Manual Therapy 20 (2015) 402–411

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Systematic review

The reliability of physical examination tests for the diagnosis of anterior cruciate ligament rupture – A systematic review

Toni Lange [a, 1], Alice Freiberg [b, 1], Patrik Dröge [b], Jörg Lützner [c], Jochen Schmitt [a, b], Christian Kopkow [a, *]

[a] University Hospital Carl Gustav Carus Dresden, Center for Evidence-Based Healthcare, Fetscherstr. 74, 01307 Dresden, Germany
[b] Technical University Dresden, Medical Faculty, Institute and Policlinic of Occupational and Social Medicine, Fetscherstr. 74, 01307 Dresden, Germany
[c] University Hospital Carl Gustav Carus Dresden, Centre for Orthopaedics and Trauma Surgery, Fetscherstr. 74, 01307 Dresden, Germany

ARTICLE INFO

Article history:
Received 8 January 2014
Received in revised form
3 November 2014
Accepted 6 November 2014

Keywords:
Anterior cruciate ligament
Lachman test
Reliability
Systematic review

ABSTRACT

Study design: Systematic literature review.
Background: Despite their frequent application in routine care, a systematic review on the reliability of clinical examination tests to evaluate the integrity of the ACL is missing.
Objectives: To summarize and evaluate intra- and interrater reliability research on physical examination tests used for the diagnosis of ACL tears.
Methods: A comprehensive systematic literature search was conducted in MEDLINE, EMBASE and AMED until May 30th 2013. Studies were included if they assessed the intra- and/or interrater reliability of physical examination tests for the integrity of the ACL. Methodological quality was evaluated with the Quality Appraisal of Reliability Studies (QAREL) tool by two independent reviewers.
Results: 110 hits were achieved of which seven articles finally met the inclusion criteria. These studies examined the reliability of four physical examination tests. Intrarater reliability was assessed in three studies and ranged from fair to almost perfect (Cohen's $k = 0.22–1.00$). Interrater reliability was assessed in all included studies and ranged from slight to almost perfect (Cohen's $k = 0.02–0.81$). The Lachman test is the physical tests with the highest intrarater reliability (Cohen's $k = 1.00$), the Lachman test performed in prone position the test with the highest interrater reliability (Cohen's $k = 0.81$). Included studies were partly of low methodological quality. A meta-analysis could not be performed due to the heterogeneity in study populations, reliability measures and methodological quality of included studies.
Conclusion: Systematic investigations on the reliability of physical examination tests to assess the integrity of the ACL are scarce and of varying methodological quality.

1. Background

Rupture of the anterior cruciate ligament (ACL) is a severe knee injury, which is often acquired during sporting activities (Alentorn-Geli et al., 2009; Gianotti et al., 2009), and often results in instability. This subsequently leads to meniscal and cartilage lesions and/or functional impairments (Lohmander et al., 2004; Michalitsis et al., 2013; Sri-Ram et al., 2013). Early diagnosis and treatment is therefore necessary. To diagnose an ACL tear, information obtained from patient history, physical examination, arthrometric testing, imaging techniques, and sometimes arthroscopy are combined. Since the physical examination is essential in this process, several reviews have already assessed the validity of physical examination tests for ACL rupture (Malanga et al., 2003; Scholten et al., 2003; Benjaminse et al., 2006; van Eck et al., 2013). Accordingly, the Lachman test seems to be the test with the highest validity (Benjaminse et al., 2006). However, if physical examination tests are applied, validity and also reliability are of concern (Karanicolas et al., 2009; Scholtes et al., 2011). Assessing the reliability of a test should be a primary concern in diagnostic research, because a poor reliability has a negative influence on the test accuracy, and a test will not be valid if it does not measure consistently (Atkinson and Nevill, 1998). Physical tests with insufficient reliability might be the reason for varying data regarding the validity of physical tests, due to poor test conduction (e.g. trained versus untrained rater,

* Corresponding author. Tel.: +49 (0)351 4586492; fax: +49 (0)351 4587238.
E-mail address: Christian.Kopkow@uniklinikum-dresden.de (C. Kopkow).
[1] The first two authors (A.F. and T.L.) contributed equally to this work.

http://dx.doi.org/10.1016/j.math.2014.11.003
1356-689X/

variation in test execution due to raters). Empirical research has shown that variability among raters influences diagnostic accuracy (Whiting et al., 2004, 2013). Reliability refers to the ability of a measurement to differentiate between subjects or objects, and agreement represents how identical scores or ratings are (Kottner et al., 2011). Reliability can be further divided into (Kottner et al., 2011):

- Intrarater reliability (agreement): the same subjects or objects are assessed through the same rater using the same scale, classification, instrument or procedure at different times.
- Interrater reliability (agreement): the same subjects or objects are being assessed through different raters using the same scale, classification instrument or procedure (e.g. a physical examination test).

For clinicians it might be of interest if their diagnosis made is in agreement with the diagnosis made be colleagues, which refers to interrater reliability (de Vet et al., 2013). This is often the case in daily practice, when patients are referred or being diagnosed by different persons with varying experience or training in test conduction.

Since ACL injury is a very common knee injury (Zeng et al., 2013), and physical examination tests for the assessment of ACL rupture are being applied in routine care very frequently (Jensen, 1990; van Eck et al., 2013), not only the validity should be summarized systematically and assessed critically. In addition, also reliability as another domain of measurement properties (Mokkink et al., 2010), should be in the focus of systematic reviews of physical examination tests.

2. Objective

To our knowledge, there have been no published systematic reviews conducted to determine the intra- and interrater reliability of physical examination tests for ACL injuries. Accordingly, the objective of this systematic review was to summarize and evaluate intra- and interrater reliability research on physical examination tests used for the diagnosis of ACL tears.

3. Methods

3.1. Data sources and search strategy

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines during the whole searching and reporting process of this review (Moher et al., 2009). The PRISMA statement aims to improve the reporting of systematic reviews and meta-analyses.

We executed a comprehensive systematic literature search in the following databases via the Ovid interface until May 30th, 2013: MEDLINE from 1946, EMBASE from 1974, and the Allied and Complementary Medicine Database (AMED) from 1985. The search strategy included terms about clinical tests for the ACL, the integrity of the ACL and the test measurements (Fig. 1). In addition, we screened the reference lists of all eligible articles for further relevant studies (Cooperman et al., 1990; Duggan and Ross, 1991; Fleming et al., 1992; Johnson et al., 2004; Wiertsema et al., 2008; Peeler et al., 2010; Mulligan et al., 2011).

3.2. Inclusion and exclusion criteria

Studies assessing the intra- and/or interrater reliability of a physical examination test were included if written in English or German. Since there is no consensus which reliability measures are preferable and are of most concern for clinical practice (Sim and Wright, 2005; de Vet et al., 2013), studies were not excluded based on the used reliability measures. Studies on patients of every age and setting were considered eligible. The investigated ACL ruptures could have been acute or chronic as well as partial or complete. The criteria therefore were adopted by the original definition of the authors of each study. Ruptures could furthermore be isolated or in combination with other knee injuries.

We excluded studies if the physical examination test was performed under anesthesia. Animal studies and cadaveric studies were excluded, as well as studies that used device supported examinations or clinical test series but also if the authors made generic terms such as “physical examination” so that one could not identify a specific test. If a study did not provide the name or the description of the physical test or did not reference a source for further description of the test, studies were considered ineligible. If the physical examination test was named, but no description on how the test was performed was given, we assumed that the authors performed the test in the usual manner and included the study.

3.3. Selection of studies and data abstraction

The identified titles and abstracts of studies were screened independently by two reviewers (A.F. and T.L.). Subsequently, full texts were checked independently for eligibility by the two

#	Searches	Results
1	(lachman$ or "active lachman$" or "anterior drawer" or "pivot shift" or "fibular head" or "loe test" or "loss of extension test").mp.	3784
2	(pathолog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instability or laxity).mp.	4138266
3	(knee or anterior cruciate ligament or acl).mp.	257306
4	(diagnos$ or "physical examination" or examin$ or sign$ or test$ or manual$ or man?euv$).mp.	19167589
5	"reproducibility of results"/ or (reliabilit$ or reproduca$).mp. or ((inter or intra) and (rater or examin$ or tester or observer)).mp. or (inter-rater or inter-examin$ or inter-tester or inter-observer or intra-rater or intra-examin$ or intra-tester or intra-observer).mp.	670349
6	1 and 2 and 3 and 4 and 5	175
7	remove duplicates from 6	110

Fig. 1. Search strategy.