DIAGNOSTIC ACCURACY OF THE LEVER SIGN IN DETECTING ANTERIOR CRUCIATE LIGAMENT TEARS: A SYSTEMATIC REVIEW AND META-ANALYSIS

Kristin Abruscato, SPT¹

Kelsie Browning, SPT¹

Daniel Deleandro, SPT¹

Quinn Menard, SPT¹

Mark Wilhelm, PT, DPT¹

Amy Hassen, PT, DPT, OCS, MTC¹

ABSTRACT

Background: The anterior cruciate ligament (ACL) is one of the most commonly injured ligaments in the knee. With the prevalence of ACL tears increasing, there is a growing need for clinical tests to rule in and rule out a suspected tear. A new clinical test for detecting ACL tears has been introduced with preliminary studies showing promising results.

Hypothesis/Purpose: To systematically review and analyze information from the current literature on the diagnostic accuracy of the Lever Sign test for the use of diagnosing anterior cruciate ligament (ACL) injuries in a clinical setting.

Study Design: Systematic review and meta-analysis

Methods: A computerized search of PubMed, Cinahl, Scopus, and Proquest databases as well as a hand-search was completed on all available literature using keywords relating to the diagnostic accuracy of the Lever Sign Test. A quality assessment was performed on each article included in this review utilizing the Quality Assessment of Diagnostic Accuracy Studies (QUADAS).

Results: Eight articles were included, with only three studies exhibiting high quality, however the study samples were heterogenous. Included studies indicated that the Lever Sign test is both sensitive and specific in diagnosing ACL tears. Pooled sensitivity and specificity were 0.77 and 0.90, respectively. The negative likelihood ratio is 0.22 and the positive likelihood ratio is 6.60.

Conclusion: The Lever Sign test is comparable to other clinical tests used in current practice to detect an ACL rupture. The pooled data from current available literature on the Lever Sign indicate that a positive or negative test should result in a moderate shift in post-test probability. This test may be used in addition to other tests to rule in and rule out the presence of an ACL rupture.

Level of Evidence: 2a- Systematic Review of Level 2 diagnostic studies

Key Words: Anterior cruciate ligament, diagnostic accuracy, knee, Lelli test, Lever sign test, movement system

CORRESPONDING AUTHOR

Amy Hassen

Assistant Professor

Walsh University

2020 E. Maple St.

North Canton, OH, 44702

E-mail: ahassen@walsh.edu

¹ Department of Physical Therapy, Walsh University, North Canton, OH, USA

The authors indicate no conflicts of interest in the reporting of this systematic review and meta-analysis.

INTRODUCTION

Anterior cruciate ligament (ACL) tears are a common injury with a high prevalence occurring during athletic competitions.¹ The incidence of ACL tears among athletes has been reported at 68.6 per 100,000 persons per year.² A higher incidence of ACL tears has been noted in females compared to males when participating in the same sport.³ Anterior cruciate ligament injuries in athletes are more commonly due to non-contact mechanisms and are a result of deceleration and/or pivoting motions required in athletics.⁴ The ACL acts to resist posterior translation of the femur on the tibia, thus providing a large amount of stability to the knee joint.⁵ Early diagnosis is necessary to determine the best course of care and reduce the risk of further injury.

Clinical tests with favorable specificity and sensitivity are needed to determine if additional testing or imaging is warranted. Currently there are three diagnostic tests that are commonly used to assess for ACL tears including the Lachman's test, anterior drawer test, and the pivot shift test. Of the tests previously stated, the Lachman test has long been reported to be the most sensitive and specific and should be included in every examination of a suspected ACL tear.⁶⁻⁷ Both the Lachman and pivot shift tests require the clinician to stabilize one segment of the leg while manipulating the other segment. This may prove challenging and produce inaccurate test results if the patient's leg is of a larger girth or weight than the clinician is able to support. Therefore, a test that places the clinician in a biomechanical advantage may prove more practical.

Recently, a new clinical test for ACL was introduced called the Lever Sign test. The test is performed with the patient lying in supine and the examiner's closed fist under the proximal third of calf. With the knee slightly flexed, the examiner's opposite hand then applies a downward force just proximal to the knee joint. In an intact ACL, the heel should rise off the table, indicating a negative test. A disruption in the ACL will result in the patient's heel remaining on the table, indicating a positive test.⁸ It has been found that at 30° of knee flexion, the greatest amount of anterior tibial translation occurs relative to the femur. By placing the fist under the calf, the knee is positioned in approximately 30° of flexion,

placing the ACL origin and insertion at a maximal distance from each other. This orientation will maximally stress the ACL, explaining why an ACL deficient knee will not be able to overcome the posterior force directed at the femur.⁹

With the prevalence of ACL tears increasing, diagnostic tests with good utility are needed in order to assist the clinician in a diagnosis. At the time of submission, there were no previously published systematic reviews analyzing the diagnostic accuracy of the Lever Sign test. The purpose of this systematic review and meta-analysis is to determine the diagnostic accuracy of the Lever Sign test.

METHODS

Protocol and Registration

A systematic review protocol was registered with the International Prospective Register of Systematic Reviews – PROSPERO with registration number: CRD42018082534. This systematic review used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which is a 27-item checklist used to ensure optimal design and standardized reporting of systematic reviews and meta-analyses.¹⁰

Eligibility Criteria

The inclusion criteria set for this review included 1) studies that were completed on human subjects and 2) studies that evaluated the diagnostic accuracy of Lever Sign test for ACL rupture as compared to MRI or arthroscopic surgery, both of which are accepted reference standards. All peer-reviewed full-text articles of any study design were eligible for inclusion. Studies included subjects of any age and acuity of injury. Studies were excluded if the full text was not written in English.

Search Strategies

A comprehensive search was performed on PubMed, CINAHL, Scopus, and ProQuest in December 2017 and updated in April 2018 for articles relating to the diagnostic accuracy of the Lever Sign test. A search strategy was developed for PubMed utilizing variations of the keywords Lever Sign test, ACL rupture, and diagnostic accuracy, which can be found in Appendix 1. The search strategy was then

adapted for use in each specific database utilizing database-specific article indexing. The Walsh University library database was used to conduct a hand search for grey literature and other eligible articles in order to complete the comprehensive systematic search.

Study Selection

Titles and abstracts were retrieved from each database and duplicates were eliminated. Two review authors independently screened titles and abstracts found in the search with respect to the inclusion and exclusion criteria. The two review authors discussed discrepancies until consensus was achieved. Full text copies were then obtained for the remaining articles and were reviewed independently by two other authors for inclusion. Disagreements were discussed between the two full-text review authors and consensus was reached. Inter-rater reliability for title/abstract and full-text inclusion were calculated using Cohen's Kappa.¹¹ The interpretation of the Cohen's unweighted kappa statistics used were: < 0 = poor, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1 = almost perfect.¹¹

Risk of Bias

To identify the risk of bias in each individual article included in the review, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was utilized and consists of 14 items that rate the overall quality of an article using the study's internal and external validity. The review authors gave a rating of "yes" (sufficiently covered) "no" (insufficiently covered) or "unclear" (insufficient detail included for open interpretation) for each item. The QUADAS tool has been reported to be reliable when assessing the strengths and weaknesses of diagnostic accuracy study.^{12,13} Articles that scored above a 10/14 were considered high quality and those below were considered low quality. While there is not an absolute qualification for high and low quality scores when using the QUADAS, these numbers are generally accepted in literature.¹⁴⁻¹⁶ Two review authors independently performed the quality assessment for each article and discrepancies were resolved by consensus. Inter-rater reliability for risk of bias was calculated using Cohen's Kappa.¹¹

Data Extraction

Two review authors extracted data independently from the included articles into a standardized form to ensure consistency of the data extraction process. The data were then substantiated by two review authors to ensure accuracy. Extracted information included: examiner skill level; patient age; patient gender; comparison standard; side of involvement; time since injury; study methodology; outcomes; diagnostic accuracy statistics including sensitivity and specificity. In the event that data components were not reported within the article, authors from included studies were contacted and a formal request was made for additional data. The definitions and calculations for diagnostic accuracy statistics can be found in Table 1.

Statistical Analysis

Pooling of results data for the quantitative synthesis was performed using Open Meta-analyst.¹⁷ Raw data from studies were input into a 2x2 table and Open Meta-analyst generated independent sensitivity and specificity values using the DerSimonian-Laird Random-Effects model. Cochran's Q and the I² statistic were assessed to determine the degree of heterogeneity. If I² was > 50%, a random effects model was used. The weighted average of pooled statistics was calculated and overall diagnostic accuracy values were determined. The diagnostic odds ratio (DOR) was calculated to determine the overall diagnostic power of the test. The I² statistic was used to determine variance across studies and indicate overall heterogeneity.

Outcomes and Summary Measures

Sensitivity represents the amount of true positives that the test was able to detect and specificity represents the amount of true negatives that the test was able to detect. Likelihood ratios (+/-) are used to determine the chance that an individual has the diagnosis after the test is performed.¹⁶ Further information regarding other statistical measures used can be found in Table 1.

RESULTS

Study Selection

The systematic electronic search of PubMed, CINAHL, Scopus, ProQuest, and a hand search resulted

Table 1. Definitions and Calculations of Diagnostic Accuracy Statistics.

Statistical Measure	Definition	Calculation
Sensitivity	Probability a positive test represents the population with the pathology	$TP / (TP + FN)$
Specificity	Probability that a negative test represents the population without the pathology	$TN / (FP + TN)$
Positive likelihood ratio	Ratio of population with the pathology and a positive test to population without the pathology and a positive test	$SN / (1 - SP)$
Negative likelihood ratio	Ratio of population with the pathology and a negative test to population without the pathology and a negative test	$(1 - SN) / SP$
Positive predictive value	Portion of population with the pathology and a positive test	$TP / (TP + FP)$
Negative predictive value	Portion of population without the pathology and a negative test	$TN / (FN + TN)$
Accuracy	Proportion of the population who were correctly identified as having or not having the pathology	$(TP + TN) / (TP + FP + FN + TN)$
Interrater reliability	Ability of two or more examiners to repeat the test	$K = (P_o - P_e) / (1 - P_e)$
Prevalence	Portion of the population that presently have the pathology	$TP + FP / N \times 100$

TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative, SN=Sensitivity, SP=Specificity, N=Total Number of Subjects, K=Cohen's Kappa, PO=Observed Proportionate Agreement, PE=Probability of Random Agreement

in a total of 1,383 articles. After duplicates were removed the total yielded 1,305 articles. The title and abstracts of these articles were then screened and generated 18 full text articles to be screened ($\kappa=0.83$; 95% CI 0.67-0.98). After full-text screening was completed ($\kappa=1.00$; 95% CI 1.00-1.00), eight articles were found to meet the inclusion criteria.^{8,17-24} The summary of the literature search can be seen in Figure 1.

Study Characteristics

There were a total of 977 subjects between the eight included studies. Of these, 648 were males and 329 females. All of the studies were cohort design studies with subject sizes ranging from 33 to 400 and a mean of 122. The mean age range of patients included in the studies was 23 to 42 years old. Magnetic resonance imaging (MRI) was the reference standard used in two studies,^{8,22} arthroscopy was the reference standard for three studies,^{19,21,24} and both were used in three of the studies.^{18,20,23} Another study created an additional reference standard requiring at least two out of three findings consisting of 1) positive MRI, 2) excessive laxity of more than 3mm as measured on a KT-1000™ arthrometer (measures tibial translation), and 3) a positive findings on an independent examination that was performed at a later time.²⁰ All study characteristics are summarized in Table 2.

Risk of Bias

Five of the included studies^{8,18,20,21,23} were rated as low quality with a high risk of bias. The QUADAS ratings for each study can be found in Table 3. The most common item on QUADAS receiving a “No” or “Unclear” rating was Item 11. Studies in which this item was not met did not provide sufficient information regarding the individual's involvement in the study who was interpreting the reference standard, therefore, it could not be determined if he/she had knowledge of the index test outcome. The Cohen's unweighted kappa calculated for inter-rater reliability of the QUADAS assessment was 0.72, indicating substantial agreement (95% CI 0.58-0.86).

Results of Individual Studies

True negative, true positives, false negatives, and false positives for each of the studies can be found in Table 4. Sensitivity, specificity, positive and negative likelihood ratios, positive predictive values (PPV) and negative predictive values (NPV), and posttest probability values are presented in Table 5. All studies reported a sensitivity within the range of 0.37-1.00.^{8,18-24} Specificity was reported/calculated in five studies and was between 0.50-1.00.^{8,20-24} Positive and negative likelihood ratios were reported/calculated in five studies and were between 1.35-801.00 and 0-0.87, respectively.^{8,20-24} Positive predictive

Figure 1. Flow Diagram of Literature Screening process.

values were reported/calculated in six studies and were between 0.47-1.00.^{8,19-24} In addition, NPV were reported/calculated in five of the studies and ranged between 0.57-1.00.^{8,20-24} Posttest probabilities for positive and negative tests were reported/calculated in five of the studies and ranged between 0.48-infinity and 0-0.43, respectively.^{8,20-24}

Pooled Results

Results were compiled and run through Open Meta-Analyst software revealing a pooled sensitivity and specificity of 0.77 and 0.90 respectively, with a p-value of <0.01. Pooled data can be found in Figures 2-6. Heterogeneity across studies was high, therefore a random-effects model was used. Results of the meta-analysis indicated a high sensitivity (0.77) and high specificity (0.90) for the Lever Sign test. The negative likelihood ratio is 0.22 with a p-value <0.01 and the positive likelihood ratio is 6.60 with a p-value <0.01. The diagnostic odds ratio is 40.70 with a p-value of <0.01. Heterogeneity of the studies

(I²) was calculated for each statistic; sensitivity and specificity were reported as 89.64 and 74.16, negative and positive likelihood ratios were reported at 77.29 and 79.28, and heterogeneity for diagnostic odds ratios was 87.69. Meta-analysis results and forest plots for each statistic can be seen in Figures 2-6.

DISCUSSION

The purpose of this systematic review and meta-analysis was to assess the diagnostic accuracy of the Lever Sign test. Based on the results of a systematic search of the literature, the Lever Sign has shown favorable diagnostic accuracy numbers in detecting ACL tears. The reported sensitivity and specificity for the Lever Sign was shown to be comparable to sensitivities and specificities for anterior drawer test, pivot shift test, and Lachman's test.^{7,25,26} The Clinical Practice Guidelines on Knee Stability and Movement Coordination Impairments: Knee Ligament Sprain Revision 2017 recommends the use of the pivot shift and Lachman's in every suspected

Table 2. Study Characteristics.

Author, Year	Sample Size	M	F	Study Design	Mean Age (Range)	Reference Standard	Exclusion Criteria	Inclusion Criteria
Chong ¹⁹ , 2017	33	21	12	Prospective Cohort Study w/o blinding	30.75 (11-62)	Arthroscopy	<ul style="list-style-type: none"> past knee injury infection on affected side within 72 hours of injury chronic knee pain associated ligament injuries patients complaining of hip/ankle/foot symptoms 	<ul style="list-style-type: none"> U/L knee injury (not sustained <72 hours prior to exam) that resulted in symptomatic instability at 2 selected facilities no prior knee injuries on the involved side or ACL reconstruction/repair No surgical procedures on involved knee 6 wks prior to data collection
Devici ¹⁸ , 2015	117	96	21	Cohort	25.8+-5.9 (17-45)	MRI/Arthroscopy	<ul style="list-style-type: none"> medial meniscus posterior root tear bilateral ACL tear multiple ligament injuries past arthroscopic surgery 	<ul style="list-style-type: none"> ACL tears determined by arthroscopic procedure
Jarbo ²³ , 2017	102	44	58	Cohort	23 (15-66)	MRI/Arthroscopy	<ul style="list-style-type: none"> without MRI evaluation 	<ul style="list-style-type: none"> individuals w/CC of acute knee pain who were examined within 4 wks of their injury or onset
Lelli ⁸ , 2014	400	281	119	Cohort	26.43+-14.9	MRI	<ul style="list-style-type: none"> cartilage defects multi-ligamentous injuries meniscal injuries prior reconstructions of the affected ACL 	<ul style="list-style-type: none"> definitive MRI diagnosis of U/L ACL rupture (partial or complete)
Lichtenberg ²⁴ , 2018	94	57	37	Cohort	34+-15	Arthroscopy	<ul style="list-style-type: none"> malignancies, systemic diseases, CNS disorders, complaints of knee locking, previous (partial) ruptures of the ACL 	<ul style="list-style-type: none"> minimum age of 16 years trauma to the knee Indication for knee arthroscopic surgery
Massey ²² , 2017	91	61	30	Cohort	28+-7 (16-60)	MRI	<ul style="list-style-type: none"> past knee ligamentous reconstruction fracture of the distal femur or proximal tibia bilateral knee injuries known cruciate ligament injury 	<ul style="list-style-type: none"> subjects presenting after a knee injury w/subjective swelling, an objective effusion, and uninjured, contralateral knee (no past injury or surgery).
Mulligan ²⁰ , 2017	60	38	22	Cohort	42+- 13.4 (18-65)	Arthroscopy or Positive finding in 2/3: <ul style="list-style-type: none"> MRI >3mm translation w/KT-1000 independent physician exam 	<ul style="list-style-type: none"> possible fracture based on Ottawa knee rules past knee joint arthroplasty suspicion of PCL involvement knee surgery in last 6 months serious underlying pathology/illness. 	<ul style="list-style-type: none"> subjects were between the ages of 18-65 w/complaint of knee pain rated as less than 7/10 on a verbal numerical rating scale. subjects possessing at least 20-120° ROM were eligible for inclusion.
Thapa ²¹ , 2015	80	50	30	Cohort	32.12 (21-42)	Arthroscopy	<ul style="list-style-type: none"> NONE 	<ul style="list-style-type: none"> aged 20-45 w/knee symptoms of giving way/locking/pain following injury
M=Number of Males, F=Number of Females, U/L=Unilateral, Wks=Weeks, W/=With, CC=Chief Complaint, CNS= Central Nervous System, ROM=Range of Motion, PCL=Posterior Cruciate Ligament								

ACL tear. This recommendation along with other literature indicates that a thorough history and clinical examination by an experienced clinician may be just as accurate as an MRI in diagnosing ACL tears.^{25,26} The CPG reports sensitivity for the pivot shift test between 0.24-0.95, specificity between 0.95-0.98, positive likelihood ratio between 4.37-16.42, and negative likelihood ratio between 0.38-0.84. For Lachman's, the sensitivity is reported between 0.85-0.95, specificity between 0.94-0.95, positive likelihood ratio between 1.39-40.81, and negative likelihood ratio 0.02-0.52.⁷ A meta-analysis reported sensitivity and specificity of the anterior drawer test to be between 0.38-0.63 and 0.81-0.91, respectively. The positive likelihood ratio for anterior drawer was reported as 4.50 and the negative likelihood ratio reported as 0.22.⁶ A third clinical test with high

specificity and sensitivity may be a valuable addition to the clinical examination to rule in as well as rule out ACL tears. A more useful tool in clinical practice may be likelihood ratios, which determine how accurate a test is in predicting a condition. Specificity and sensitivity only take into account two components of the contingency table, meaning that a diagnostic test could potentially have a large amount of false negatives while still having good specificity. Likelihood ratios take into consideration all aspects of the contingency table meaning that a large number of false negatives will largely impact the value. A pooled sensitivity and specificity of 0.77 and 0.90 indicates that the Lever Sign test is good to rule in and rule out ACL tears, however likelihood ratios of 6.60 and 0.22 will only result in a moderate shift in the post-test probability.

Table 3. QUADAS (Quality Assessment of Diagnostic Accuracy Studies) Results of Multiple Lever Sign Studies.

QUADAS Domain*																	
Author, Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	QUADAS Score	Quality Interpretation	
Chong ¹⁹ , 2017	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	U	Y	N	U	10	High	
Deveci ¹⁸ , 2015	N	Y	Y	Y	Y	Y	Y	Y	U	N	U	U	U	U	7	Low	
Jarbo ²³ , 2017	Y	Y	Y	U	Y	U	Y	Y	N	Y	U	Y	U	U	8	Low	
Lelli ⁸ , 2014	N	Y	Y	U	Y	Y	Y	Y	N	Y	Y	N	U	U	8	Low	
Lichtenberg ²⁴ , 2018	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	U	Y	Y	U	11	High	
Massey ²⁰ , 2017	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	12	High	
Mulligan ²¹ , 2017	Y	Y	N	U	Y	N	U	Y	N	Y	N	Y	Y	Y	8	Low	
Thapa ²⁰ , 2015	Y	Y	Y	U	Y	Y	Y	Y	Y	U	U	Y	U	U	9	Low	

N=No, Y=Yes, U=Unsure

*=1. Was the spectrum of participants representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described? 3. Was the reference standard likely to classify the target condition correctly? 4. Was the period between performance of the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample receive verification using the reference standard? 6. Did participants receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test? (that is, the index test did not form part of the reference standard) 8. Was the execution of the index test described in sufficient detail to permit its replication? 9. Was the execution of the reference standard described in sufficient detail to permit its replication? 10. Were the index test results interpreted without knowledge of the results of the reference standard? 11. Were the reference standard results interpreted without knowledge of the results of the index test? 12. Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice? 13. Were uninterpretable, indeterminate or intermediate test results reported? 14. Were withdrawals from the study explained?

Table 4. Lever Sign Test Results Compared to Gold Standard.

Author, Year	TN	FN	TP	FP
Chong ¹⁹ , 2017	0	4*	29	0
	0	1*‡	32	0
	0	6†	27	0
	0	0†‡	33	0
Jarbo ²³ , 2017	46	19	32	5
Lelli ⁸ , 2014	400	0	400	0
Lichtenberg ²⁴ , 2018	46	25	16	0
Massey ²² , 2017	16	12	59	4
Mulligan ²⁰ , 2017	26	15	9	10
Thapa ²¹ , 2015	40	5	30	5

TN=True Negative, FN=False Negative, TP=True Positive, FP=False Positive
 *=Experienced Orthopedic Surgeon, †=Experienced Orthopedic Physician Assistant, ‡=Under Anesthesia

Preliminary research has shown similar sensitivity and specificity numbers for the Lever Sign test and mechanics of this test may allow for a more accurate result in situations where the patient's limb

is too heavy for the clinician to support. All studies included in this review had more male subjects than female subjects, which may influence the clinical application of the test due to ACL tears being

Table 5. Summary of Results.

Author, Year	SN	SP	LR+	LR-	PPV	NPV	PP	PP+	PP-	DOR	Overall Accuracy
Chong ¹⁹ , 2017	.88*	UTD	UTD	UTD	100%	0%	100%	UTD	UTD	UTD	100%
	.97*‡	UTD	UTD	UTD	100%	0%	100%	UTD	UTD	UTD	100%
	.82†	UTD	UTD	UTD	100%	0%	100%	UTD	UTD	UTD	100%
	1.00†‡	UTD	UTD	UTD	100%	0%	100%	UTD	UTD	UTD	100%
	.94§	UTD									
Deveci ¹⁸ , 2015	.98	UTD									
Jarbo ²³ , 2017	.63	.90	6.4	0.41	87%	71%	50%	86.49%	29.23%	15.49	76.47%
Lelli ⁸ , 2014	1	1	∞	0	100%	100%	50%	∞	0%	641,601.00	100%
Lichtenberg ²⁴ , 2018	.39	1	36.9	.61	100%	64.79%	∞	∞	35.21%	60.18	71.26%
Massey ²² , 2017	.83	.8	4.15	0.21	94%	57%	78.02%	93.65%	42.86%	19.67	82.42%
Mulligan ²⁰ , 2017	.38	.72	1.35	0.87	47%	63%	40%	48%	36%	1.56	58%
Thapa ²¹ , 2015	.86	.91	7.71	0.16	85.71%	88.89%	43.75%	85.71%	11.11%	48.00	87.50%

DOR= Diagnostic Odds Ratio, UTD=Unable to Determine, ∞=Infinity, SN=Sensitivity, SP=Specificity, LR+=Positive Likelihood Ratio, LR-=Negative Likelihood Ratio, PPV=Positive Predictive Value, NPV= Negative Predictive Value, PP=Post-test Probability, PP+=Positive Post-test Probability, -PP=Negative Post-test Probability
 *=Orthopedic Surgeon, †=Orthopedic Physician Assistant, ‡=Under Anesthesia, §=Pre-anesthesia, || = Anesthesia

Figure 2. Forest Plot of Sensitivity. Abbreviations: C.I. = confidence interval; TP = true positive, FN = false negative. **Figure 3. Forest Plot of Specificity.** Abbreviations: C.I. = confidence interval; TN = true negative, FP = false positive.

Figure 4. Forest Plot of Negative Likelihood Ratios. Abbreviations: C.I. = confidence interval; FN = false negative; TN = true negative; Di- = disease absent; Di+ = disease present..

Figure 5. Forest Plot of Positive Likelihood Ratios. Abbreviations: C.I. = confidence interval; TP = true positive; FP = false positive; Di- = disease absent; Di+ = disease present.

Figure 6. Forest Plot of Diagnostic Odds Ratios. Abbreviations: C.I. = confidence interval; TP = true positive; TN = true negative; FP = false positive; FN = false negative.

more prevalent in females.⁴ Leg girth differences between males and females may impact sensitivity and specificity due to males tending to have a larger and heavier leg than females. A heavier leg will require more force to be placed through the distal

thigh during the test, and results may be inaccurate if this force is not enough to overcome limb resistance. Certain studies excluded subjects if any concomitant tears were suspected or known or if they had a previous ACL tear on the same limb. Meniscal

and medial collateral ligament (MCL) tears commonly accompany an ACL tear. It is also not uncommon for an athlete to re-tear a reconstructed ACL. In addition, pain, swelling, and muscle guarding are common in patients with acute ACL tears and also were not adequately addressed in the research. Conducting the Lever Sign test on these patients may yield different sensitivity and specificity numbers.

There were several limitations noted in the included studies. Limited studies were available for inclusion and sample sizes varied from small to large. Methodological procedures that predisposed studies to bias were inclusion criteria of confirmed ACL tears, comparing results of uninjured contralateral leg, unclear blinding of examiners, and missing 2x2 tables. The study with the largest sample size only included patients who were confirmed to have an ACL tear via MRI.⁸ The contralateral uninjured leg was assessed using the Lever Sign test and included the results in their 2x2 table. No false negatives or false positives were reported, resulting in a sensitivity and specificity of 1.00. However, this study was not the only one to include patients with a confirmed ACL tear and compare the contralateral uninjured leg. Within studies, there was a lack of false positives found, but it is unclear if this is due to participants with false positives being removed from the study. Authors from three studies^{18,19,23} were contacted for missing data, however only data from two of the authors was obtained. Without the data from the third author, specificity was unable to be calculated and therefore not included in the meta-analysis. Bias was noted within the studies with inadequate descriptions of the examiners performing the test. Exclusion criteria introduced additional limitations within the studies such as excluding patients with concomitant tears or a previous ACL tear on the leg to be tested.

The Lever Sign test is relatively new, therefore, there is limited research on the topic. Strong exclusion criteria were unable to be developed in this systematic review due to the limited number of studies on the topic. Studies were only excluded if they were not written in English, which limits the ability to control the quality of the included studies. Without strong exclusion criteria, study quality was a major limitation. The inclusion criteria developed for this systematic review lead to high Cohen's kappa scores,

with title and abstract screening showing substantial agreement and the full text screening showing almost perfect agreement. While the QUADAS test has been shown to have low interrater reliability in the literature,¹² the kappa score for this systematic review showed substantial interrater agreement.

It is recommended that further research be completed on the diagnostic accuracy of the Lever Sign test. Studies should include a larger number of female participants in order to better represent the overall population of ACL tears. Future research should be completed examining the effect of examiner's hand size on the accuracy of the test. Differences in hand size may place the leg in varying positions, not always maintaining 30° of knee flexion. In this instance, the ACL is not at maximum tension which may yield inconsistent results between examiners. Partial, full thickness, and concomitant tears, as well as acuity, pain, and swelling may also play a role in the accuracy of the Lever Sign test and are parameters that should be included in future studies. It is also recommended that more studies be conducted that do not include the patient being tested under anesthesia in order to improve the clinical applicability of this test. There is currently not enough data to determine any differences in likelihood ratios or sensitivity/specificity when testing under anesthesia vs testing in the awake patient. Tests that can be used to both rule in and rule out pathologies are ideal in clinical practice and research indicates that the Lever Sign test is both specific and sensitive, however quality evidence is currently lacking. Recommendations to reduce bias in future research should be aimed to include larger sample sizes, increased external and internal validity, inclusion of patients with an unknown ACL status, and blinding of examiners to both reference standard and index test results.

CONCLUSIONS

There are a variety of different clinical tests that can be useful in diagnosing ACL rupture. Clinicians utilize clinical tests with a high sensitivity to rule out a diagnosis following a negative test, and specificity to rule in a diagnosis following a positive test. While there is still no single clinical test that is consistently accurate, the results of this study indicate that the Lever Sign test, when used with other tests, may

help provide the clinician with a more complete clinical picture about the status of the ACL.

REFERENCES

1. Nessler T, Denney L, Sampley J. ACL injury prevention: What does research tell us?. *Curr Rev Musculoskelet Med*. 2017;10(3):281-288.
2. Rugg CM, Wang D, Sulzicki P et al. Effects of prior knee surgery on subsequent injury, imaging, and surgery in NCAA collegiate athletes. *Am J Sports Med*. 2014;42(4):959-964.
3. Moses B, Orchard J, Orchard J. Systematic review: Annual incidence of ACL injury and surgery in various populations. *Res Sport Med*. 2012;20:157-179.
4. Renstrom P, Ljungqvist A, Arendt E, et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Br J of Sports Med*. 2008;42(6):394-412.
5. Dargel J, Gotter M, Mader K, et al. Biomechanics of the anterior cruciate ligament and implications for surgical reconstruction. *Strategies Trauma Limb Reconstr*. 2007; 2(1):1-12.
6. Van Eck C.F., van den Bekerom M.P.J., Fu F.H., et al. *Knee Surg Sports Traumatol Arthrosc*. 2013;21: 1895-1903
7. Logerstedt DS, Snyder-Mackler L, Ritter RC, et al. Knee stability and movement coordination impairments: Knee ligament sprain revision 2017. *J Orthop Sports Phys Ther*. 2017;47(11):A1-A47.
8. Lelli A, Di Turi RP, Spenciner DB, et al. The "Lever Sign": a new clinical test for the diagnosis of anterior cruciate ligament rupture. *Knee Surg Sports Traumatol Arthrosc*. 2016;24:2794-2797.
9. Reuben JD, Rovick JS, Schrager RJ, et al. Three-dimensional dynamic motion analysis of the anterior cruciate ligament deficient knee joint. *Am J Sports Med*. 1989;17:463-471.
10. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151:264-269.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
12. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;(3):25.
13. Hollingworth W, Medina LS, Lenkinski RE, et al. Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool. *Acad Radiol*. 2006;13(7):803-810.
14. Cook C, Hegedus E. Diagnostic utility of clinical tests for spinal dysfunction. *Manual Therapy*. 2011;16:21-25.
15. Hegedus Ej, Cook C, Hasselblad V, et al. Physical examination tests for assessing a torn meniscus in the knee: A systematic review with meta-analysis. *J Orthop Sports Phys Ther*. 2007;37(9):541-550.
16. Hegedus Ej, Goode A, Campbell S, et al. Physical examination tests of the shoulder: A systematic review with meta-analysis of individual tests. *Br J Sports Med*. 2008 Feb;42(2):80-92.
17. Wallace, Byron C., Issa J. Dahabreh, Thomas A. Trikalinos, Joseph Lau, Paul Trow, and Christopher H. Schmid. Closing the gap between methodologists and end-users: R as a computational back-end. *J Stat Softw* 2012;49(5):1-15.
18. Deveci A, Cankaya D, Yilmaz S, et al. The arthroscopical and radiological corelation of lever sign test for the diagnosis of anterior cruciate ligament rupture. *Springerplus*. 2015; 4:830.
19. Chong AC, Whitetree C, Priddy MC, et al. Evaluating different clinical diagnosis of anterior cruciate ligament ruptures in providers with different training backgrounds. *Iowa Orthop J*. 2017;37:71-79.
20. Mulligan EP, Anderson A, Watson S, et al. The diagnostic accuracy of the lever sign for detecting anterior cruciate ligament injury. *Int J Sports Phys Ther*. 2017 Dec;12(7):1057-1067.
21. Thapa SS, Lamichhane AP, Mahara DP. Accuracy of Lelli test for anterior cruciate ligament tear. *J Inst Med*. 2015;37(2):91-94.
22. Massey PA, Harris JD, Winston LA, et al. Critical analysis of the lever test for diagnosis of anterior cruciate ligament insufficiency. *Arthroscopy*. 2017;33(8):1560-1566.
23. Jarbo KA, Hartigan DE, Scott KL, et al. Accuracy of the lever sign test in the diagnosis of anterior cruciate ligament injuries. *Orthop J Sports Med*. 2017;5(10).
24. Lichtenberg MC, Koster CH, Teunissen LPJ, et al. Does the lever sign test have added value for diagnosing anterior cruciate ligament ruptures? *Orthop J Sports Med*. 2018;6(3):2325967118759631.
25. Madhusudhan T, Kumar T, Bastawrous S, et al. Clinical examination, MRI and arthroscopy in meniscal and ligamentous knee injuries – a prospective study. *J Orthop Surg Res*. 2008;3(1):19.
26. Kocabey Y, Tetik O, Isbell WM, et al. The value of clinical examination versus magnetic resonance imaging in the diagnosis of meniscal tears and anterior cruciate ligament rupture. *Arthroscopy*. 2004;20(7):696-700.

APPENDIX 1. DETAILS OF PUBMED SEARCH STRATEGY

1 reproducibility of results [Mesh Term]	20 diagnostic test approval [Text Term]
2 reproducibility of results [Text Term]	21 lever sign test [Text Term]
3 sensitivity and specificity [Mesh Term]	22 lelli test [Text Term]
4 sensitivity and specificity [Text Term]	23 lelli's test [Text Term]
5 diagnostic accuracy [Text Term]	24 diagnostic techniques [Text Term]
6 interrater reliability [Text Term]	25 diagnostic procedure [Text Term]
7 intrarater reliability [Text Term]	26 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR
8 validity [Text Term]	20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26
9 result reproducibility [Text Term]	27 anterior cruciate ligament injuries [Mesh Term]
10 sensitivity [Text Term]	28 anterior cruciate ligament injuries [Text Term]
11 specificity [Text Term]	29 anterior cruciate ligament/injuries [Mesh Term]
12 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9	30 anterior cruciate ligament/injuries [Text Term]
OR 10 OR 11	31 anterior cruciate ligament [Mesh Term]
13 physical examination [Mesh Term]	32 anterior cruciate ligament [Text Term]
14 diagnostic techniques and procedures [Mesh	33 full ACL tear [Text Term]
Term]	34 partial ACL tear [Text Term]
15 diagnosis, differential [Mesh Term]	35 acute ACL injury [Text Term]
16 diagnostic test approval [Mesh Term]	36 chronic ACL injury [Text Term]
17 physical examination [Text Term]	37 ACL Rupture [Text Term]
18 diagnostic techniques and procedures [Text	38 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR
Term]	34 OR 35 OR 36 OR 37
19 diagnosis, differential [Text Term]	39 12 AND 36 AND 38