<sup>1</sup>

<sup>1</sup> Division of Physiotherapy,  
University of the Witwatersrand.

an intervention. An outcome measure should also be related to the intervention (APA position statement 2003).

The American Knee Society Clinical Rating System (AKSCRS) is one among the most commonly used outcomes for total knee arthroplasty patients (Stavem and Arnesen 2005; Lingard et al 2001). It is a dual rating system developed by Insall et al (1989). It is also known as the Knee Society Rating System (KSRS) or Knee Society Clinical Rating System (KSCRS). For the purpose of clarity the rating system hereafter will be referred to as the AKSCRS. The AKSCRS has two components, the knee score and the functional score. The system was designed to score the knee joint itself (knee score) and its function (functional score) separately, thus avoiding the impact of functional and age related health problems on the knee joint itself. The knee score is based on the subjective assessment of pain and objective measurement of stability, range of motion, flexion contracture, extension lag and alignment at the knee joint. The individual scores are combined to give the knee a score which ranges from 0 to 100. The functional score is a composite score of walking, climbing up and down stairs and use of assistive devices. The knee score has been shown to be valid and responsive (Lingard et al in 2001). The functional score of the system has been shown to be less responsive (Lingard et al in 2001) and is not explored further in this paper. For the purpose of clarity, the knee score will be mentioned hereafter as Knee Society Knee Score (KSKS).

Wright and Feinstein (1992) discussed the common causes of variability in orthopaedic measurements. They stated that patient, procedure and clinician variability are the common causes for unreliable measures. Patient variability can be reduced by selecting measurement tools appropriate to patient conditions. Procedural variability can be reduced by using the same instruments and standardising measurement procedures. Clinician variability can be reduced by repeated practice and experience of the examiner in the measurement skills used.

The classification of some commonly used outcome measures, based on their type, validity, reliability and dual rating design (design which measures structure and function of the joint separately) are shown in Table 1.

From Table 1 it can be seen that if researchers are in need of a joint specific outcome measure, that has been shown to be valid and reliable, the Knee Score component of the AKSCRS is a good option.

The aim of this study was to assess whether the KSKS can be reliably used by physiotherapists in evaluating the knee joint in post TKA patients. This was achieved by establishing intra- and inter-tester reliability of two qualified physiotherapists using the KSKS.

#### MATERIALS AND METHODS

This was a correlational study. Ethical clearance for this study was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand. Patients who agreed to take part in the study signed a consent form and were assigned numerical codes on the data sheet, ensuring anonymity.

#### Sample

Two qualified physiotherapists partici-

pated and are referred to as examiner A and examiner B. The study was conducted at the arthroplasty clinic in a Gauteng hospital. The Knee Society Knee Score was administered on patients who met the inclusion criteria and gave consent for participation.

#### Inclusion criteria:

- Patients aged between 45 – 75 years who were attending the clinic for their six weeks to one year postoperative follow up visit

#### Exclusion criteria:

- Subjects who were not walking (independently or without walking aids) prior to surgery / severe pain
- Pre-existing septic arthritis or conditions that may compromise TKA outcome (Charcot's joint, Paget's disease, severe osteoporosis)
- Patients with a neurological disorder that may affect the outcome of TKA
- Patients with infectious diseases or metastatic disease

The sample size required was taken as 30 based on the assumptions of normality that a minimum sample size of n = 30, is essential to be able to assess agreement.

#### Procedure

The principal researcher (Examiner A) and Examiner B evaluated the patients independently of one another. The

**Table 1: Classification of outcome measures based on their type, validity, reliability and dual rating design.**

Outcome Measures	Type of measure	Validity	Reliability	Dual rating system
WOMAC (Kreibich et al 1996; Davies 2002)	Disease specific	Yes	Yes	No
SF-36 (Aitken and Bohannon 2001; Lingard et al 2001)	General health	Yes	Yes	No
HSSKRS (Davies 2002)	Joint specific	No	Yes	No
Bristol Knee Score (Davies 2002)	Joint specific	No	Yes	No
OKS (Davies 2002; Dawson et al 1998)	Disease specific	No	Yes	No
AKSCRS (Insall et al 1989)	Joint specific	Yes	Yes	Yes

patients were taken into a room with a plinth and a chair with back support along with the researcher (examiner A) and an observer. The observer was a qualified physiotherapist. The same room, chair, plinth and goniometer were used for all measurements and for the full duration of the study.

Examiner A took all measurements, immediately followed by examiner B. The procedures were repeated by examiner A and examiner B with a stipulated time interval not less than 45 minutes between their first and second measurements. The examiners did not record the measure directly but, gave the actual measures (eg. degrees of ROM) to the independent observer who entered it into the data sheet, minimising examiner bias. The principal researcher completed the scoring after data collection was complete.

#### PROCEDURES FOLLOWED FOR MEASURING EACH COMPONENT OF THE SCORE

##### Pain

The patients were asked, “Do you have any pain in your operated knee?” If they answered “yes” they were asked, “Is your pain mild, moderate or severe”. If they had mild pain, they were asked whether they had pain while using stairs and walking. In cases where they had moderate pain, they were asked whether the pain was continuous or occasional.

##### Range of motion (ROM)

A universal goniometer was used to measure the range of motion at the knee joint. The patients were in supine lying. The head was supported by a pillow, with the hip in neutral and the knee extended (Clarkson and Gilewich 1989). The goniometer axis was placed over the lateral condyle of the femur with its stationary arm parallel to the longitudinal axis of the femur pointing towards the greater trochanter and the movable arm parallel to the longitudinal axis of the fibula pointing to the lateral malleolus. The measurement was noted down as initial ROM. If the initial ROM was not 0° the reading was taken as degree of flexion contracture. The patients were instructed to take their

heel towards their buttock and the examiner assisted the movement to feel the end range and measured range of motion. The patients were instructed to inform the examiner if they felt any pain or discomfort in their knee and the movement was stopped at that point. The measurement was recorded in the data sheet by the observer.

##### Stability

The Lachman’s test (Petty and Moore 1998) and Valgus-Varus stress test (Magee 1997) was used to assess the anteroposterior and mediolateral stability respectively. The amount of translation of the tibia over the femur during Lachman’s, and the amount of angulation at the knee joint during the Valgus-Varus test experienced by the examiner were conveyed to the observer and noted on the data sheet. These were clinical measurements of what the examiner experienced during the tests.

##### Extension lag

The patients were positioned supine at the end of the plinth, with the knee hanging flexed over the end of the plinth, with a towel roll underneath the distal thigh. The patients were asked to actively extend the knee and range was measured using the goniometer as active extension ROM. The difference between active extension ROM and the passive extension ROM was recorded as the degree of extension lag (Stillman 2004) by the observer.

##### Alignment

Measurements of the degree of valgus and varus at the knee joint were obtained from the surgeon.

#### STATISTICAL ANALYSIS

Reliability was assessed by making use of an Intraclass Correlation Coefficient (ICC) (John 2004). ICC (h) is the number obtained from the statistical analysis which ranges from zero to positive or negative one. The closer the value of ICC is to one, the closer the relationship between the two variables (Hicks 1995).

#### RESULTS

Thirty patients were initially included in this study. Two patients were excluded due to severe pain and one patient was excluded as she was unwilling to participate once the testing began. In three patients, both knees met the inclusion criteria, therefore 30 knees were examined. The alignment scores for two knees were missing from the database and therefore the following results are from the scores of 28 knees. An average of one hour was the time between the first and second measurements and in no case was the time less than 45 minutes.

##### Intra-rater reliability

The total scores obtained by individual examiners during their assessments with KSKS were used to establish the intra-rater reliability of the KSKS. The first set of scores obtained by examiner A were compared and correlated with the second set of scores obtained by examiner A. The same procedure was followed for examiner B. Individual items on the KSKS were also subjected to analysis. The ICC (h) for intra-rater reliability for the individual items as well as the total score is shown in Table 2.

**Table 2: The Intraclass Correlation Coefficient of intra rater reliability for the individual items from Examiner A and Examiner B.**

Item on the KSKS	Intraclass Correlation Coefficient (h)	
	Examiner A	Examiner B
Knee ROM	0.96	0.94
AP stability	0.82	0.80
ML stability	0.72	0.82
Flexion contracture	0.95	0.89
Extension lag	0.65	0.87
Total	0.95	0.71

\* AP stability – anterior posterior stability, ML stability – medial lateral stability

**Table 3: The Intraclass Correlation Coefficient of inter rater reliability for the individual scores during test 1 and test 2, between Examiners A and B.**

Item on the KSKS	Intraclass Correlation Coefficient (h)	
	Test 1 (A & B)	Test 2 (A & B)
Knee ROM	0.85	0.82
AP stability	0.47	0.45
ML stability	0.32	0.24
Flexion contracture	0.58	0.64
Extension lag	0.54	0.76
Total	0.67	0.66

\*ROM –range of motion, AP stability – anterior posterior stability, ML stability – medial lateral stability

Examiner A showed excellent correlation and examiner B showed good correlation for the KSKS (.90 ≤ excellent, .70 to .89 = good, .50 to .69 = moderate, .50 ≥ poor).

**Inter-rater reliability**

The total scores obtained by examiners A and B during their test 1 and test 2 were used to estimate the inter-rater reliability of the KSKS. The set of scores obtained from test 1 by examiner A and examiner B were correlated. Similarly the set of scores obtained from test 2 by examiner A and examiner B were also correlated. Individual items on the KSKS were also subjected to analysis. The ICC (h) measuring inter-rater reliability of the individual items as well as the total score is shown in Table 3.

Most of the individual items in the scoring system showed a poor correlation between the two examiners. Overall, the examiners showed moderate correlation between the KSKS during test 1 and test 2.

**DISCUSSION**

Practice of assessment and evaluation in physiotherapy has been emphasised not only for the purpose of quality service, but also for audit and research advances (Stavem and Arnesen 2005; Kreibich et al 1996). It has become essential for the physiotherapist to assess the effectiveness of a treatment using an outcome measure which is valid and reliable. Besides improving the quality of health care services, reliable outcome measures enhance the quality of trials in which they are used (John 2004). It is important for physiotherapists to communi-

cate their findings in the same terms as other health professionals to facilitate the team-approach to patient management.

The aim of this study was to establish the intra and inter-rater reliability of two physiotherapists using the KSKS. The common method of test-retest reliability was implemented by administering the KSKS at two different times with an average time interval of 60 minutes between them. The time interval in this study was not so short that the memory of the previous test biased the performance of the examiner (Thomas and Stewart 2005) and not too long so that there were no changes in the attributes which were being measured (Finch 2002; Campbell et al 1999).

The argument that good overall reliability was biased by the memory of previous measurement as the time interval (one hour) between the two tests was shorter than that of the time interval (two hours) of a previous study (Liow et al 2000) could be made. In the current study, the influence of memory was minimised by recruiting an independent observer to note down the measurements from the tests with the intention that if the examiner was not actually writing down the measurement it would less easily be committed to memory.

In our study, examiner A showed excellent intra-rater reliability (h=0.95) and examiner B showed good intra-rater reliability (h = 0.71). In a similar study (Liow et al 2000) the knee score was administered by six examiners with varying experience on 29 subjects. The study showed considerable variations in intra-rater reliability which was attri-

buted to poor experience of the examiners and a lack of training in administering the tool. They also found that the examiners with more than three years experience showed relatively higher intra-rater reliability. In our study, both the physiotherapists had more than four years experience and were trained in the assessment tool prior to the study, which may have contributed to the reliability. The results of this study revealed a moderate inter-rater reliability between the examiners during test 1 (h=0.67) and test 2 (h=0.66). Ryd et al (1997) reported low inter-rater reliability with a standard deviation of 26 for the knee score which is larger than that that reported in this study (SD = 16). The physiotherapists in our study standardised the measurement procedures through repeated practice, training and discussion prior to the study. This is supported by Liow et al (2000). To complete this discussion, individual components of the score will be discussed.

Intra-rater reliability for ROM was excellent for both examiners (h=0.96 and 0.94). Inter-rater reliability for ROM was also good (h=0.85 and 0.82). Analysis of the flexion contracture component of the KSKS, showed excellent (h=0.95), and good (h=0.89) intra-rater reliability. This is significantly higher than in the previous study by Liow et al 2000 (Kappa=0.52). Moderate inter-rater reliability was found between the examiners during test 1 and test 2 (h=0.58 and 0.64). This is relatively higher than that reported between experienced staff by Liow et al 2000 for the same component (Kappa=0.19). It is of interest that knee ROM and flexion contracture showed very little variation. This may be because both were controlled passively by the examiners and the measurements are taken using goniometry, which has been found to be a reliable measure (Smith and Walker 1983; Gajdosik and Bohannon 1987).

When analysing the stability components of the KSKS, both examiners showed good intra-rater reliability in antero-posterior stability and medio-lateral stability. A previous study (Liow et al 2000) demonstrated moderate correlation in an experienced examiner with a Kappa value = 0.50. In our study,

poor inter-rater reliability was found between the examiners. It is postulated that this is due to the subjective nature of the testing procedure. In contrast, the measurements from a goniometer are more objective and showed highest inter-rater reliability among the items in the KSKS.

An interesting finding was that the inter-rater reliability of extension lag improved from test 1 ( $h=0.54$ ) to test 2 ( $h=0.76$ ). This may be attributed to a learning effect, or due to repeated reinforcement from the examiners.

In conclusion the results of this study showed good intra-rater reliability and moderate inter-rater reliability for the KSKS when conducted by two experienced physiotherapists. Physiotherapists working in the field of osteoarthritis or total knee replacement rehabilitation should consider using this measure in the clinical setting as well as in research.

## REFERENCES

- Aitken DM, Bohannon RW. 2001. Functional independence measure versus short form-36: relative responsiveness and validity. *International Journal of Rehabilitation Research*. 24(1): 65-68.
- APA (Australian Physiotherapy Association) position statement. 2003. Clinical justification and outcome measures. November. 1-3
- Bhave A, Mont M, Tennis S, Michele N, Starr R, Etienne G. 2005. Functional problems and treatment solutions after total hip and knee joint arthroplasty. *The Journal of Bone and Joint Surgery*. 87: 9-21.
- Campbell MJ, Machin D. 1999. *Medical statistics a common sense approach*. Third edition. John Wiley and Sons Ltd. England: 28-29.
- Clarkson HM, Gilewich GB. 1989. *Musculoskeletal Assessment-Joint Range of Motion and Manual Muscle Strength*. First edition. Williams and Wilkins. Baltimore: 286.
- Davies AP. 2002. Rating systems for total knee replacement. *The Knee*. 9: 261-266.
- Dawson J, Fitzpatrick R, Murray D, Carr A. 1998. Questionnaire on the perceptions of patients about total knee replacement. *Journal of Bone and Joint Surgery (Br)* 80-B: 63-69
- Finch E, Brooks D, Stratford P, Mayo N. 2002. *Physical rehabilitation outcome measures-A guide to enhanced clinical decision making*. Second edition. Lippincott, Williams & Wilkins. Ontario: 28-31.
- Gajdosik RL, Bohannon RW. 1987. Clinical measurement of range of motion – Review of goniometer emphasizing reliability and validity. *Physical Therapy*. 67(12): 1867 – 1872.
- Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, Katz B, Bombardier C, Heck D, Freund D. 1998. Health-related quality of life after Knee Replacement. *Journal of Bone and Joint Surgery (Am)*. 80(2): 163-173.
- Hicks CM. 1995. *Research for Physiotherapists- Project Design and Analysis*. Second edition. Churchill Livingstone. Singapore. 57-58.
- Insall JN, Dorr LD, Scott RD, Scott WN. 1989. Rationale of the Knee Society Clinical Rating System. *Clinical Orthopaedics and Related Research*. Nov (248): 13-14.
- John LM. 2004. The role of measurement reliability in clinical trials. *Clinical Trials*. 1: 553-566
- Kreibich DN, Vaz M, Bourne RB, Rorabeck CH, Kim P, Hardie R, Kramer J, Kirkley A. 1996. What is the best way of assessing outcome after total knee replacement? *Clinical orthopaedics and related research*. 331: 221-225.
- Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB. 2001. Validity and Responsiveness of the Knee society Clinical Rating System in comparison with the SF-36 and WOMAC. *Journal of Bone and Joint Surgery (Am)* 83-A (12): 1856-1864.
- Liow RYL, Walker K, Wajid MA, Bedi G, Lennox CME. 2000. The reliability of the American Knee Society Score. *Acta Orthop Scand*; 71 (6): 603–608.
- Magee DJ. 1997. *Orthopedic Physical Assessment*. Third edition. W.B. Saunders. Pennsylvania: 539-547.
- McAuley J, Harrer M, Ammeen D, Engh G. 2002. Outcome after Total Knee Arthroplasty in Patients with Poor Preoperative Range. *Clinical Orthopaedics and Related Research*. 404: 203-207.
- Petty NJ, Moore AP. 1998. *Neuromusculoskeletal Examination and Assessment*. A Handbook for Physiotherapist. First edition. Churchill Livingstone. London: 295.
- Rajan R, Pack Y, Jackson H, Gillies C, Asirvatham R. 2004. No need for outpatient physiotherapy following total knee arthroplasty: A randomized trial of 120 patients. *Acta Orthopaedica Scandinavia*. 75(1): 71-73.
- Ryd L, Karrholm J, Ahlvin P. 1997. Knee scoring systems in gonarthrosis: evaluation of interobserver variability and the envelope of bias. *Acta Orthopedic Scandinavia*. 68: 41-46.
- Stavem K, Arnesen Q. 2005. Use of hip and knee clinical scoring systems in prosthesis surgery in Norwegian hospitals. *International Orthopaedics (SICOT)*. 29: 301-304.
- Stillman BC. 2004. Physiological Quadriceps lag. Its nature and clinical significance. *Australian Journal of Physiotherapy*. 50: 237-241.
- Thomas SJ, Stewart A. 2005. Test-retest reliability, inter rater reliability and internal consistency of the post operative physiotherapy discharge scoring tool. Unpublished research report. University of the Witwatersrand. Health Sciences Library. Johannesburg.
- Walsh M, Kennedy D, Stratford P, Woodhouse L. 2001. Perioperative performance of women and men following total knee arthroplasty. *Physiotherapy Canada (spring)*: 92-100.
- Wright JG, Feinstein AR. 1992. Improving the reliability of orthopaedic measurements. *Journal of bone and joint surgery*. 74B: 287-291.