

https://dspace.org/

Intra-and inter-rater reliability of the knee society knee score, post total knee arthroplasty at Johannesburg Hospital.

Gopal, Sudhakar

2008-05-14T07:24:52Z
http://hdl.handle.net/10539/4797

Downloaded from a DSpace repository. https://dspace.org/

LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter will discuss the importance of outcome measures, their characteristics and factors influencing them as well as common outcome measures used following TKA. The literature will be reviewed in such a way as to discuss whether suitable outcome measures were used in the studies. In addition, the KSKS, its components and importance are discussed as well as the context in which the KSKS has been used as an outcome measure. Data bases searched included Pubmed, Pedro and the Cochrane collaborators. In addition a hand search was undertaken.

2.2 Importance of outcome measures

The use of outcome measures in health care services is gaining momentum. Objective assessment of intervention has become significant due to the culture of audit and research advances (Stavem and Arnesen 2005; Kreibich et al 1996).

According to Australian Physiotherapy Association (APA) position statement 2003, the significance of outcome measures is:

1. “To evaluate and document the demonstrable benefits of physiotherapy in relation to treatment goals”.
2. “Evaluate the need for continued physiotherapy to maintain or improve patient status”.
3. “Identify, document and act on factors that may compromise treatment outcomes or predict poor outcomes”.

Increased practice of physiotherapy assessment and evaluation has been emphasized by physiotherapy associations for quality service and professional accreditation (Charted society of Physiotherapy 2005).

Physiotherapists should use outcome measures that are appropriate to the patient's treatment goals (APA position statement 2003) and to compare baseline measurements with postoperative measurements in order to predict effects of interventions.

2.3 Characteristics of outcome measures

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 an intervention. An outcome measure must be valid, reliable and responsive (Kreibich et al 1996), to be used in intervention studies and controlled trials to detect change and should be related to the intervention (APA position statement 2003).

Validity refers to the extent to which the outcome tool measures the component that it is designed to measure (Kreibich et al 1996; Finch 2002). There are three basic types of validity - content validity, criterion validity and construct validity (Johnston et al 1992). Content validity refers to whether the tool measures the component items of a relevant feature that an examiner wants to assess in a subject. Criterion validity refers to the extent to which the assessment of variables using a tool is similar to that obtained by standard procedures which are set to assess the same variable.

Construct validity refers to the extent to which the tool produces similar results to the preformed theories about domains which cannot be measured using standard procedures (Finch 2002).

Reliability refers to the extent to which consistent results are produced after repeated measures (Seale and Barnard 1998) and hence reliability of the tool and examiners must be established. There are two types of reliability - interrater reliability and intrarater reliability. Interrater reliability refers to the extent to which a tool produces consistent results when used by different examiners during same instance of examination and mainly prevents an examiner from being biased on the tool (Johnston et al 1992; Hicks 1995). Intrarater reliability or test retest reliability refers to the extent to which a tool produces consistent results when used by the same examiner during two different times (Johnston et al 1992) provided there is no intervention done between the assessments.

A responsive outcome measure is said to predict changes in patient's impairments due to interventions over a period of time (Kreibich et al 1996). It is the character of a tool to measure even small changes in the patients condition which is important to establish the effectiveness of a treatment.

Hence outcome measures are tools that enable the treating physiotherapist to undertake an evaluation of physiotherapy treatment. They are used to evaluate changes in patients level of disability through impairments, activity limitation,

participation restrictions and quality of life. But there are certain factors that have to be considered while selecting outcome measures to prevent variability in results.

2.3.1. Common problems encountered in orthopaedic outcome measures

There are possible sources of error encountered while using outcome tools which must be considered during their administration to prevent false information or results or reports. In 1992, Wright and Feinstein discussed the common causes of variability in orthopaedic measurements. They stated that patient variability, procedure variability 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. They also stated that the conversion criteria used to categorise the actual score of the measurements into groups or categories may also play an important role in variability of the measures.

Brinker et al (1997) investigated the demographic biases on four commonly used knee scoring systems (Hospital for Special Surgery Knee Rating System (HSSKRS), Hungerford and Kenna, KSRS, System of Hofmann et al). The demographic features studied were age, gender, race and relative body weight. The component scores of the scoring system were normalized by dividing the obtained score by maximum possible score. Significant difference was found in normalised total scores according to age (greater than 80 years with lower scores) and family income (below poverty

line) and associated medical conditions (more than two major medical conditions). Gender, race and relative body weight did not have a significant influence on normalised total and component scores. Poverty level was based on the education, income and occupational status of the patients.

Hence the sources of common variability (Wright and Feinstein 1992) may arise from selection of tool, procedure and skills of the examiner and they should be minimised before using a tool. Other factors such as age, income and associated medical conditions also found to have influence over rating systems (Brinker et al 1997) must be minimised by correlating the results of groups based on such factors. There are several outcome measures available for post total knee arthroplasty which are not excempted from variability and hence the tool has to be selected based upon the aim of the study in which it has to be implemented. Some common problems following TKA and the outcome measures will now be discussed.

2.4. Common problems following TKA

There are various complications and impairments following TKA which need to be considered during follow up and the treatment should be directed towards minimising their effect on patients function. In 2005, Bhave et al studied the impairments following total knee arthroplasty and their interventions. Fifty one unilateral arthroplasty patients were examined by radiographs, movement analysis of gait using video, isokinetic strength testing, sensory conduction test and electromyography. Quadriceps weakness, flexion contractures, knee flexion deficit were the most common impairments detected in early post operative period (six to nine weeks).

Other impairments include peroneal nerve entrapment, limb length differences, malalignment of foot, superficial neuroma and ligamentous laxity.

Even though TKA has been said to bring about decrease in pain (Hawker et al 1998; McAuley et al 2002) and improvement in function (Hawker et al 1998; Walsh et al 2001), the chances for developing impairments as mentioned above are high. Thus it becomes essential to continuously assess patients postoperative to monitor any chances of developing such impairments. The KSKS includes components which measure the above mentioned impairments such as stability, flexion contracture, alignment and extension lag. There are other commonly used outcome measures for total knee arthroplasty, with their own advantages and variability. These will be discussed.

2.5 Common outcome measures used following TKA

2.5.1. Common outcome measures in practice

The outcome measures which are commonly used post TKA can be broadly classified as joint specific measures (focus on physical characteristics of joints such as movement, stability, alignment, pain and function) (Miner et al 2003), disease specific measures and general health measures (Hawker et al 1995).

Following are some commonly used outcome measures following total knee arthroplasty.

- Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)
- Short Form-36 (SF-36)

- Bristol Knee Score
- Oxford Knee Score (OKS)
- HSSKRS
- AKSCRS

The classification of 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 2.1.

Table 2.1. 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	Disease specific	Yes	Yes	No
SF-36	General health	Yes	Yes	No
HSSKRS	Joint specific	No	Yes	No
Bristol Knee Score	Joint specific	No	Yes	No
OKS	Disease specific	No	Yes	No
AKSCRS	Joint specific	Yes	Yes	Yes

WOMAC is a self administered disease specific measure in the form of questionnaire. It is designed with questions categorised into three sections which are stiffness, pain and physical function. The patients respond to questions in a five point linkert scale

ranging from 0 standing for worst and 5 for best. The validity and reliability of the tool has been established (Kreibich et al 1996; Davies 2002).

The SF-36 is reliable and valid (Aitken and Bohannon 2001) measure for patient's general health status. It involves a spectrum of health concepts categorised into physical and cognitive components (Lingard et al 2001). It also involves subject's comparison of their present health status with that of 12 months ago in the form of self administered questionnaires. Responses will be from "much better" to "much worse".

The Bristol Knee Score (Davies 2002) is a joint specific measure developed in 1970 and later modified in 1980 which increased the total score of 50 to 100. Twenty seven points were assigned to knee function, 40 points to pain and 33 for combined movement, alignment and stability (Davies 2002). Although it has good reliability (Davies 2002) the validity of the score has not been established. The knee and functional scores are not separated and hence there is a chance of functional deterioration affecting the total score.

The Oxford Knee Score (Davies 2002) is a disease specific outcome tool for total knee replacement. It is a self administered questionnaire with 12 questions which have to be rated in a five point scale. It includes questions regarding pain (during night and standing), physical abilities and functional activities (Davies 2002). The tool has been found to be consistent and responsive with its results.

The HSSKRS (Davies 2002) was published by Insall during 1974. It was designed to obtain a maximum possible score of 100 divided into six categories (pain-30, function of knee joint-22, ROM of knee-22, muscle strength-10, flexion deformity-10, and

instability-10). Negative scores included for use of walking aids, extension lag and valgus-varus deformity. The function and structure of the knee joint are scored together, and there is possible impact of functional deterioration on decreased total score (Davies 2002). The subjects are classified based on the scores as excellent (>85), good (70-84), fair (60-69 and poor (<60). The score has been shown to have a good overall reproducibility (Davies 2002). HSS has been replaced by American Knee Society Clinical Rating System.

2.5.2. American Knee Society Clinical Rating System

The AKSCRS is one of the most commonly used tool for total knee arthroplasty patients. It is a unique dual rating system. Insall et al (1989) developed this rating system for total knee replacement which was said to be an upgrade of HSSKRS addressing its problem of combining knee and functional score. The system was designed to score knee joint and its function separately with each scoring a maximum of 100, thus avoiding impact of functional and age related health problems on the knee joint itself. The KSKS consists of six parameters which include positive score for pain, ROM, stability and negative score for flexion contracture, extension lag, and malalignment.

The functional score consists of two parameters which are walking and stair climbing. The patient obtains a maximum score if he walks unlimited and climbs up and down stairs unaided. Deductions are made if assistive devices are used. Liow et al (2000) reported interobserver and intraobserver variations of knee score. Hence it is necessary to establish reliability of the examiners before applying it in controlled trials.

Lingard et al (2001) proved the validity and responsiveness of knee score while the functional score was found to be less responsive.

2.5.3. Significance of American Knee Society Clinical Rating System

In addition to being a dual rating system, the AKSCRS is a common and widely accepted tool due to its easy use. Stavem and Arnesen (2005) studied the prevalence of outcome measure used for hip and knee replacement procedures in various hospitals. Nine percent of the hospitals (n=56) performing knee procedures used standardised outcome measures for assessments and majority of them were university hospitals. The AKSCRS (80%) was the most frequently used for knee replacements.

In a prospective study, Konig et al (1997), assessed patients pre-surgery and three months, one and two years post TKA using AKSCRS. They emphasised the need for a dual rating system in assessment of the knee joint. Preoperative predictors such as knee score, function score, pain stability, ROM, walking distance, climbing stairs, body mass index, patient category, age, gender, implant type were compared with postoperative pain, knee score and function score. Functional score declined after a period of two years. Knee score remained constant. This is due to age and associated medical conditions deteriorating general health. They found that dual rating system of the score was important in preventing the knee score being affected by functional deteriorations due to aging and general health problems.

Davies (2002) reviewed the scoring systems available for total knee arthroplasty, and their advantages. The HSSKRS, Knee Function Assessment Chart, AKSCRS, Knee Rating Score Card, Guide to Recording Information about Knee Replacement, Hungerford and Kenna, Bristol Knee Score, OKS, SF-36 and WOMAC were reviewed. They analysed the rating systems on the basis of ease to use and validation. HSSKRS, AKSCRS, OKS, SF-36 and WOMAC were validated. They were categorized as good (WOMAC, SF-36, OKS), fair (AKSCRS) and poor (HSSKRS) on the basis of ease to use. HSSKRS has been replaced by AKSCRS.

Miner et al (2003) investigated the relationship between function and structure of the knee joint following TKA. Six hundred and eighty four patients were assessed preoperatively, three and four months postoperatively. WOMAC and knee ROM were used as outcome measures. They found that there is a poor correlation between function score and knee ROM. They found that the poor correlation was due to the compensation available from the hip joint. Therefore knee ROM may be limited even in patients who function better. Hence it is important to use a rating system such as AKSCRS which has two separate scores for structure and function of the knee joint.

Liow et al (2000) assessed the reliability of the AKSCRS. Observers with varying level of experience (2 to 36 months) administered the score on 29 subjects. The interobserver and intraobserver reliability of AKSCRS was assessed by two measurements taken with a time interval of two hours between them. The higher the experience of the examiners, the more consistent they were with their measurements.

Reliable use of AKSCRS requires administration by experienced and trained individuals.

There are other outcome measures such as inpatient medication, length of hospital stay, complications following TKA (Davies et al 2003; Denis et al 2006), period of rehabilitation following arthroplasty (Aitken and Bohannon 2001) satisfaction visual analogue scale and pain visual analogue scale (Bullen et al 2001). The six minute walk test and thirty second stair climb test (Kreibich et al 1996) are tests to measure functional capacity of the subjects. These outcome measures are all subjected to variations and therefore the examiners should be aware of possible variations and characteristics of an outcome measure.

From the above studies it is clear that the AKSCRS is a commonly used outcome measure following TKA (Stavem and Arnesen 2005). It is valid (Davies 2002) and the dual rating system prevents the knee score from being affected by the functional score (Konig et al 1997; Ali and Mangaleshkar 2006). It has been found easy to use when compared with the former rating system (HSSKRS) of the knee society (Davies, 2002). As a drawback, the reliability was found to vary with an examiner (Liow et al 2000) which makes it essential for the examiners to prove their reliability before implementing the tool.

2.6. Literature on characteristics of outcome measures used post TKA

The three main features of an outcome tool, which are said to be the reliability, validity and responsiveness, should be considered when selecting them for studies.

Kreibich et al (1996) conducted a study to establish the responsiveness of six different outcome measures following TKA. Outcome measures were randomly used on patients prior and after three and six months post TKA. The AKSCRS and WOMAC were most responsive and SF-36 and Time Trade Off were the least responsive.

Lingard et al (2001) investigated the validity and responsiveness of the AKSCRS by comparing it with WOMAC and SF-36. Six hundred and ninety seven patients were examined using KSCRS, WOMAC and SF-36. Validity of the AKSCRS was compared with that of WOMAC and SF-36 as the later two were validated tools. They found that AKSCRS has good convergent construct validity. The KSKS was found to be more responsive than the function score.

Aitken and Bohannon (2001) studied the responsiveness of SF-36 and Functional Independence Measure (FIM) in 28 patients by correlating with other patient outcomes (physiotherapy treatments, occupational therapy treatments and length of stay). They found that FIM to be more responsive than the SF-36.

Gajdosik and Bohannon (1987) reviewed the reliability of goniometer and the factors affecting it. They found the universal goniometer to be more reliable than any other instruments assigned for joint motion measurement with high correlation coefficient

showing good reliability (Rothstein et al 1983). They also found that one measurement is enough instead of doing repeated measures (Liow et al 2000). Reliability of the universal goniometer is found to be affected by complex joint motions, force applied using passive measurement, experience and accuracy of examiner and patient condition (oedema, adhesions, strength deficits, muscle hypertrophy).

Thus the studies insist that the outcome measures should be responsive, valid, reliable and acceptable (Kreibich et al 1996). The characteristics of the tool must be established in order to be used widely in randomised controlled trials.

2.7. Appropriate use of outcome measures in studies on post TKA patients

Irrespective of the nature of the studies (interventional or observational), the outcome measures must be chosen to measure the variables which are of interest to the examiners. While measuring structure of the knee joint all the variables such as stability, flexion contracture, extension lag and alignment must be included. The outcome meaures should be matched to the interventions used.

The studies about post total knee arthroplasty and the outcome measures used are listed in the Table 2.2, for the ease comparison of differences between them.

Table 2.2 Studies on total knee replacement (TKR) and the outcome measures used.

S.No	Articles	Intervention	Outcome Measures used	
			Joint specific	**Non Joint specific**
1	**Oldmeadow et al, 2002**	• TKR • Exercises • Gait training • CPM • Cryotherapy	Knee ROM	• Length of hospital stay • Discharge destination • IOWA level of assistance scale
2	**Davies et al, 2003**	• TKR • CPM • Slider board therapy • Exercises	Knee ROM	• Cost of hospital stay and medications • Readmission Development of complication • WOMAC pain score and function score
3	**Dennis et al, 2006**	• TKR • CPM • Active Knee ROM Exs • Isometric knee extension Exs • Weight bearing/Gait training	Knee ROM	• Timed up and go • WOMAC • Length of hospital stay
4	**Salman and Mangaleshkar, 2006**	• TKR	AKSCRS	--

5	**March et al, 2004**	• TKR • Mobilisation • CPM	--	• WOMAC • SF-36
6	**Vaquero et al, 2005**	• TKR	HSSKRS	--
7	**Gandhi et al, 2006**	• TKR	AKSCRS	--
8	**March et al, 2004**	• Mobilisation • CPM	--	• WOMAC • SF-36
9	**Rajan et al, 2004**	• TKR	Knee ROM	--

Davies et al (2003) investigated the effectiveness of continuous passive motion (CPM) therapy and slider board therapy in comparison with exercises. The interventions were aimed to improve the function of knee joint. Three groups of patient received exercise therapy in common and two groups received either slider board therapy or CPM in addition to exercise. The outcome measures used were cost for hospital stay and medications, readmission, development of complications, knee range of motion and WOMAC pain score and function score. The development of complications and readmission due to joint infection, cost of hospital stay and medications were not suitable to measure the specific status of the knee joint, which is what the intervention was aimed at. WOMAC is a subjective measure from the patient and therefore not appropriate for measuring physical features of a joint which must be more objectively measured. Even though the ROM was measured, other components of knee joint such as stability, contracture, extension lag and alignment were not measured. The reliability of the examiners in measuring ROM was not considered.

Oldmeadow et al (2002) used length of hospital stay, discharge destination, knee range of motion and IOWA level of assistance scale, to study the patient outcomes following TKA. Even though knee ROM, IOWA were found to be appropriate, the length of hospital stay and discharge destination were not joint specific. IOWA is again a subjective measure from therapists and does not measures the physical features of the knee joint. The reliability of the examiners in measuring ROM was not considered and other components of the knee joint were not measured.

Dennis et al (2006) studied the effect of CPM compared to conventional physiotherapy following TKA. The outcome measures used were active knee flexion and extension ROM, timed up & go test, WOMAC, and length of hospital stay. The treatment addressed joint integrity and patient function. The patient function was adequately assessed, but the only outcome measure to assess joint integrity was ROM. The other components such as strength, swelling, alignment were not measured.

March et al in 2004 compared patient costs and functional outcomes between unilateral and bilateral TKA patients following surgery. Mobilisation and CPM were interventions used postoperatively during stay at hospital. WOMAC and SF-36 were self administered functional outcome measures used during preoperative and later at 3 months interval for the first postoperative year. The outcome measures used can not measure ROM which is a physical measure of the knee joint which would have been affected by the interventions.

Caracciolo and Giaquinto (2005) attempted to find out predictive factors of the outcomes following TKA by comparing non operated outpatients with operated patients. 53 patients who underwent TKA were subjected to assessment preoperative, during admission to a rehabilitation centre, at discharge and in six months duration using WOMAC. No objective physical measure specific to knee joint were used in the study to compare with the functional measures of the knee joint.

Kennedy et al (2002) compared the differences in self-reported and physical outcome measures based on gender and group differences in patients waiting for TKA. One-thousand and forty-four patients with unilateral knee osteoarthritis were assessed one month preoperative. Self Paced Walk Test (SPW), Timed up and go test were functional physical outcome measures and Visual Analogue Scale (VAS), Lower Extremity Activity Profile (LEAP) were patient self perceived measures. Outcome measures mentioned are suitable to measure the variables of interest.

In 2006 Salman and Mangaleshkar, in their attempt to assess the surveillance of rotating-platform type prosthesis in TKA used AKSCRS to predict knee and functional score after 4-12 years of surgery. Even though the AKSCRS is a total score of objective measures of range, flexion contracture, stability, extension lag and alignment of the knee joint, the reliability of the examiners were not established in using the AKSCRS.

Vaquero et al in 2005 analysed the effect of age on the results of TKA. HSSKRS was used as the outcome measure which includes pain, walking, transfer, flexion

deformity, and instability, ROM, and muscle strength. Even though HSSKRS is a joint specific objective measure reliability of the examiner who administered the tool was not reported.

Gandhi et al (2006) investigated the predictive factors responsible for limited ROM in knee at one year post TKA in a retrospective study on 1216 TKA. The KSKS and OKS were recorded at preoperative, 6 weeks, 6 months and 1 year post operative. The reliability of examiners in administering the KSKS was not mentioned.

Rajan et al in 2004 investigated the effect of outpatient physiotherapy with that of inpatient physiotherapy services on patients with flexion contracture following TKA. Knee ROM was the only outcome measure used. The extension lag and alignment measurements which indirectly contribute towards the problem were not measured.

Bullen et al, 2001 attempted to establish the relationship between subjective and objective measures obtained from post total knee arthroplasty patients. Pain and satisfaction visual analogue scale were the subjective measures used while AKSCRS, WOMAC and Knee Society Roentgenographic evaluation system were the objective measures. The outcome measures were suitable to the variable of measurement. The reliability of the examiners was not considered in using the AKSCRS.

In the above mentioned studies, even though knee ROM (Dennis et al in 2006; Davies et al 2003; Oldmeadow et al 2002) was used as an objective measure, other features of the knee joint, such as stability, alignment and extension lag were not

considered. Most studies have used either functional outcome measures (Kennedy et al 2002; Caracciolo and Giaquinto 2005) or subjective measures (Cross et al 2000; March et al 2004) which may have influenced other factors such as compensation available at the hip joint (Miner at al 2003). Moreover the examiners reliability was not proved in most of the studies (Salman and Mangaleshkar 2006; Vaquero et al 2005; Gandhi et al 2006; Bullens et al 2001), which leaves the reliability of the data collected to be questionable.

2.8. Summary

In this chapter the importance and characteristics of outcome measures and their possible variations were discussed. In addition, the common outcome measures available for total knee arthroplasty patients and whether they were suitable to the studies in which they were used were reviewed. Most of the studies did not measure stability, extension lag, flexion contracture and alignment even though knee range of motion was measured. Even the studies which used scoring systems to include knee structure (stability, flexion contracture etc) did not consider the intra-rater and the inter-rater reliability of the examiners who administered them.