Assessing participation in the ACL injured population: Selecting a patient reported outcome measure on the basis of measurement properties

CrossMark

Robert Letchford [a,c,*], Valerie Sparkes [a,b], Robert W.M. van Deursen [a,b]

[a] School of Healthcare Sciences, Cardiff University, Second Floor, Cardigan House, Heath Park Campus, Cardiff CF14 4XN, United Kingdom
[b] Arthritis Research UK Biomechanics and Bioengineering Centre, Cardiff University, Cardiff, United Kingdom
[c] Aneurin Bevan Health Board, Physiotherapy Department, Royal Gwent Hospital, Cardiff Road, Newport NP20 2UB, Gwent, United Kingdom

ARTICLE INFO

ABSTRACT

Background/aim: A return to pre injury activity participation remains a common but often elusive goal following ACL injury. Investigations to improve our understanding of participation restrictions are limited by inconsistent use of insufficiently investigated measurement tools. The aim of this study was to follow the consensus based standards for the selection of health measurement instruments (COSMIN) guideline to provide a comparative evaluation of four patient reported outcomes (PROMs) on the basis of measurement properties. This will inform recommendations for measuring participation of ACL injured subjects, particularly in the United Kingdom (UK) National Health Service (NHS).
Methods: Thirteen criteria were compiled from the COSMIN guideline. These included reliability, measurement error, content validity, construct validity, responsiveness and interpretability. Data from 51 subjects collected as part of a longitudinal observational study of recovery over the first year following ACLR was used in the analysis.
Results: Of the thirteen criteria, the required standard was met in 11 for Tegner, 11 for International Knee Documentation Committee (IKDC), 6 for Cincinnati Sports Activity Scale (CSAS) and 6 for Marx. The two weaknesses identified for the Tegner are more easily compensated for during interpretation than those in the IKDC; for this reason the Tegner is the recommended PROM.
Conclusions: The Tegner activity rating scale performed consistently well in respect of all measurement properties in this sample, with clear benefits over the other PROMs. The measurement properties presented should be used to inform implementation and interpretation of this outcome measure in clinical practice and research.
Level of evidence: Level II prospective study.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Short term success of interventions for the anterior cruciate ligament (ACL) injured knee has been defined by a symptom free return to participation in the individual's chosen activities [1–3]. However, recent publications of rehabilitative [1] and surgical [2,3] interventions, have demonstrated lower rates of success than has been previously expected [4]. Whilst a multifactorial interaction between physical, physiological, psychological and social factors has been proposed [4], investigations of these are limited by inconsistencies in the measurement of participation outcomes [2,5].

No gold standard measure for participation outcomes exists, however patient reported outcome measures (PROMs) have become widely accepted in the literature. A recent systematic review demonstrated that PROMs are inconsistently adopted and that the four most commonly reported (Tegner, Cincinnati sports activity scale (CSAS), Marx and International knee documentation committee (IKDC)) lack a comprehensive exploration of measurement properties [5]. There has been considerable debate regarding terminology and methodology for the assessment of measurement properties of PROMs [6,7]. The COSMIN group (Consensus based standards for the selection of health measurement Instruments) have published an international consensus guideline that goes a significant way to resolving this debate, and offers a framework for the conduct and reporting of such studies [8].

This study therefore aimed to; following the COSMIN guideline, provide a comparative evaluation of the measurement properties of these four PROMs. This will inform recommendations for participation PROMs for ACL injured subjects, particularly in the United Kingdom (UK) National Health Service (NHS).

2. Materials and methods

2.1. The four patient reported outcome measures (PROMs)

The *Tegner activity rating scale* [9] is a single item with 11 responses ranked by activity type and intensity on an ordinal scale between 0 and

* Corresponding author at: Floor 2 Cardigan House, School of Healthcare Sciences, Cardiff University, Heath Park, Cardiff CF14 4XN, United Kingdom. Tel.: +44 7973661467.
E-mail address: letchfordrh@cardiff.ac.uk (R. Letchford).

10. Investigations for measurement properties [9–13] have reported adequate evidence for test–retest reliability, measurement error and known groups validity [5].

The *Cincinnati Sports Activity Scale* (CSAS) [14,15] is a single item with 12 responses ranked by activity type and frequency on an ordinal scale between 0 and 10. There is adequate evidence only for reliability [5].

The *Marx activity rating scale* [16] has four items each with four responses. These are ranked by frequency on an ordinal scale between 0 and 4 and summed to a maximum score of 16. There is adequate evidence for reliability and convergent/divergent validity [5].

The *International Knee Documentation Committee (IKDC)* includes participation on both the knee evaluation [17] and subjective knee form [18]. It is a single item with five responses ranked between 1 and 5 by the intensity and type of activity. No studies of measurement properties of the activity rating section were identified [5].

2.2. Data collection

Data were collected as part of a prospective longitudinal observational study investigating recovery following primary hamstring autograft ACL reconstruction. All patients attending our unit for the above procedure between January 2011 and July 2013 were invited to participate. Subjects were excluded only if additional surgical procedures that altered the standard rehabilitation programme were performed (e.g. microfracture). Data were collected prior to surgery and 1, 2, 3, 6 and 12 months following surgery. Pre operative data were collected on average 25 (standard deviation, $SD = 34$) days before surgery; this was on average 19 ($SD = 17$) months following injury. Retrospective measures of pre injury participation were also collected at the pre operative visit, however for 22 subjects they were delayed due to later inclusion of the IKDC in the study. All four participation PROMs were provided in the original published format [9,14,16,18]. Additional data were collected simultaneously at each visit for use in the validity and interpretation analysis; measures of knee function including the Lyshlom knee scale, IKDC Subjective knee form (IKDC SKF) and Visual Analogue Scale (VAS) for pain and a global rating of change score (GRCS) for participation. Ethical approval was received from the South Wales Research Ethics Committee (Reg: 10/WSE04/48).

Where missing data occurred, the Missing Completely at Random (MCAR) assumption was assessed using Little's MCAR test [21] and differences in baseline participation and demographics between the subjects with and without missing data were explored [22]. Listwise deletion was used when the MCAR assumption was supported [22,23].

2.3. Measurement properties

The definitions and methodological guidelines established by the COSMIN group [8,19] were fully adopted. These are summarised below.

2.3.1. Reliability and measurement error

Test–retest reliability was calculated from repeated measures from a convenience sample of 35 subjects (on the basis of the COSMIN recommendations) [19] who completed each PROM at consultation, then repeated them two and four days later [20] and returned them in a sealed envelope. Intraclass correlation coefficients (ICC) were considered acceptable when values were >0.8 for group and >0.9 for individual analysis [19,24,25]. Measurement error was calculated from this repeated measures data and considered acceptable when smallest detectable change (SDC) for both individual and group analysis [19,26,27] was lower than one category of change. Due to differences in the PROM's scoring structures this was one point change on the Tegner, Marx and IKDC, but 5 points on the CSAS.

2.3.2. Validity and responsiveness

Relevance of items and responses to the construct, population and purpose of the instrument (content validity) were assessed by cross matching the items on each PROM to the participation domain of the World Health Organisation International Classification of Functioning, Disability and Health (WHO ICF) [28], published national sports participation data [29] and ACL injury risk data [30]. The PROM with the greater number of ICF domains and qualifiers, greatest number of high risk and high frequency participation sports was considered to best represent the diverse population of ACL injury and was therefore preferred. Questionnaire development (Item generation and reduction) was previously assessed by our systematic review [5] and was therefore not further investigated in this study. Only the Marx scale was considered adequate on this criterion [5].

Since there is no gold standard measure for participation, construct rather than criterion methods were applied to assess validity and responsiveness [8]. Hypothesis testing of relationships between known groups (healthy, ACL injured and ACL reconstructed), convergent (the four participation PROMs) and divergent (knee function and pain) constructs was used to define validity. Direction and magnitude of the change in scores over time, as subjects passed from healthy to injured and reconstructed were used to define responsiveness. The following hypotheses were generated.

For convergent validity: Since all four PROMs measure the same construct Hypothesis one states that they would correlate highly ($r > 0.7$). For divergent validity: Whilst function is considered a primary limiter of participation [2,3,31,32] the 'knee abuser' is known to continue to participate despite impairments [14]. Hypothesis two therefore stated there would be a moderate (0.4–0.7) correlation between function and participation. Participation is known to reduce with age [3,33] and Hypothesis three stated a moderate (0.4–0.7) inverse correlation with age. In an active population there is no reason to expect BMI to influence participation and Hypothesis four stated a low correlation ($r < 0.4$) with BMI. Correlations were calculated using Spearman's r and interpreted using Dancey and Reidy (2004) categorisation [34], coefficients between 0.7 and 0.9 are considered strong, 0.4 to 0.6 are considered moderate and 0.1 to 0.3 are considered weak.

For known groups validity and responsiveness: Participation is known to be restricted following ACL injury and is expected to improve but not resolve 12 months after surgery [2,3]. Since most rehabilitation schedules consider return to pre injury activities after six months [40] change in participation between six and 12 months is expected to be greater than between injury and six months. Therefore Hypothesis five for validity stated that pre operative scores will be lowest, becoming sequentially greater at six months, 12 months and highest before injury. Hypothesis six for responsiveness stated that change scores between pre-injury and pre-operative will be negative and larger than the positive changes occurring between both pre-operative and six months or six and 12 months. These differences were tested using Freidman's ANOVA with post hoc Wilcoxon signed rank test and Bonferroni correction. Effect size (r) was calculated from the t statistic and interpreted following the categorisation of Cohen (1998); >0.5 is considered large, 0.3–0.5 moderate and <0.3 small. Adequate known group's validity and responsiveness were achieved if hypotheses were confirmed with appropriate effect size.

2.3.3. Interpretability

Floor and ceiling effects were considered significant when >15% of scores were located within the lowest or highest category of the scale, absence of significant effects is preferred [36]. Minimally important change (MIC) is the subject of considerable debate, with both distribution and anchor based methods described [7,37]. The visual anchor based MIC distribution provides an appropriate compromise, fulfilling the requirements of both methods [7]. The minimal change and much change levels of the global rating of change score (GRCS) were used to define important change. Receiver Operator Characteristic (ROC) curve was used to define the optimal cut off point, for both directions of change at 'minimally' and 'much' changed levels of the GRCS [38]. The area under the curve (AUC) is used as a summary statistic [19,39]