Systematic Review

Assessing activity participation in the ACL injured population: a systematic review of activity rating scale measurement properties

Robert Letchford[1,2,3], Kate Button[1,2,4], Valerie Sparkes[1,2], Robert WM van Deursen[1,2]

[1]Research Centre for Clinical Kinaesiology, School of Healthcare Studies, Cardiff University, UK, [2]Arthritis Research UK Biomechanics and Bioengineering Centre, Cardiff University, UK, [3]Aneurin Bevan Health Board, Royal Gwent Hospital, UK, [4]Cardiff and Vale University Health Board, University Hospital of Wales, UK

Background: Participation is an important factor in assessing both the requirement for and outcome from anterior cruciate ligament reconstruction. Many patient-reported rating scales exist; however, measurement properties have not been well established.

Objective: To provide a systematic review to identify currently available activity rating scales for anterior cruciate ligament injured subjects and to evaluate current knowledge of their measurement properties.

Methods: Systematic searches of four databases (Medline, AMED, EMBASE, and CINAHL) without date or language restriction, using terms structured around the PICOS system were completed on 1st March 2011. Citation tracking, reference screening and contact with lead authors of key papers completed the search strategy. Studies using participation rating scales were identified to assess frequency of reporting and cited validation. Studies assessing one or more psychometric properties of the identified rating scales were subject to independent data extraction and critical appraisal by two independent authors using published tools.

Results: Thirty-one rating scales were identified from 241 outcome studies. Most scales were inadequately developed or validated prior to use. Only three scales (Tegner, CSAS, and Marx) had psychometric analysis in eight studies of mixed quality. Only the Tegner scale has adequate assessment of reliability, validity, and responsiveness. The use of type, intensity, and frequency variables in the identified scales is discussed.

Conclusions: The Tegner scale has been adequately validated; however, other rating scales require further validation. A comprehensive comparative analysis of clinical applicability and psychometric testing of existing scales, including clinically useful statistics, is required.

Keywords: Anterior cruciate ligament, Participation, Patient-reported outcome, Systematic review

Introduction

The World Health Organization International Classification of Functioning, Disability and Health (WHO ICF) describes four core measures of health: structure, function, activity, and participation.[1] The outcome of anterior cruciate ligament (ACL) reconstruction is reported with robustly validated measurement tools for structure (e.g. magnetic resonance imaging),[2] function (e.g. International Knee Documentation Committee subjective knee form),[3] and activity (e.g. functional hop tests).[4] However, participation is more complex to measure, and tools are less clearly validated.

Outcomes in the participation domain assess 'involvement in life situations'[1] a measure of what a person is doing. This is important in ACL injury since both clinicians[5] and patients[6] consider participation restrictions an important factor when deciding on management and resolution of these restrictions is an expected outcome of surgical reconstruction[7,8] that is related to patient satisfaction.[9] For researchers participation is important since it may influence the significance of outcomes from other domains of the ICF.[7] For instance, outcomes from the function domain are frequently based on symptoms which can be dependent on the type and intensity of activities in which a person participates. Similarly, activity based measures are only significant if the ability to execute a task also translates into the ability to participate in

that task. Accurate assessment and reporting of participation is therefore required for clinical decision making and appropriate interpretation of clinical and research outcomes.[10–12]

Whilst many methods for assessing participation have been suggested, patient-reported outcomes have gained acceptance in the ACL literature and clinic. These tools are commonly referred to as activity rating scales, a name that was established prior to the current ICF classification that can lead to confusion over which ICF domain is being assessed. Despite their name, activity rating scales provide a quantification of what a person is doing and therefore fulfil the participation domain of the ICF. The commonly stated aim of anterior cruciate ligament reconstruction, to restore participation at pre-injury levels,[13] can be assessed adequately with a yes or no grading to define success for that individual. However, given that many patients may not share this aim,[14] or will never achieve it[15,16] a more sophisticated tool is often required. Similarly, when considering the impact which activity participation may have on the outcome of an intervention the tool must be capable of both quantifying an individual's level of participation and defining groups for comparative analysis.[17]

Authors of the first rating scales[10,11,18–20] should be commended for identifying the importance of assessing participation in the ACL injured population and for developing rating scales. However, by current standards, many were poorly developed and insufficiently investigated before being introduced. There are fears that inadequately validated patient-reported outcomes may in part be responsible for the large variations seen in published outcomes.[21] Psychometric testing has been applied to a range of outcome instruments and now forms the standard by which such tools are assessed.[22,23] This assessment should include clinical practicality and applicability, reliability, validity and responsiveness as minimum requirements.[22,23] Many reviews of patient-reported knee rating instruments have been conducted,[24–30] few have addressed participation rating[31] and none have focused on measurement properties of activity rating scales.

A well validated measurement tool must also be useful in the clinical setting and capable of influencing intervention decisions. For this the tool must measure those aspects of participation that are of most concern to patient and clinician. The components of participation assessed by each tool need to be understood in the context of how each may influence the outcome, and therefore how they can be appropriately applied in the clinical environment.

This systematic review will identify activity rating scales used in the ACL injured population, examine the available evidence for psychometric properties,

and discuss the clinical implications of tools based around different components of participation.

Study Aims

1. Systematically identify all currently available activity rating scales for assessment of participation in ACL injured subjects.
2. Where evidence exists, provide a systematic critical appraisal of each scale which will include an evaluation of design, psychometric properties, and clinical implications of the components assessed.

Methods

A two stage systematic search was conducted with no date or language restriction in four medical databases (Medline, AMED, EMBASE, and Cinahl) up to the 1st March 2011. Keywords were developed using the PICOS (participants, interventions, comparators, outcomes, and study design) system and adapted for each database to account for differences in index terms. The full search strategy for Medline is detailed in Table 1. Keywords included; participants: Anterior Cruciate Ligament, ACL, Anterior Cruciate Ligament reconstruction, Anterior Cruciate Ligament rupture; interventions: outcome assessment, questionnaire, measurement, treatment outcome; comparators: none were specified; outcomes: sport, participation, exercise, physical activity, return to sports; study design: reliability, validity, responsiveness and psychometrics.

The first search combined participant, intervention and outcome keywords to identify all studies reporting activity participation as an outcome of interventions for ACL injury. Studies were abstract filtered for appropriateness on the following criteria; a study that includes a patient-reported tool or questionnaire, measuring physical activity participation in adult

Table 1 Full search strategy used in Medline

1	Exp* Anterior cruciate ligament	7652
2	ACL.mp†	6456
3	1 OR 2	10 289
4	Exp 'reproducibility of results'	207 148
5	Reliability.mp	71 199
6	Validity.mp	76 552
7	Responsiveness.mp	67 759
8	EXP Psychometrics	43 078
9	4 OR 5 OR 6 OR 7 OR 8	348 327
10	EXP Outcomes assessment (healthcare)	505 855
11	EXP Questionnaires	227 746
12	Rating scales.mp	50 927
13	Measurement.mp	330 919
14	EXP treatment outcome	471 671
15	10 OR 11 OR 12 OR 13 OR 14	1 056 053
16	EXP Sports	88 865
17	Return to sport.mp	208
18	EXP recovery of function	19 977
19	Participation.mp	92 438
20	Physical activity.mp	34 869
21	EXP exercise	53 931
22	16 OR 17 OR 18 OR 19 OR 20 OR 21	259 219
23	3 AND 15 AND 22	362
24	3 AND 9 AND 15 AND 22	22

Note: *EXP=explode search term; †mp=keyword search.

ACL injured or reconstructed subjects. Full text articles were collected only for those studies where the rating scale was not identifiable from the abstract. Rating scales were extracted and registered in an electronic database. Reference screening for additional outcome studies and papers of psychometric testing was completed.

The second search strategy was to identify studies reporting psychometric properties of the identified rating scales. Scales were accepted on the basis of frequent use or the presence of reference to psychometric analysis. The initial search was combined with the study design keywords, and reference screening; additional searches using keyword, title and author were completed for each of the accepted scales. Papers were reference screened and the initial publication of each tool was citation tracked. Authors of accepted scales were contacted by e-mail with the search results and a request for information of any further published or unpublished material. Each psychometric study was abstract screened for appropriateness for one or more of the following criteria; describing tool development; detailing instructions for use of the tool; or reporting at least one psychometric property. Full text articles were obtained for all psychometric studies.

Critical appraisal and data extraction were completed using published tools designed specifically for psychometric studies; a detailed description of their use has been previously described.[23] Both tools have been used in recent reviews of patient-reported outcome instruments; for neck pain,[23] shoulder function,[32] and obesity management.[33] They have been investigated for inter-rater agreement ($k=0.43$ to 0.92) and inter-rater reliability (ICC=0.91: 95%CI=0.86–0.94).[23] Appraisal and extraction processes were completed separately, by two independent reviewers (RL and KB). If disagreement was irresolvable by a consensus on a second review of the manuscripts, plans were made for a third reviewer to be consulted and consensus agreed.

The critical appraisal tool assesses the quality of reporting for 12 criteria spread across five categories: study question; study design; measurement; analysis; and recommendation (see Table 2). Each criterion is scored 0, 1 or 2 points with a high score indicating high quality of reporting. The scores can be summed and expressed on a scale from excellent to poor, although analysis of each individual criterion is encouraged.[23]

The following data were extracted from each study following the published guidance;[23] population, intervention, clinical usefulness/practicality (readability, interpretability, time to administer, administration burden, and cultural applicability), reliability [relative reliability reported with intraclass correlation coefficient (ICC), absolute reliability reported with standard error of the mean (SEM) and minimal detectable change (MDC)], content/structural validity (internal consistency, content validity, floor and ceiling effects, factorial validity, item response, and Rasch analysis), construct/criterion validity (known groups, convergent, divergent, longitudinal, concurrent criterion,

Table 2 Evaluation criteria from the critical appraisal tool. Reproduced with permission from MacDermid *et al.*[23]

Evaluation criteria	Score		
Study question	2	1	0
1. Was the relevant background work cited to define what is currently known about the psychometric properties of measures under study, and the potential contributions of the current research question?			
Study design			
2. Were appropriate inclusion/exclusion criteria defined?			
3. Were specific psychometric hypothesis identified?			
4. Was an appropriate scope of psychometric properties considered?			
5. Was an appropriate sample size calculated?			
6. Was an appropriate retention/follow up obtained?			
Measurements			
7. Were specific descriptions provided of the techniques used to collect the measurements reported?			
8. Did measurement procedures use standardized techniques (and other methods required) to minimize potential sources of error/misinterpretation in the individual measures taken from the study?			
Analysis			
9. Were analyses conducted for each specific hypothesis or purpose?			
10. Were appropriate statistical tests conducted to obtain point estimates of the psychometric property?			
11. Were appropriate ancillary analysis done to describe properties beyond the point estimates (CI, benchmark comparisons, SEM/MID)			
Recommendations			
12. Were the conclusions/clinical recommendations supported by the study objectives, analysis and results?			
Subtotals (of columns 1 and 2)			
Total score (sum of subtotals/24 × 100)			
If an item is deemed appropriate then you can sum the score of items/2× number of items × 100.			

and predictive criterion) and responsiveness/clinical change [responsiveness reported with effect size (ES), standard response mean (SRM) and minimal clinically important difference (MCID)].

Previous appraisal of activity rating scales has provided well reasoned and valid criticism.[31] Two of these criticisms are of particular importance, are unrelated to psychometrics, and would not be captured by the data extraction tool. These are the ability to detect the knee abuser (a person who participates in a high level of activity despite significant knee symptoms) and the ability to detect alterations in activity unrelated to the knee. These two criteria were assessed during data extraction on a simple yes/no scale.

Results

Results of the search strategy are presented in Fig. 1. After initial title and abstract screening 325 outcome studies were reviewed, nine were excluded (five review papers, one discussion paper, one retracted by publisher, one not available in British library, and one French language paper), leaving 316. Of these 75 (24%) did not report activity participation, leaving 241 studies using 37 different rating scales. Six scales did not meet inclusion criteria and were excluded; four rated symptoms with activity (Knee Outcome Survey Sports Activity Scale, Knee Osteoarthritis Outcome Scale, Orthopadischen Arbeitsgruppe Knie, ACL Return to Sport Index); the activity participation
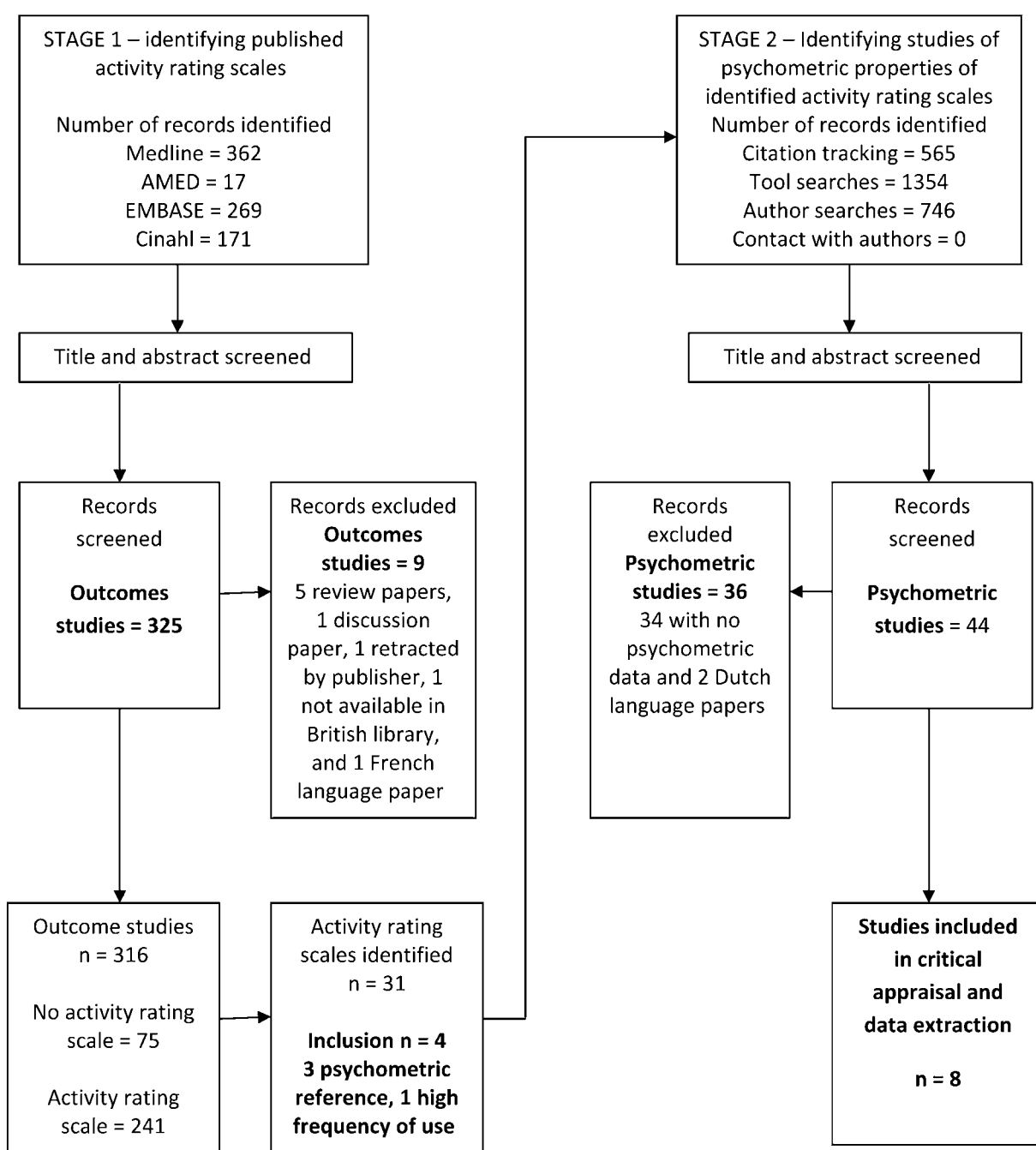

Figure 1 Flowchart showing systematic search process

rating was not reportable as a separate subscale in one (ACL Quality of Life); and one was not published in English (Innsbruck).

A total of 31 rating scales were identified. They can be grouped into three themes; scales that assess return to sport based on either pre-injury comparisons or time,[9,10,34–50] scales which rank sports or activities[7,18,19,51,52] and those which rank knee motions and forces.[10,17,20,53,54] Only five scales take account of frequency of participation,[10,17,42,47,52] and only three the intensity.[7,18,19]

The most frequently used scale is the Tegner activity rating scale ($n=121$), followed by IKDC activity rating ($n=38$), Marx activity rating scale ($n=7$) and Cincinnati sports activity scale ($n=5$). Twenty-five authors used their own scales with no reference to validation of psychometric properties. The Marshall score was used in three studies[34,55,56] and the University of California activity scale in one[57] with no cited validation.

Studies of psychometric properties ($n=8$) were identified for just three of the scales; Tegner activity rating scale ($n=5$),[11,12,58–60] Cincinnati sports activity rating scale ($n=2$),[10,61] and Marx activity rating scale ($n=1$).[17] The authors of each of these scales were contacted by e-mail for missing publications or unpublished data, all replied but no new material was identified. Due to the high reporting rate of the IKDC activity scale and the international standing of the authors, this scale was included in the second search strategy and an IKDC team member was contacted (Dr JJ Irrgang), but no study specifically assessing the activity rating scale was identified.

During critical appraisal there was agreement between the two reviewers on 70 of the 84 questions (83%); where there was disagreement it was by a maximum of one point and all were fully resolved by consensus. The quality of the studies was variable (see Table 3), with none scoring greater than 90% which is the criteria to be classified as excellent. Three were ranked as very good (71–90%);[58–60] two as good (51–70%);[12,61] one as fair (31–50%);[17] and two poor ($\leqslant 30\%$).[10,11]

There were no disagreements between the reviewers in the data extraction. The results are presented in Table 4 and data for the two additional criteria in Table 5. The results will be described in detail for each of the tools below.

International Knee Documentation Committee (IKDC) Activity Rating Scale

The IKDC activity rating scale did not meet the primary inclusion criteria for the review since there were no psychometric studies available. However, as the second most reported tool, used in 15.8% ($n=38$) of the identified outcome studies, it was entered into search two to assess its validation status. The IKDC was formed in 1987 by a group of leading knee surgeons with the goal of determining a common terminology for the evaluation of knee ligament injuries. Activity participation was included on the resulting knee evaluation form[54] and the subsequently developed subjective knee form.[3] The scale uses a four point scale (Level 1 jumping, pivoting, hard cutting, football, soccer; Level 2 heavy manual work, skiing, tennis; Level 3 light manual work, jogging, running; level 4 sedentary work and activities of daily living) to describe activity participation pre-injury, pre-treatment and present, with a yes/no statement to establish whether changes are knee related. Whilst both the evaluation form and subjective knee form has received appropriate validation, no papers assessing psychometric properties of the activity rating section as an independent measure were identified. The scale will detect knee abusers when used in conjunction with the subjective knee form and will identify non-knee related changes.

Tegner Activity Rating Scale

The Tegner activity rating scale was introduced in 1985 as a measure of handicap as described by the World Health Organization international classification of impairment, disability and handicap.[7] Designed to complement the Lysholm knee scale[11] it was intended to assess changes in activity participation for a single subject over time rather than differences between subjects.[7,62] The authors should be praised as

Table 3 Results of the systematic critical appraisal of each psychometric paper

| Study | Scale | Item evaluation criteria | | | | | | | | | | | | Total % |
		1	2	3	4	5	6	7	8	9	10	11	12	
Barber-Westin *et al.* (1999)[61]	CSAS	2	1	1	1	1	2	1	0	0	1	1	2	54
Briggs *et al.* (2006)[58]	Tegner	1	1	2	2	1	2	1	1	2	2	2	2	79
Briggs *et al.* (2009)[59]	Tegner	2	1	2	1	1	2	2	1	2	2	2	2	83
Briggs *et al.* (2009)[60]	Tegner	2	2	0	1	1	NA	2	2	2	2	1	2	77
Noyes (1989)[10]	CSAS	1	0	0	0	0	0	2	0	1	0	0	0	17
Marx *et al.* (2001)[17]	Marx	1	1	1	1	0	1	1	0	1	1	0	1	38
Paxton *et al.* (2003)[12]	Tegner	1	2	1	1	0	1	1	1	2	2	0	2	58
Tegner and Lysholm (1985)[11]	Tegner	0	1	0	0	0	0	0	0	1	1	0	1	16

Note: Each of the 12 questions scored from a maximum 2 and minimum 0.

Table 4 Psychometric data extracted from each of the included studies

Study	Tegner and Lysholm (1985)[11]	Paxton et al. (2003)[12]	Briggs et al. (2006)[58]	Briggs et al. (2009)[59]	Briggs et al. (2009)[60]	Noyes (1989)[10]	Barber-Westin et al. (1999)[61]	Marx et al. (2001)[17]
Population	ACL deficient (n=73)	Acute (n=110) and recurrent (n=43) patellofemoral dislocation	Meniscal injury	ACL reconstruction (n=505)	Healthy (n=500) 255 M; 245 F	Chronic knee injured (n=59)	Knee injured (n=50)	Knee injured (n=50)
	55 M; 21 F		Isolated or combined (n=122) 77 M; 45 F	Age 37 years (18–77 years)	Age 41 years (18–85 years)		28 M; 22 F	27 M; 23 F
							Age 36 years (13–65 years)	Age 32.8 years (12–57 years) item generation/ reduction only Healthy (n=40)
			Age 48 years (14–76 years)				Healthy (n=50) 22 M; 28 F Age 34 years (20–59 years)	27 M; 13 F Age 33.7 years (18–50 years)
Reliability	N=76 Retest	N=119 Retest 21 days (13–42)	N=122 Retest <4 weeks	N=50 Retest <4 weeks	–	–	N=100 Retest 4–13 days	N=40 Retest 1 week
	ICC= 0.90–0.97	ICC=0.92	ICC=0.817 (95%CI= 0.75–0.87) SEM=0.4 MDC=1	ICC=0.82 (95%CI= 0.66–0.89) SEM=0.64 MDC=1			ICC=0.98	ICC=0.97
Content validity	–	N=153	N=80	N=687	–	–	–	Excellent item generation and reduction process
		Floor=0% Ceiling=0%	Floor=2.5–5.5% Ceiling=0–2%	Floor=8% Ceiling=3%				
Known groups validity	–	–	N=80 All 8 hypothesis significant at P<0.05	N=687 All 6 hypothesis significant at P<0.05	–	–	–	
Convergent validity	–	–	–	–	Lysholm stairs (p=0.20; P=0.01) and squat (p=0.14; P=0.001)	–	–	All at P<0.05 Tegner r=0.66 Daniel r=0.52 CSAS r=0.67
Divergent validity	–	Knee rating scales r=0.20–0.54	–	–	Inverse correlation with age P=−0.44 P=0.01 Differed by gender P=0.001	–	–	Age r=−0.42; P=0.02
Criterion validity	–	–	N=80 SF-12 physical score r=0.46; P<0.05	N=170 SF-12 physical score p=0.2; P<0.05 IKDC p=0.22; P<0.001	–	–	–	–
Responsiveness	–	–	N=80 Isolated ES=0.61 SRM 0.6 Combined ES=0.83 SRM 0.704	N=505 ES=1.0–1.1 SRM=0.84–1.2	–	–	–	–

Note: N=number of subjects; M=male; F=female; ICC=intraclass correlation coefficient; SEM=standard error of measurement; MDC=minimal detectable change; ES=effect size; SRM=standard response mean; IKDC=international knee documentation committee. Blank sections indicate that no data were presented.

Table 5 Capability of rating scale to detect the knee abuser and non-knee related changes in activity participation

Scale	Detect knee abuser	Detect non-knee related change
IKDC	Yes	Yes
Tegner	Yes	No
CSAS	Yes	Yes
Marx	Yes	No

being the first to develop such a tool, which has now become the most frequently used participation rating scale for various knee disorders.[58]

The original manuscript[11] contains a briefly explained rationale. The development process is inadequately described with no literature review presented. Whilst patients were used in the item generation there is minimal description of item reduction. The text states the use of a troublesome rating which is not clearly defined. The scale is scored out of 10 points using the Gutman method. This method is applied to ranked scales, the score is given to the highest selected level, and it is inferred that lower ranked levels are also included. Brief descriptors of competitive and/or recreational sports, occupations and activities of daily living are provided to describe each level. Participation in national and international soccer defines the top score of 10 and currently on sick leave due to knee injury a low score of 0. It is suggested that results should be presented as a retrospectively reported pre-injury and prospectively reported post-injury and post-intervention score.[7] Rodkey *et al.*[62] have proposed a method to enable changes in the results to be normalized across diverse populations, which they term a Tegner index.[62] The index is calculated from pre-injury, preoperative and postoperative scores and indicates the percentage of lost activity level that is regained following an intervention.

Five studies reporting psychometric properties of the Tegner scale were identified, of which three were assessed as very good;[58–60] one good;[12] and one poor.[11] Study samples included healthy individuals,[60] and patient groups with ACL injury and reconstruction,[59] meniscal injury[58] and patellofemoral dislocation.[12] There is no evidence for clinical usefulness/practicality of the tool. Whilst it has been suggested that the European bias of the listed sports and activities may hinder understanding particularly in the USA, no evidence for administration burden, readability or interpretation is presented. There is evidence of excellent test–retest reliability from four studies,[11,12,58,59] with reported ICCs ranging from 0.82 to 0.97, SEM from 0.4 to 0.64 and MDC of 1. There is good evidence for validity of the scale with reports of acceptable (<30%) floor and ceiling effects in three studies,[12,58,59] known groups validity in two studies,[58,59]

convergent validity (Lysholm stairs and squat subscales) in one study,[59] divergent validity (knee rating scales, age, and gender) in two studies[12,59] and criterion validity (SF-12 physical and IKDC) in two studies.[58,59] There is evidence for acceptable responsiveness with moderate to large effect sizes and SRMs reported in two studies.[58,59] Usefulness and practicality are inadequately investigated and reports of responsiveness lack the clinically useful statistic of minimal clinically important difference. The scales developer was involved in two of the studies whilst three were performed by independent groups.

The scale has been criticized for: ill defined use of arbitrary descriptors such as competitive and recreational;[31] the ill defined method in which sports are ranked;[31] simultaneous assessment of occupational and sporting activity;[31] occupational activities that are not comparable to sporting activity at the same level;[31] a dominance of European sports leading to a lack of cross cultural application,[58] all of which remain valid. The scale is capable of detecting knee abusers when used in conjunction with the Lysholm scale, but does not allow for detection of non-knee related changes in activity.

Cincinnati Sports Activity Scale (CSAS)

First published in 1989 as one of 13 subscales that form the Cincinnati Knee Rating System, the CSAS is based around two criteria, frequency of participation and specific knee functions. The original manuscript[10] offers a very well reasoned and evidenced basis for development of the tool, but lacks evidence of item generation and reduction techniques or psychometric testing.

The resulting tool is a 12-point scale scored out of a possible 100 points using a Gutman method. The dominant criterion, frequency, is assessed through four levels: level one 4–7 days/week, level two 1–3 days/week, level three 1–3 times/month and level four no sports. Each frequency level is divided by specific knee functions; category one activities are jumping, pivot cutting; category two are running, twisting and turning; category three are none of these. Points are allocated to each group in descending order of frequency and knee function, a maximum 100 for level one pivoting sports and zero for severe problems with level 4 activities.

Two studies reporting psychometric properties of the CSAS were identified, of which one was assessed as good[61] and one poor.[10] There is little evidence for clinical usefulness and practicality. It is suggested that data can be presented in five different ways:[10] as a percentage of the population at each level to describe groups; an individual's level; change in level; agreement with patient's goals; or level deemed acceptable to the clinician, which makes it confusing

for clinicians when reporting and interpreting results. There is no evidence for readability or interpretability and there are concerns that terms such as pivoting and cutting require more explanation than is given. There is evidence of excellent test–retest reliability (ICC=0.98) in one study;[61] however, no other psychometric data are available for the CSAS despite a comprehensive analysis of validity and responsiveness for the other subscales of the Cincinnati Knee Rating System.[61] It should also be noted that the reliability data are from two sample subsets consisting of 50 healthy and 50 chronic knee injured subjects, not from the larger ACL reconstructed group used for validity testing of the other subscales. There is no evidence for validity or responsiveness. Both studies were conducted within the research team of the tool developer.

When used with the appropriate counterparts of the Cincinnati Knee Rating System, the CSAS will detect knee abusers and non-knee related changes in activity. The CSAS scale is highly dependent on frequency of participation at the expense of both intensity and type of task.

Marx Activity Rating Scale

First published by Marx *et al.* this tool was developed to enable the comparison of activity participation between study groups and aid generalizability of research findings, although it has subsequently been recommended[30] and used[63] as a longitudinal outcome score. The original manuscript[17] provides an appropriately reasoned and defined development process, there is adequate review of the literature, and appropriate item generation (consultation with specialists and 20 knee injured patients) and reduction (50 knee injured patients ranking for importance and severity).

The scale is based upon the criterion of specific knee functions and frequency of participation. Each of four knee functions (running, cutting, decelerating, and pivoting) are rated on a five point scale of frequency (<1 time in a month, 1 time per month, one time in a week, 2 or 3 times in a week or 4 or more times a week) and scored between zero and four points. The scale is scored by adding the scores to give a total out of a possible 16 points, with a higher score indicating more frequent participation.

One study reporting psychometric properties of the Marx activity rating scale was identified, which was appraised as fair.[17] This study establishes usefulness and practicality in a knee injured patient group. There is evidence for excellent test-retest reliability (ICC=0.97), correlation with three other activity rating scales (Tegner, Daniel,[20] and CSAS) demonstrates evidence for convergent validity and an inverse correlation with age demonstrates evidence for divergent validity. All psychometric analyses were

conducted in a small healthy sample (*n*=40). Reliability data are limited as SEM, MDC and confidence intervals (CI) are not reported, and there are no responsiveness data, which can be explained by the initial intention not to use the tool for longitudinal follow up. The study was conducted by the tool developer.

The scale is very specific to those activities which challenge the ACL injured athlete and is very frequency dependent. Descriptions are usefully worded, to eliminate ambiguity and facilitate comprehension. There is a concern that the scale will not adequately differentiate those people who remain active in non-knee aggravating activities from those that give up sports entirely, creating a possible floor effect that requires investigation. The IKDC subjective knee form is recommended as an adjunct scale[17] which would allow detection of the knee abuser, but there is no way of detecting non-knee related changes.

Discussion

The systematic search strategy was comprehensive and compares well to recently reported searches for the Tegner activity scale,[59,63] providing reassurance that results were not affected by missing studies. There were fewer than the 208 papers reported by Lysholm and Tegner[29] and 200 reported by Briggs *et al.*;[59] however, this would be expected since both these searches included all knee conditions, whilst the current review was restricted to ACL injured subjects. A horizon analysis could be used to estimate missed articles, but was not performed in this study. It is possible that selective reporting bias limits the presentation of negative results in published studies. There was evidence of missing data for some hypotheses in one of the included studies.[61] It is possible that the review was affected by publication bias; however, contact with authors in the field indicated that they were not aware of any unpublished data relating to their scales.

Language restrictions were not applied to the searches, and of the 30 foreign language papers identified in the first search only one was excluded due to lack of an English abstract. Two Dutch language psychometric papers identified in the reference screening of one paper[64] were not included. Both these papers reported on the CSAS, with the Neeb *et al.*[64] report presenting reliability data (*r*=0.64–0.84)[65] and correlation of the Tegner and CSAS with the Lysholm score (Kendall Tau b of 0.29 and 0.47 respectively).[66] This reliability data are slightly lower than that reported by Barber-Westin *et al*,[61] but does not substantially influence the results of this review. Translation of these articles was not attempted, which is a potential weakness in the study design.

None of the studies scored excellent on the critical appraisal. The three highest scoring studies (very good) were also the most recent and conducted by the same research team.[58,59,60] The introductory texts to all three scales scored badly with two in the poor[10,11] and one in the fair[17] category. These early studies were completed before psychometrics had gained recognition in the development of patient-reported outcomes in orthopaedics, which may explain why earlier studies performed poorly and that quality has improved with time.

Analysis of the appraisal data shows particular deficiencies in three criteria. Sample size was poorly justified in all studies with none including an appropriately described power analysis. Standardizing of measurement procedures was often not clearly described, leaving doubt over how data were collected and handled. Finally, there was inadequate use of appropriate ancillary analysis to describe properties beyond the point analysis. Test–retest reliability is frequently reported; however, data are often restricted to the intraclass correlation coefficient and there is inadequate reporting of confidence intervals and the clinically useful standard error of measurement and minimum detectable change. Various measures of validity are reported but none of the scales have been comprehensively assessed with all the recommended methods. Responsiveness is inadequately reported and when data are available it is limited to effect sizes and standard response means; only two studies[58,59] made reasonable attempts to do this, but neither presented the clinically useful statistic of minimal clinically important differences. All four scales are capable of detecting knee abusers when combined with the recommended partner knee rating scales, only the CSAS and IKDC are capable of detecting non-knee related changes.

Of the 31 scales identified only the Tegner activity rating scale has received adequate investigation of psychometric properties. The results should be interpreted with the benefit of its associated knee rating scale, the Lysholm scale, and investigators should be mindful of the impact of non-knee related causes of score changes. Despite rigorous development processes there is currently insufficient psychometric evidence for the CSAS or Marx scales. Whilst all scales appear to fulfil the clinical usefulness and practicality requirements for administration burden, this is not reported in the studies. There are concerns over the readability and interpretability of technical terms in the CSAS and cultural applicability of some specified sports in the Tegner.

Each of the identified tools measures one or more components of participation from three primary themes, type of activity, frequency, and intensity. Each will be described and the clinical implications discussed.

Type of Activity

The description of activity type falls in one of three categories, return to sports, ranked sports or knee stress. Return to sports is assessed in one of two distinct ways. First, a retrospective comparison with pre-injury levels, using either subjective descriptors ranked on three or four point scales,[9,34–37,39,40,50] a VAS scale[49,50] or a simple yes/no answer[38,40,46] which is highly dependent on the ability to accurately recall previous activity. Second, the time taken to return to sport or activity,[41,43–45,48] which is reliant upon accurately defining what constitutes a 'returned' state, something which is not easily standardized, particularly amongst recreational athletes.

Activities and sports have been listed either with a pre-determined rank[7,18,53] or for the subject to provide a numerical quantification of participation frequency or symptom provocation.[19,51] The primary concern with this type of scale is the criteria selected to perform the ranking of activities, where agreement is often lacking.[31] The importance of condition specificity should also be considered in the ranking since there are sports where demands on the knee are extremely high, but ACL injury may have a small effect (e.g. road cycling). Another concern is how a subject responds when their sport is not listed,[31] since there is no guidance on how the ranks are assigned.

Knee stress is often used as an alternative to ranking of named sports[10,17,18,20,52,54] on the basis that it represents the required level of dynamic knee stability and risk of injury. This eliminates the need to list sports as individuals can assess the motions they undergo in any activity and rank them.[17] Activities involving pivoting, cutting, and rapid deceleration are considered to be ranked above those which do not. It is important to consider the manner in which this is communicated to subjects as technical descriptors such as cutting and pivoting are potentially not familiar terminology to most subjects and may require explanation.[30] When condition specific scales use these motions in isolation, a floor effect is likely, since all activities that do not include these will be ranked at the bottom of the scale.

Intensity

Intensity has been measured in three distinct ways. Firstly through competitive level[18,49,52] with the assumption that competitive sports are more intense than those played recreationally, although the interpretation of 'competitive' is likely to be ambiguous. Secondly by using a subjective descriptor such as strenuous, stressful, moderate or light.[53,54] Thirdly, a reflection on pre-injury level,[49] which has a much stronger reference point and will serve well to define return to sport for an individual, but will not allow for comparison between individuals.

Frequency

Frequency is most often measured in units of times per week or month,[10,17,52] only one scale measures hours per week.[42] Whilst frequency is undoubtedly an important factor in determining participation, there are limitations in its application. Scales highly dependent on frequency may assess someone who participates in low risk activity on a frequent basis at the same level or above someone who participates in high risk activity infrequently. In the case of ACL deficiency a successful adaptor could be ranked at a higher activity level than their pre-injury status despite significant reduction in participation.

Whilst type, frequency, and intensity are undoubtedly important components of participation assessment, it is not yet clear which combination or method of assessment provides the analysis with most significance to clinical decisions or functional outcomes. The identified scales measure each component with different emphasis; the IKDC is heavily based on activity type, the Tegner combines type and intensity and both the Marx and CSAS are dependent on frequency. Further investigation to determine which of these components most significantly impacts on patient satisfaction and functional outcomes would enable tools to be selected on the basis of clinical utility as well as measurement properties.

Conclusion

Activity participation is important for selecting interventions for ACL injured subjects and establishing outcome for ACL reconstruction, and therefore requires appropriately validated measurement tools. This systematic review identified many rating scales which have three common measurement components (type, intensity, and frequency) and considerable variation in the methods for describing them. The choice of components and methods will have implications for the outcome, and clinical application. Currently there is no method which is clearly superior for application in the clinical setting. Unfortunately, only the Tegner activity rating scale has received appropriate investigation of the recommended psychometric properties. There is a need for further psychometric investigation of the Marx, IKDC, and CSAS scales, which should include a comprehensive analysis of clinical usefulness, reliability, validity, and responsiveness. Future studies should include calculations of MCID and MDC to be useful to clinicians and patients. A comparative analysis of measurement properties of these four scales, and the relationship of each to patient satisfaction and functional outcome is required if recommendations for a single standard of reporting is to be achieved.

References

*Indicates the most important references.

1 WHO ICF Checklist, World Health Organisation September 2003.
2 Behairy NH, Dorgham MA, Khaled SA. Accuracy of routine magnetic resonance imaging in meniscal and ligamentous injuries of the knee: comparison with arthroscopy. Int Orthop 2009;33(4):961–7.
3 Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, *et al*. Development and validation of the international knee documentation committee subjective knee form. Am J Sports Med 2001;29(5):600–13.
4 Reid A, Birmingham TB, Stratford PW, Alcock GK, Griffin JR. Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. Phys Ther 2007;87(3):337–49.
5 Eitzen I, Moksnes H, Snyder-Mackler L, Engebretsen L, Risberg MA. Functional tests should be accentuated more in decision for anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthrosc 2010;18:1517–25.
6 Swirton LR, Renstrom P. Factors effecting outcome after anterior cruciate ligament injury: a prospective study with a six year follow up. Scand J Med Sci Sports 2008;18:318–24.
7 Tegner Y, Lysholm J, Odensten M, Gillquist J. Evaluation of cruciate ligament injuries. Acta Orthop Scand 1988;59(3):336–41.
8 Kvist J. Rehabilitation following anterior cruciate ligament injury: current recommendations for sports participation. Sports Med 2004;34(4):269–80.
9 Kocher MS, Steadman JR, Briggs K, Zurakowski D, Sterett WI, Hawkins RJ. Determinants of patient satisfaction with outcome after anterior cruciate ligament reconstruction. J Bone Joint Surg Am 2002;84:1560–72.
10* Noyes FR. A rationale for measuring activity level in anterior cruciate ligament injury. Clin Orth Rel Res 1989;246:238–49.
11* Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. Clin Orth Rel Res 1985;198:43–9.
12* Paxton EW, Fithian DC, Stome ML, Silva P. The reliability and validity of knee specific and general health instruments in assessing acute patella dislocation outcomes. Am J Sports Med 2003;31(4):487–92.
13 Ardern CL, Webster KE, Taylor NF, Feller JA. Return to pre injury level of competitive sport after anterior cruciate ligament reconstruction surgery: two thirds of patients have not returned by 12 months after surgery. Am J Sports Med 2011;39(3):538–43.
14 Heijne A, Axelsson K, Werner S, Biguet G. Rehabilitation and recovery after anterior cruciate ligament reconstruction: patients' experiences. Scand J Med Sci Sports 2008;18:325–35.
15 Holm I, Oiestad BE, Risberg MA, Aune AK. No difference in knee function or prevalence of Osteoarthritis after reconstruction of the anterior cruciate ligament with 4 strand hamstring autograft versus patella tendon bone autograft: a randomised study with 10 year follow up. Am J Sports Med 2010;38(3):448–54.
16 Gobbi A, Francisco R. Factors effecting return to sports after anterior cruciate ligament reconstruction with patella tendon and hamstring graft: a prospective clinical investigation. Knee Surg Sports Traumatol Arthrosc 2006;14:1021–8.
17* Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. Am J Sports Med 2001;29(2):213–8.
18 Straub T, Hunter RE. Acute anterior cruciate ligament repair. Clin Orth Rel Res 1988;227:238–50.

19 Seto JL, Orifino AS, Morrissey MC, Medeiros JM, Mason WJ. Assessment of quadriceps/hamstring strength, knee ligament stability, functional and sports activity levels five years after anterior cruciate ligament reconstruction. Am J Sports Med 1988;**16**(2):170–80.

20 Daniel DM, Stone ML, Dobson BE, Fithian DC, Rossman DJ, Kaufman KR. Fate of the anterior cruciate ligament injured patient. Am J Sports Med 1994;**22**(5):632–44.

21 Labs K, Paul B. To compare and contrast the various evaluation scoring systems after anterior cruciate ligament reconstruction. Arch Orthop Trauma Surg 1997;**116**:92–6.

22* Bent NP, Wright CC, Rushton AB, Batt ME. Selecting outcome measures in sports medicine: a guide for practitioners using the example of anterior cruciate ligament rehabilitation. Br J Sports Med 2009;**43**(13):1006–12.

23* MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, *et al.* Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009;**39**(5):400–17.

24 Anderson AF, Federspeil CF, Snyder RB. Evaluation of knee ligament rating systems. Am J Knee Surg 1993;**6**(2):67–73.

25 Sgaglione NA, Del Pizzo W, Fox JM, Freidman MJ. Critical analysis of knee ligament rating systems. Am J Sports Med 1995;**23**(6):660–7.

26 Weitzel PP, Richmond JC. Critical evaluation of different scoring systems of the knee. Sports Med Arthrosc Rev 2002;**10**:183–90.

27 Wright RW. Knee sports injury outcome measures. J Knee Surg 2005;**18**(1):69–72.

28 Brinker MR, Garcia R, Barrack RL, Timon S, Guinn S, Fong B. An analysis of knee evaluation instruments. Am J Knee Surg 1999;**12**(1):15–24.

29 Lysholm J, Tegner Y. Knee injury rating scales. Acta Orthop 2007;**78**(4):445–53.

30 Marx RG. Knee rating scales. Arthroscopy 2003;**19**(10):1103–8.

31* Barber-Westin SD, Noyes FR. Assessment of sports participation levels following knee injuries. Sports Med 1999;**28**(1):1–10.

32 Roy SB, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum 2009;**61**(5):623–32.

33 Forhan M, Vrkljan B, MacDermid J. A systematic review of psychometric evidence to support the use of an obesity specific quality of life measure for use with persons who have class III obesity. Obes Rev 2009;**11**:222–8.

34 Marshall JL, Fetto JF, Botero PM. Knee ligament injuries: a standardized evaluation method. Clin Orthop 1977;**123**:115.

35 Faegin JA, Blake WP. Postoperative evaluation and result recording in the anterior cruciate ligament reconstructed knee. Clin Orthop 1983;**172**:143.

36 Chick RR, Jackson DW. Tears of the anterior cruciate ligament in young adults. J Bone Joint Surg Am 1978;**60**:970.

37 Hughston JC, Barrett GR. Acute anteromedial rotatory instability: long term results of surgical repair. J Bone Joint Surg Am 1983;**65**:145.

38 Colombet P, Allard M, Bousquet V, de Lavigne C, Flurin PH, Lachaud C. Anterior cruciate ligament reconstruction using four-strand semitendinosis and gracillis tendon grafts and metal interference screws. Arthroscopy 2002;**18**(3):232–7.

39 Mykleburst G, Holm I, Maehlum S, Engerbretsen L, Bahr R. Clinical, functional and radiological outcome in team handball players 6 to 11 years after cruciate ligament injury. Am J sports med 2003;**31**(6):981–9.

40 Zarins B, Rowe C. Combined anterior cruciate ligament reconstruction using semitendinosus tendon and illiotibial tract. J Bone Joint Surg Am 1986;**68**:360.

41 Shelbourne KD, Urchl SE. Primary anterior cruciate ligament reconstruction using the contralateral autogenous patella tendon. Am J Sports Med 2000;**28**(5):651–8.

42 Hahn T, Foldspang A, Ingemann-Hansen T. Prevalence of knee instability in relation to sports activity. Scand J Med Sci Sports 2001;**11**(4):233–8.

43 Jari S, Shelbourne KD. Simultaneous bilateral anterior cruciate ligament reconstruction. Am J Sports Med 2002;**30**(6):891–6.

44 Fujimoto E, Sumen Y, Urabe Y, Deie M, Murkami Y, Adachi N, *et al.* An early return to vigorous activity may destabilize anterior cruciate ligaments reconstructed with hamstring grafts. Arch Phys Med Rehabil 2004;**85**:298–302.

45 Carey JL, Huffman GR, Parekh SG, Sennett BJ. Outcomes of anterior cruciate ligament injuries to running backs and wide receivers in the national football league. Am J Sports Med 2006;**34**(12):1911–7.

46 Bushfield BT, Kharrazi FD, Starkey C, Lombardo SJ, Seegmiller J. Performance outcomes of anterior cruciate ligament reconstruction in the national basketball association. Arthroscopy 2009;**25**(8):825–30.

47 Shah VM, Andrews JR, Fleisig GS, McMichael CS, Lemark LJ. Return to play after anterior cruciate ligament reconstruction in national football league athletes. Am J Sports Med 2010;**38**(11):2233–9.

48 Glasgow SG, Gabriel JP, Sapega AA, Glasgow MT, Torg JS. The effect of early versus late return to vigorous activities on the outcome of anterior cruciate ligament reconstruction. Am J Sports Med 1993;**21**(2):243–8.

49 Tripp DA, Stanish W, Ebel-Lam A, Brewer BW, Birchard J. Fear of reinjury, negative affect and catastrophizing predicting return to sport in recreational athletes with anterior cruciate ligament injuries at 1 year post surgery. Rehabil Psychol 2007;**52**(1):74–81.

50 Rankin AE, Kramer JF, Fowler PJ, Kirkley A. Survey of knee brace usage following anterior cruciate ligament reconstruction. Physiother Can 2000;**52**(3):215–24.

51 Giove TP, Miller SJ, Kent BE, Sanford TL, Garrick JG. Non-operative treatment of the torn anterior cruciate ligament. J Bone Joint Surg Am 1983;**65**:184–92.

52 Keays SL, Newcombe PA, Bullock-Saxton JE, Bullock MI, Keays AC. Factors involved in the development of osteoarthritis after anterior cruciate ligament surgery. Am J Sports Med 2010;**38**(3):455–63.

53 Hanley ST, Warren RF. Arthroscopic meniscectomy in the anterior cruciate ligament deficient knee. Arthroscopy 1987;**3**:59–65.

54 Hefti F, Muller W, Jacob RP, Staubli HU. Evaluation of knee ligament injuries with the IKDC form. Knee Surg Sports Traumatol Arthrosc 1993;**1**:226–34.

55 Jarvinen M, Natri A, Lehto M, Kannus P. Reconstruction of chronic anterior cruciate ligament insufficiency in athletes using a bone patella tendon bone autograft. Int Orthop 1995;**19**(1):1–6.

56 Jarvela T, Kannus P, Jarvinen M. Anterior cruciate ligament reconstruction in patients with or without accompanying injuries: a re-examination of subjects 5 to 9 years after reconstruction. Arthroscopy 2001;**17**(8):818–25.

57 Dahm DI, Wulf CA, Dajani KA, Dobbs RE, Levy BA, Stuart MA. Reconstruction of the anterior cruciate ligament in patients over 50 years. J Bone Joint Surg Br 2008;**90**(11):1446–50.

58* Briggs KK, Kocher MS, Rodkey WG, Steadman JR. Reliability, validity and responsiveness of the Lysholm knee score and Tegner activity scale for patients with meniscal injury of the knee. J Bone Joint Surg Am 2006;**88**:698–705.

59* Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The reliability, validity and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee. Am J Sports Med 2009;**37**(5):890–7.

60* Briggs KK, Steadman JR, Hay CJ, Hines S. Lysholm score and Tegner activity level in individuals with normal knees. Am J Sports Med 2009;**37**(5):898–901.

61* Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous reliability, validity and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured or anterior cruciate ligament reconstructed knees. Am J Sports Med 1999;**27**(4):402–16.

62 Rodkey WG, DeHaven KE, Montgomery WH, Baker CL, Beck CL, Hormal SE, *et al.* Comparison of the collagen meniscus implant with partial meniscectomy. J Bone Joint Surg 2008;**90**:1423–6.

63 Spindler KP. The multicentre ACL revision study (MARS). J Knee Surg 2007;**20**(4):303–7.

64 Neeb TB, Aufdemkampe G, Wagener HD, Mastenbroek L. Assessing anterior cruciate ligament injuries: the association and differential value of questionnaires, clinical tests and functional tests. JOSPT 1997;**26**(6):324–31.

65 Strik G, Aufdemkampe G, Neeb TB, Mastenbrook ML. Betrouwbaarheid en content-validiteit van de Nederlandstalige vragenlijsten bij knieklachten op basis van een voorste kruisband laesie. Ned T Fysiotherapie. 1998;**108**(1):15–20.

66 Vlasveld ES, Huis van R, Klooster van't GR, Aufdemkampe G, Moolenaar CL. De waarde van functionele tests in de beoordeling van de kniefunctie. Ned T Fysiotherapie 1993;**5**:180–9.