Introduction
Knee surgery is hugely diverse and encompasses arthroplasty, soft-tissue and cartilage regeneration procedures, the patellofemoral joint and fracture management. There is currently no globally accepted and validated outcome measurement tool for surgery relating to the knee. This lack of standardisation has resulted in a large number of unvalidated scoring systems, which have served to confuse rather than assist the surgeon's future decision-making. In order to assess the effectiveness of any intervention, an appropriate assessment system or combination of systems is necessary. Critically, the measurement tool should be both site and pathology specific.

When considering selection criteria for a scoring system, reliability, validity and responsiveness are essential properties. Reliability equates to the consistency (repeatability) of the system, it is not measured and can only be estimated. There are two ways that estimation of reliability can be performed: internal consistency and test-retest, both being important in orthopaedic scoring. Internal consistency groups the questions in a questionnaire that examines the same concept (e.g. instability after anterior cruciate ligament (ACL) reconstruction). Correlation between the groups of questions will determine if the system is reliably measuring the concept. Test-retest accounts for variation over time in stable patients. The assumption that there is no change in the underlying condition between, for example, test 1 and test 2, can be problematic with orthopaedic scoring because the time points are often widely spaced and joint function deteriorates in the interim.

Validity questions whether an instrument actually measures what is intended. Four types are commonly examined and all are relevant to orthopaedic scoring and outcome measurement. Conclusion validity asks if there is a relationship between the intervention and the observed outcome. Internal validity is similar but examines whether the outcome seen was causal. External validity looks at the ability to generalise the results of one study to other settings, a common practice in orthopaedic discussion. Finally, there is construct validity. This is the most commonly cited but most demanding concept to understand and refers to an ability to extrapolate study results to different settings. Meanwhile, responsiveness refers to a scoring system's ability to detect clinically important change over time.

This article examines the authors' preferred scoring systems in the assessment of outcome after knee surgery.

Knee degeneration
When considering the extensive knee degeneration that can be created by non-inflammatory and post-traumatic causes, the surgical options include replacement and osteotomy; arthrodesis will not be considered in this article. Significant advances have been made in prosthetic design and function. This has changed the emphasis from one of the alleviation of disabling pain and a limited return of functional activity as the primary end-point, to a more generalised improvement in quality of life and knee function. Expectations vary greatly between patients and the mismatch of experience versus expectation after knee replacement is a potent cause of patient dissatisfaction. Scoring systems in turn have evolved to accommodate more active patients at both ends of the age spectrum. As a result of earlier surgical intervention patients are now expecting not only pain relief, but also correction of any deformity added to an early return to physical and recreational activities. Currently, there is no single best outcome measure for total knee replacement. There are, however, several reliable, responsive and validated systems. The Western Ontario and McMaster University Osteoarthritis Index (WOMAC) and Oxford-12 disease-specific scores are most frequently used. The WOMAC underwent rigorous psychometric validation before its introduction and requires licensed use from the copyright holders. This may be obtained free online for educational and clinical use (www.womac.org). It is ubiquitous, easy to use and evaluates three domains; pain (five questions), stiffness (two questions) and physical function (17 questions), each weighted on a similar computation. The WOMAC Index is sensitive to change and has shown greater efficiency than most other instruments in the assessment of osteoarthritis. A seven-point reduced WOMAC scale has also been developed and retains excellent validity and repeatability in the assessment of total joint replacement. The Oxford-12 knee score (OKS), published in 1998, originally examined 12 items with a possible score of 1 to 5 for each. Scores thus ranged from 12 to 60, with 12 being the best outcome. Although simple, and ranked the highest for a disease-specific scale of reliability, content validity and feasibility of use, many have found the system unintuitive. It is now recommended that each question is scored from 0 to 4 with 4 being the best outcome. Thus, the new scoring system ranges from 0 to 48 with 48 representing the most favourable outcome. It is important that any study which incorporates the OKS clearly states which method has been used.
In contrast with the patient-assessed and equally-weighted OKS, the American Knee Society Score (AKSS) is a surgeon-assessed weighted score developed through consensus by the Knee Society in 1989. It comprises two parts, the first addressing pain, stability and range of movement. The second part examines function, with particular reference to walking distance and stair climbing. Maximum scores of 100 are possible in each section. The AKSS has been validated and is responsive and reproducible. However, it suffers from high inter- and intra-observer variation when the assessments are performed by less experienced doctors and nurses. In an attempt to isolate knee function from other factors, patients are categorised into three types: A, with no contralateral knee disease; B, with substantial arthrosis; and C, with multiple joint involvement. The final knee score is designed to be independent of other factors even in the face of declining function created by comorbidities and polyarthropathy.

The WOMAC and Oxford-12 would appear to be the most reliable and valid assessments of outcome after total knee replacement. However, with the increasing use of segmental replacements and osteotomy, scoring systems that examine higher levels of activity are required. The High Activity Arthroplasty Score (HAAS), although not specific to the knee, goes some way towards the subjective measurement of total joint replacement in patients who enjoy an otherwise active lifestyle. This system has been validated in patients receiving either total hip or knee replacements; we propose that it would be applicable to those receiving segmental replacements, high tibial and distal femoral osteotomies.

**Ligament reconstruction**

The assessment of ligamentous injury has resulted in more published outcome measures than any other area of knee surgery. Over 60 mainly unvalidated outcome measures have been produced, largely concentrating on the ACL-deficient knee. The considerable interest in ACL reconstruction and rehabilitation has generated conflicting views on the reliability, validity and sensitivity to change over time of the various scoring systems. When evaluating the results of ligament reconstruction around the knee it is important to be aware of the potential confounding factors that may affect outcome. Lower patient-reported outcomes after ACL reconstruction are strongly associated with obesity, smoking and severe chondrosis at the time of surgery. The modified Lysholm scale is one of the most commonly used scoring systems. Some have even claimed it to represent the ‘gold standard’ in the evaluation of the ACL-deficient knee.

First published in 1982, the Lysholm scale consists of eight questions, primarily aimed at the assessment of instability in younger patients. The score was designed to be physician administered and, therefore, risks the introduction of bias. It is a validated measure of disability; as such, low-demand patients tend to perform highly with this score as it does not evaluate high-performance knee stability or include a physical examination. The system focuses on the patient’s perception of function in those activities of daily living which are most important to the patient, and the patient’s functional level at various intensities of athletic activity. To complement the Lysholm score, the Tegner activity rating scale was introduced in 1985. This evaluates the patient’s level of work- and sports-activity handicap on an 11-level scale and is able to prognosticate to what physical level an individual may return, with and without reconstructive surgery. The combination of the Lysholm score and Tegner scale continues to show acceptable temporal responsiveness in the evaluation of early return to function after ACL reconstruction.

The Cincinnati knee rating system, introduced in 1983, was originally designed to assess ACL injuries but with an emphasis on the patient’s symptoms and their perception of knee function. As with the Tegner scale, the Cincinnati system examines physical abilities but in a more detailed fashion. Since its original publication, it has undergone several modifications and currently examines 11 functional components with particular reference to participation in sport. Other parameters include knee stability and radiographic findings. It is a complex and time-consuming system but has been shown to be reliable, responsive, valid and, critically, the most sensitive to change. The Cincinnati system is particularly useful in the evaluation of the multiple-ligament injured knee. The more recent literature would suggest that the most accurate instrument in the assessment of the ACL-injured knee is the Mohtadi quality-of-life (Mohtadi – Qol) questionnaire.

Designed as a disease-specific outcome measure, it is appreciably longer to complete than the Lysholm or Cincinnati questionnaires and employs a visual analogue scale in order to answer its 34 questions. There are five sections, which include the spectrum of symptoms and signs, impact on work, recreation, sport, social activities and emotional issues. It addresses those symptoms and disabilities that are felt to be most important to the patient. Unfortunately, the Cincinnati and Mohtadi systems lack evidence for internal reliability and construct validity. However, the various knee ligament-scoring systems show good correlation, in particular the Cincinnati and Mohtadi. This provides the surgeon with validated, standardised questions which should be asked of a patient with an ACL injury in order to determine knee function.

**Articular cartilage regeneration**

When addressing the impact of cartilage regeneration procedures, the Cincinnati rating system has proven popular with many groups. Recently the Knee Injury and Osteoarthritis Outcome Score (KOOS) and International Knee Documentation Committee Subjective Knee Form (IKDC) have also been proven in this area. The KOOS is an extension of the WOMAC index and was developed for younger, higher activity patients with knee injuries and arthritis. It addresses five subscales to be completed by the patient and is unique in that it reports health-related quality of life. Although proven to be valid and reliable it does need to be combined with a generic score to allow cross-study comparison.

The IKDC was initially developed as a ligament scoring system in 1987 by a group of American and European knee surgeons. They were concerned that the available scoring systems had assigned numerical values to factors that were not actually
quantifiable; arbitrary scores were then being added together for parameters which were not strictly comparable with one another. However, the current modified form is straightforward to use, is divided into documentation, qualification and evaluation sections, and examines four areas (subjective assessment, symptoms, range of movement and ligament examination). Additional information, which includes compartmental findings, donor site pathology, radiographic findings and functional abilities, are recorded but not used in the final evaluation. The qualification section is unique in that it has no numerical scores, merely a qualitative range from normal to severely abnormal. It is clearly a powerful scoring method but does not appear to be as generous a score as, for example, the Lysholm system. Rather than being cumulative, if a low grade is obtained for any section then the overall score can never be higher than this, however well the patient scores on the other parameters. Although originally designed for the assessment of ligament disruption, the IKDC has been shown to provide a superior overall measure of disability when compared with the KOOS, in patients who have undergone cartilage regeneration procedures.

**Patellofemoral pain and instability**

The assessment and management of patients with patellofemoral pain and/or instability is attracting increasing attention in the orthopaedic literature. The results of patellofemoral resurfacing and patellar realignment surgery are improving. In the analysis of anterior knee pain, the Kujala Anterior Knee Pain Scale has proven reliable and valid. It is a self-administered, weighted questionnaire that examines 13 domains, including pain and functionality. The score ranges from 0 to 100 with higher scores performing better. This system has also been used to assess the outcome after patellar dislocation. However, although valid with satisfactory test-retest reliability and superior performance to general health instruments, the Lysholm would appear to be the most sensitive scale, in particular for differentiating between patients with and without recurrent subluxations/dislocations.

**Fractures around the knee**

There is currently no validated, reliable and reproducible outcome measure for fractures around the knee, either on the femoral or tibial side. The most commonly used scoring systems quoted in the literature include the WOMAC, Short Form 36 (SF-36) and the Hospital for Special Surgery (HSS) knee score. The Rasmussen and Iowa scores address fractures around the knee specifically. The former, described in 1973, assesses subjective complaints of pain and walking capacity and clinical signs of knee extension, range of movement and stability. The latter, published in 1989, examines function, pain, gait, deformity and range of movement. With advances in the management of complex fractures around the knee, particularly on the tibial side, the development of a validated scoring system would appear to be a reasonable challenge.

**Discussion**

Outcome scoring is vital in the accurate evaluation of interventions around the knee. There has been a paradigm shift in the determinants of success over the last two decades, from those based on physical examination and radiographic variables, to a more patient-centred assessment of outcome. Modern knee surgery has allowed patients’ expectations and activity levels to increase but it remains difficult to accurately assess outcome. Evidence in the current literature confirms that few scoring systems have satisfactory levels of reliability and validity. What is clear is that those systems which employ a high degree of patient involvement, such as the KOOS, Oxford-12, Lysholm and IKDC scores perform better as a patient-based assessment tool.

This article has largely concentrated on pathology and knee-specific, patient-based systems. However, we would encourage the use of generic instruments to complement these. They have a greater potential to measure side-effects or unforeseen effects of treatment; the WOMAC in particular remains valid, reliable and responsive.

This article does not recommend a single best knee scoring system. Indeed, the holy grail of a short, easy to administer, reliable and valid global knee questionnaire does not currently exist. Consequently, knee function scores are likely to be based around components of a “Total Knee Function Questionnaire” such as that proposed by Philip Noble. We would urge that the scope of this questionnaire is extended and revalidated, in order to include patient populations other than those receiving total knee replacements.

**Mr Simon Tilley**, BSc DM FRCS (Tr & Orth)
Arthroplasty Fellow
Specialist Orthopaedic Group, Mater Clinic, Suite 1.08, Level 1, 3-9 Gillies Street, North Sydney, NSW 2060, Australia

**Mr Neil Thomas**, BSc (Hons) MB BS FRCS
Consultant Orthopaedic Surgeon
The Wessex Knee Unit, Wessex Nuffield Hospital, Winchester Road, Chandlers Ford, Eastleigh, Hampshire, SO53 2DW
The Hampshire Clinic, Basing Road, Old Basing, Hampshire, RG24 7AL
Basingstoke and North Hampshire NHS Foundation Trust, Aldermaston Road, Basingstoke, Hampshire, RG24 9NA

**References**