ARTICLE IN PRESS

Research paper

Comparison of psychometric properties of subjective structured assessment instruments of technical performance during knee arthroscopy

Karthik Vishwanathan*, Amit Patel, Ramesh Panchal

Pramukhswami Medical College, Department of Orthopaedics, Shri Krishna hospital and Pramukhswami Medical College, Charutar Arogya Mandal, Gokalnagar, Karamsad, Gujarat, 388325, India

A B S T R A C T

Objectives: Various subjective assessment global rating scales to assess proficiency of surgeons while performing knee arthroscopy have been described but it is still unclear as to which of the assessment scale is the most optimal and "gold standard". The aim of the present study was to compare and contrast psychometric properties like validity and reliability of objective assessment global rating scales for knee arthroscopy.
Methods: A systematic review was performed of articles published in Pubmed, Embase, AMED, ERIC, Proquest and CINAHL. Ten assessment scales (Arthroscopic surgery skill evaluation tool [ASSET]; Arthroscopy skills score; global ratings of arthroscopic performance; basic arthroscopic knee scoring system [BAKSS]; modified basic arthroscopic knee scoring system [modified BAKSS]; modified objective structured assessment of technical skill [modified OSATS]; modified objective assessment of arthroscopic skills [modified OAAS]); modified orthopaedic competency assessment project [modified OCAP; Imperial global arthroscopy rating scale [IGARS] and Injury grading index [IGI] fulfilled the inclusion criteria. A predefined checklist was used to compare content validity, construct - convergent validity, construct - discriminant validity, criterion validity, predictive validity, internal consistency, inter-rater reliability, intra-rater reliability and test-retest reliability of all the included assessment global rating scales.
Results: The ASSET demonstrated optimal number of satisfactory psychometric properties of content validity, construct-convergent validity, construct - discriminant validity, criterion validity, internal consistency, inter-rater reliability, intra-rater reliability and test-retest reliability. The IGI demonstrated the least number of satisfactory psychometric properties.
Conclusion: We recommend the use of Arthroscopic surgery skill evaluation tool [ASSET] as it demonstrates optimal psychometric properties. The ASSET can be used as gold standard to compare existing outcome assessment tools.

© 2017 Published by Elsevier, a division of RELX India, Pvt. Ltd on behalf of International Society for Knowledge for Surgeons on Arthroscopy and Arthroplasty.

1. Introduction

Knee arthroscopy is a specialized surgical procedure which entails a learning curve. Various methods have been devised to enhance training of the trainee surgeons like human cadaver knees, bench top simulator models and virtual reality knee simulator models with haptic force feedback capabilities.

Assessment of competency to perform knee arthroscopy can be done using subjective methods and objective methods. Subjective methods include using a global rating scale with predetermined criteria and task specific checklist having predetermined criteria. Objective methods include 3D Motion analysis metrics or 3D Motion analysis parameters and time taken to complete the surgical procedure or particular task. Evaluation using 3D Motion analysis metrics or 3D Motion analysis parameters needs external camera and other additional devices for performing the motion analysis and these devices are costly. Subjective methods of assessment are cheaper, practical and feasible to administer in the

* Corresponding author at: Department of Orthopaedics, Shri Krishna hospital and Pramukhswami Medical College, Charutar Arogya Mandal, Gokalnagar, Karamsad, Gujarat, 388325, India.
  E-mail addresses: karthik_vishwanathan@yahoo.com, karthikvg@charutarhealth.org (K. Vishwanathan), dramit2000@hotmail.com (A. Patel), dr_ramesh_panchal@yahoo.com (R. Panchal).

clinical and laboratory setting. Though subjective assessments entail risk of bias, efforts have been made to reduce the element of subjectivity by introducing a global rating scale (GRS) wherein various domains and attributes are marked on a predetermined rating assessment sheet with the aim to make the subjective assessment more objective and reduce the element of subjectivity. Of the two commonly used subjective assessment methods, the GRS is better than the task specific checklist because the GRS is able to discriminate between trainees of varying expertise level whereas the checklist is unable to discriminate between them[1].

Validity and reliability are crucial psychometric properties for any assessment tool[2].

There have been systematic reviews evaluating psychometric properties of competency assessment tool in microsurgery[3], laparoscopic surgical skills[4], laparoscopic cholecystectomy[5] and vascular surgical skills[6]. A systematic review had identified deficiency of validated outcome instruments to assess competency of arthroscopic surgery skills[7]. This was probably due to lack of having a standardized check list for evaluation of various psychometric properties like validity and reliability. Moreover, lot of literature has been available recently pertaining to assessment of competency of knee arthroscopy. Hence, the objective of the present study was to find out the objective global rating scale assessment of performance of knee arthroscopy having the most optimal psychometric properties.

2. Methods

2.1. Institutional HREC approval

The study proposal was reviewed by the hospital research ethics committee and granted exemption as it was a systematic review of published literature that was already available in the public domain.

2.2. Literature search

Literature search was performed of electronic databases comprising Pubmed, Ovid (Embase/AMED [Allied and complementary medicine]/ERIC [Education resources information centre]), Proquest and CINAHL using the following keywords: "knee arthroscopy competency"; "knee arthroscopy competence"; "knee arthroscopy skills"; "knee arthroscopy global rating scale"; "knee arthroscopy proficiency" and "objective structured assessment of technical skill". The electronic search consisted of screening of all articles published from 1990 to December 2016. The last date for performing the literature search was 20th of December 2016.

2.3. Inclusion and exclusion criteria

The studies that used a structured assessment tool (global rating scale) to evaluate performance of technical skills of novice and/or expert orthopaedic surgeons during knee arthroscopy and described validity or reliability or both of the assessment tools were included in the systematic review. Non-English language studies, studies that did not report validity or reliability of the global rating scale assessment method or used irrelevant statistical method, studies focussing on non-technical skills like communication skills, studies dealing with proficiency assessment of arthroscopy of joints other than knee joint, conference proceedings and literature reviews or expert opinions were excluded from the systematic review.

Table 1
Checklist to evaluate psychometric properties of GRS in individual studies.

Psychometric property	Positive rating (+) Adequate	Doubtful rating (?)	Negative rating (−) Inadequate	Nil rating (0)
Content validity	Explicit mentioning of the process of choosing domains relevant to knee arthroscopy surgical skils [Literature review/textbook review/expert knee arthroscopy surgeons/sports medicine fellowship trained surgeons opinion obtained using either Delphi method or focus group discussion] [9]	No involvement of experts in the development process [13]	Experts consider items in the final version of GRS to be irrelevant OR consider the GRS to be incomplete	No information on content validity
	Experts consider all items in the GRS to be relevant and consider the final version of the GRS to be complete	Doubtful method. Modification of pre-existing GRS by addition/deletion		Process of selection of various domains in assessment of knee arthroscopy not mentioned
Construct − convergent validity	Correlation with related constructs/instruments measuring the same construct $>/= 0.50$ [11–13]	Correlations determined with unrelated construct	Correlation with instruments measuring the same construct/related construct < 0.50 [11–13]	No information on construct − convergent validity
	Statistical significant ($p </= 0.05$) result and there was prespecified hypothesis	Statistical significant result but no pre-specified hypothesis	Statistically insignificant result	
Construct - discriminant validity	Able to discriminate various groups. Able to show significant difference between groups	Doubtful design or method	Unable to discriminate various groups statistically despite adequate design and method.	No information on construct − discriminant validity
Criterion validity	Correlation coefficient $>/= 0.70$ [11]	No convincing argument that gold standard is indeed gold std/Doubtful design and method	Correlation with "gold standard" < 0.70 despite adequate design and method [11]	No information on criterion validity
Internal consistency	Cronbach's alpha between 0.70 and 0.95 [11]	Doubtful design or method	Cronbach's alpha < 0.70 or > 0.95 despite adequate design and method [11]	No information on internal consistency
Reliability (Inter-rater reliability/Intra-rater reliability/Test retest reliability)	ICC or Weighted kappa $>/= 0.70$ [11] OR Pearson's correlation coefficient $>/= 0.80$ [12,13]	Doubtful design or method (e.g time interval between readings not mentioned)	ICC or Weighted kappa < 0.70 [11] OR Pearson's correlation coefficient < 0.80 despite adequate design and method [12,13]	No information on reliability