SYSTEMATIC REVIEW

DIAGNOSTIC ACCURACY OF PHYSICAL EXAMINATION TESTS OF THE ANKLE/FOOT COMPLEX: A SYSTEMATIC REVIEW

Braun Schwieterman, DPT[1]
Deniele Haas, DPT[1]
Kirby Columber, DPT[1]
Darren Knupp, DPT[1]
Chad Cook, PT, PhD, MBA, FAAOMPT[1]

ABSTRACT

Background: Orthopedic special tests of the ankle/foot complex are routinely used during the physical examination process in order to help diagnose ankle/lower leg pathologies.

Purpose: The purpose of this systematic review was to investigate the diagnostic accuracy of ankle/lower leg special tests.

Methods: A search of the current literature was conducted using PubMed, CINAHL, SPORTDiscus, ProQuest Nursing and Allied Health Sources, Scopus, and Cochrane Library. Studies were eligible if they included the following: 1) a diagnostic clinical test of musculoskeletal pathology in the ankle/foot complex, 2) description of the clinical test or tests, 3) a report of the diagnostic accuracy of the clinical test (e.g. sensitivity and specificity), and 4) an acceptable reference standard for comparison. The quality of included studies was determined by two independent reviewers using the *Quality Assessment of Diagnostic Accuracy Studies 2* (QUADAS-2) tool.

Results: Nine diagnostic accuracy studies met the inclusion criteria for this systematic review; analyzing a total of 16 special tests of the ankle/foot complex. After assessment using the QUADAS-2, only one study had low risk of bias and low concerns regarding applicability.

Conclusion: Most ankle/lower leg orthopedic special tests are confirmatory in nature and are best utilized at the end of the physical examination. Most of the studies included in this systematic review demonstrate notable biases, which suggest that results and recommendations in this review should be taken as a guide rather than an outright standard. There is need for future research with more stringent study design criteria so that more accurate diagnostic power of ankle/lower leg special tests can be determined.

Keywords: Ankle, leg, physical examination, sensitivity and specificity; systematic review

Level of Evidence: 3a

CORRESPONDING AUTHOR

Chad Cook, PT, PhD, MBA, FAAOMPT
Professor and Chair
2020 East Maple Street
Walsh University
North Canton, OH 44720
Email: ccook@walsh.edu

[1] Walsh University, North Canton, OH, USA

INTRODUCTION

Ankle/foot complex injuries account for a significant number of all orthopedic injuries sustained within the United States.[1,2] These injuries may present with numerous clinical manifestations, resulting in the need for a complex ankle/lower leg examination.[3,4] In order to diagnose and properly manage conditions of the ankle/lower leg, a comprehensive clinical examination is required. Compared with diagnostic imaging methods, physical examination tests or orthopedic clinical special tests have historically been an integral part of clinical examinations and are presently used by a variety of medical professionals as a less costly means of information gathering and confirmation of hypotheses.[5]

Currently, there is little comprehensive research on the diagnostic accuracy of clinical special tests of the ankle/foot complex as most studies have investigated statistical measures of diagnostic imaging, such as ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI);[6,7] tests that are not always cost-effective or available in all clinical settings. Individual studies have investigated the diagnostic accuracy of selected clinical special tests but there are no synthesized, systematically analyzed studies that have described the value of each of these tests. Additionally, many scholastic texts, reviews, and tutorials exist to explain examination techniques, including clinical special tests, which are specific to ankle/foot pathology; however, such reviews merely serve to educate readers on the appropriate application of the tests rather than delving into the statistical measures or utility of each. In a tutorial by Sizer and colleagues[3] diagnostic accuracy measures were reported with the explanation of clinical special tests of the ankle/foot, but the tutorial did not investigate the quality of the studies that obtained the diagnostic accuracy statistics, nor was it all-encompassing of the current, most widely used clinical special tests of the ankle/foot. Similarly, although Cook and Hegedus[8] have assigned diagnostic utility scores to clinical special tests of the ankle/foot complex, this scholastic text lacks the methodology to recognize it as "systematic" in nature.

Consequently, clinicians are often left to assume that the majority of clinical special tests targeting the ankle/lower leg are on an "equal playing field"

with respect to diagnostic accuracy, which in fact may not be the case.[9] Since physical examination tests are designed to rule in or rule out diagnoses and improve decision making, it is imperative to utilize those special tests which yield the best diagnostic abilities. Accordingly, the purpose of this systematic review is to investigate the diagnostic accuracy of ankle/foot complex specials tests. Our goal is to provide clinicians with enough information to determine which physical examination tests are the most appropriate for use in clinical practice.

METHODS

Study design

This systematic review used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines during the search and reporting phase of the research process. The PRISMA statement includes a 27-item checklist and a four-phased flow diagram designed to prospectively improve reporting of systematic reviews and meta-analyses.[10] The PRISMA guidelines were created for use in summarizing randomized controlled trials but can be used for multiple forms of research methodologies.[11]

Eligibility Criteria

Articles were eligible for the review if each included the following: 1) a diagnostic clinical special test of musculoskeletal pathology in the ankle/foot complex; 2) description of the clinical test or tests; 3) a report of the diagnostic accuracy of the clinical test (e.g. sensitivity and specificity) or a report of values that allowed for appropriate calculation; and 4) an acceptable reference standard for comparison, defined as either imaging or surgical confirmation of the diagnosis. Studies were excluded from the review if the diagnostic clinical test was utilized for fractures or if the data were compiled into a cluster, such as in the case of a clinical prediction rule.[12]

Search

Individualized, computer-based search strategies for PubMed, CINAHL, Scopus, ProQuest Nursing & Allied Health Source, SPORTDiscus, and the Cochrane Library were conducted on November 19. 2012. PubMed was searched using a comprehensive search strategy that included search terms related to ankle/foot complex physical examination tests. Lim-

its for the search included; 1) humans and 2) studies published in the English. All remaining databases were searched using comparable strategies (Appendix 1). The authors of this study also completed a hand search by individually searching known physical examination tests of the ankle/foot complex in the databases mentioned above. References of relevant articles and appropriate textbooks were also reviewed in order to find additional studies that met the inclusion criteria of this systematic review.

Study Selection

The review process was performed by two independent authors (using a third author to resolve disagreements) for the, 1) title search, 2) abstract search, and 3) full text search. Reasons for excluding full texts articles were documented in the PRISMA flow diagram (Figure 1). To determine the inter-rater reliability for agreement between title, abstract, and full text reviewers, Fleiss kappa values were calculated. Kappa scores are interpreted as poor (< 0.20), fair (0.21-0.40), moderate (0.41-0.60), strong (0.61-0.80), or near complete agreement (> 0.80).[13]

Data collection process

Data was extracted from each article by one author and a second author verified the information regarding diagnostic accuracy measures of each physical examination test. For the qualitative assessment, diagnostic accuracy (sensitivity and specificity) was reported either through direct transfer from each article or through calculations performed by the authors using 2 x 2 tables. Sensitivity is the probability of a positive test result in someone with the pathology,
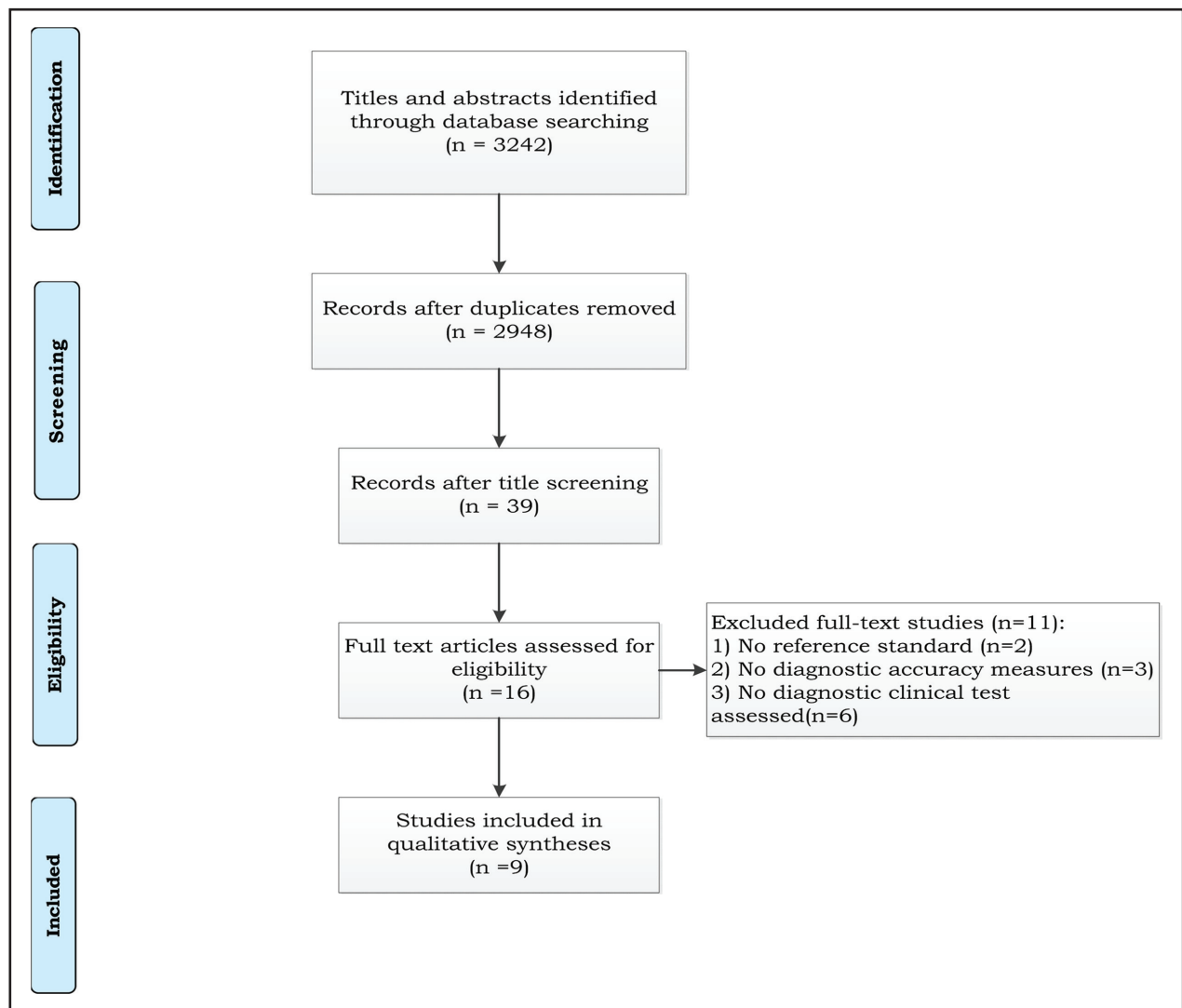
Figure 1. *Study flow for the systematic review.*

whereas specificity is the probability of a negative test result in someone without the pathology.[6] Traditionally, tests which have high sensitivity values are able to correctly identify individuals with the pathology; thus, if the examiner obtains a negative test result, then they can confidently rule out the hypothesized diagnosis. For this reason, these physical examination tests are used early in the examination process to screen out possible pathologies. On the other hand, tests which are determined to have higher specificity values are able to correctly identify individuals without the pathology; thus if an examiner obtains a positive test result, they can rule in the hypothesized diagnosis. These tests are used at the end of the physical examination to confirm pathology.

Description of these tests and measures are outside the scope of this systematic review and the reader is suggested to review each of the included papers for proper procedural description of each of the tests.

Risk of Bias
Each full-text article was reviewed independently by two authors and scored with the *Quality Assessment of Diagnostic Accuracy Studies 2* (QUADAS-2) tool.[14] Group consensus was used to decide the score in cases of disagreement on scoring,. The QUADAS-2 is a retrospective tool that comprises four domains: patient selection, index test, reference standard, and flow and timing. All domains are assessed for risk of bias and the first three domains are assessed for applicability by indicating a "low", "high", or "unclear" rating. In the QUADAS-2, "applicability" refers to whether certain aspects of an individual study are matching or not matching the review question. The QUADAS-2 does not utilize a comprehensive quality score, rather an overall judgment of "low," "high," or "unclear" risk. In order to have an overall judgment of "low risk of bias" or "low concern regarding applicability," a study must be ranked as "low" on all relevant domains. If a study receives a "high" or "unclear" rating in one or more domains, then it may be judged as "at risk of bias" or having "concerns regarding applicability."

Synthesis of results
To determine the accuracy with which each physical examination test identifies each pathology and subsequently its clinical usefulness, sensitivity and specificity as well as positive and negative likelihood

ratios (LRs) were calculated. Both positive likelihood ratios (LR+) and negative likelihood ratios (LR-) are calculated from the data provided from the sensitivity and specificity findings. Positive likelihood ratios greater than 1 increase the post-test probability that the target condition is present, and the higher the positive likelihood ratio the greater this increase. Negative likelihood ratios closer to 0 decreases the probability of the target disorder with a negative finding, and the smaller the negative likelihood ratio, the greater the decrease in probability.[6]

For those physical examination tests which have a sensitivity or specificity of 1.00, corresponding likelihood ratios (LR+ for specificity of 100, LR- for sensitivity of 100) were noted as infinity (INF). A positive finding with a test that has a LR+ >10 generates a large change in post-test probability, 5-10 moderately influences post-test probability, 2-5 generates a small change (sometimes important) in probability, and 1-2 influences post-test probability to a small degree (rarely important). A negative finding with a test that has a LR- <0.1 generates a large shift in post-test probability, 0.1-0.2 moderately shifts the pre to post-test probability, 0.2-0.5 generates small, sometimes important changes, and 0.5-1.0 alters probability to a small, rarely important degree.[15]

RESULTS

Study selection
The database searches resulted in a total of 3,242 total citations that were reviewed for inclusion. After screening, 16 full-text articles were reviewed and five were deemed eligible.[16-20] A detailed hand search was performed and four additional articles[21-24] met the inclusion criteria, resulting in a total of nine studies included in this systematic review (Figure 1). The calculated Kappa scores for the inter-rater reliability of title reviews, abstract reviews, and full-text reviews were .65 (95% CI = 0.50, 0.77), .89 (95% CI = 0.57, 0.99), and 1.00 (95% CI = 0.37, 1.00) respectively. Two papers required third party resolution for clarification.

When using the QUADAS-2 tool to review the nine included articles, it was determined that only one study[18] had low risk of bias and low concern regarding applicability. Three[19,21,22] of the nine included articles were found to be at risk for bias, but had low

concerns regarding applicability. The final five studies[16,17,20,23,24] were judged to be at risk of bias and as having concerns regarding applicability (Table 1).

The reference standard varied throughout the ten studies. See Table 2 for a list of reference standards used for each study.

Lateral ankle sprain

Only one[22] of the ten articles addressed the diagnostic accuracy of physical examination tests for lateral ankle sprains. When scored on the QUADAS-2, the Hertel et al[21] article was determined to be at risk for bias, but had low concern regarding applicability. This study investigated the diagnostic accuracy of three physical examination tests: the anterior drawer test, medial talar tilt stress test, and medial subtalar glide test. The authors of this systematic review calculated the sensitivity and specificity of the anterior drawer test to be .58 and 1.00 respectively. LRs for

this physical examination test were calculated to be Infinite (LR+) and 0.42 (LR-). It was found that the sensitivity and specificity of the medial talar tilt stress test was .50 and .88 respectively (LR+ = 4.00, LR- = 0.57) while those same diagnostic accuracy measures were found to be .58 and .88 (sensitivity and specificity, respectively) and 4.67 (LR+) and 0.48 (LR-) for the medial subtalar glide test (Table 3).

Anterior ankle impingement

Only one study[23] which included the forced dorsiflexion test investigated diagnostic accuracy of a special test for anterior ankle impingement. It was determined that this study was at risk of bias and had concerns regarding applicability. Sensitivity and specificity measure for this physical examination test were reported by the authors to be .95 and .88 respectively From this data, the authors of this systematic review calculated the LR+ to be 8.06 and the LR- to be 0.06 (Table 3).

Table 1. *Presentation of QUADAS-2 results.*

Study	RISK OF BIAS					APPLICABILITY CONCERNS			
	Patient Selection	Index Test	Reference Standard	Flow and Timing	**Overall Risk of Bias**	Patient Selection	Index Test	Reference Standard	**Overall Applicability Concerns**
Abouelea[16]	☹	?	☹	☺	High	☹	☺	☺	High
Beumer[17]	☹	?	☺	☹	High	☺	☺	☹	High
de Cesar[18]	☺	☺	☺	☺	Low	☺	☺	☺	Low
DeOrio[19]	☺	?	☺	?	High	☺	☺	☺	Low
Hertel[21]	☹	?	☺	?	High	☺	☺	☺	Low
Kinoshita[20]	☹	?	?	☹	High	☺	☺	☹	High
Maffulli[22]	☹	☹	☹	☹	High	☺	☺	☺	Low
Molloy[23]	☹	?	☺	☺	High	☹	☺	☺	High
Oloff[24]	☺	☹	☹	☹	High	☺	☺	☹	High
☺Low Risk ☹High Risk ? Unclear Risk									

Table 2. *Reference standards used in the included studies.*

Study	Reference standard used in the study
Abouelela *et al.*[16]	Basic Motor Nerve Conduction at the Tibial Nerve
Beumer *et al.* [17]	Diagnostic imaging and arthroscopy
de Cesar *et al.* [18]	MRI
DeOrio *et al.* [19]	MRI
Hertel *et. al.* [21]	Stress fluoroscopy
Kinoshita *et al.* [20]	Electromyography
Mafulli [22]	Surgical confirmation for all subjects with the diagnosis; MRI and ultrasound in all subjects without the diagnosis
Molloy *et. al.* [23]	Arthroscopy
Oloff *et. al.* [24]	Surgical confirmation

Table 3. *Diagnostic accuracy of ankle/lower leg special tests.*					
Article	Sample Size	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)
Anterior Drawer Test					
Hertel et al. [21]	20 subjects	.58 (.29-.84)	1.00 (.60-1.00)	INF	.42 (.21-.81)
Medial Talar Tilt Stress Test					
Hertel et al [21]	20 subjects	.50 (.22-.78)	.88 (.47-.99)	4.00* (.59-27.25)	.57* (.31-1.04)
Medial Subtalar Glide Test					
Hertel et al [21]	20 subjects	.58 (.29-.84)	.88 (.47-.99)	4.67* (.70-31.04)	.48 (.24-.96)
Forced Dorsiflexion Test					
Molloy et al [23]	73 subjects	.95 (.81-.99)	.88 (.72-.96)	8.06 (3.20-20.30)	.06 (.01-.23)
Tinel's Sign					
Oloff et al [24]	19 feet	.58 (.33-.79)	NA	NA	NA
Triple Compression Stress Test					
Abouelela et al [16]	145 feet	.86 (.76-.92)	1.00 (.93-1.00)	INF	.14 (.08-.24)
Dorsiflexion-Eversion Test Kinoshita et al [20]					
Numbness	144 feet	.25 (.14-.41)	1.00 (.95-1.00)	INF	.75 (.63-.89)
Pain	144 feet	.57 (.41-.71)	1.00 (.95-1.00)	INF	.43 (.31-.61)
Tenderness	144 feet	.98 (.86-1.00)	1.00 (.95-1.00)	INF	.02 (.00-.16)
Cotton Test					
Beumer et al [16]	28 ankles	.25	NA	NA	NA
External Rotation Test					
Beumer et al [17]	294 ankles	NA	.99 (.97-.99)	NA	NA
de Cesar et al [18]	56 subjects	.20 (.04-.56)	.85 (.71-.93)	1.31* (.32-5.41)	.94* (.69-1.30)
Fibular Translation Test					
Beumer et al [17]	322 ankles	.75 (.55-.89)	.88 (.84-.91)	6.30 (4.32-9.19)	.28 (.15-.54)
Syndesmosis Squeeze Test					
de Cesar et al [18]	56 subjects	.30 (.08-.65)	.93 (.81-.98)	4.60 (1.08-19.55)	.75* (.50-1.13)
Posterior Tibial Edema Sign					
DeOrio et al [19]	49 subjects	.86 (.71-.94)	1.00 (.56-1.00)	INF	.14 (.07-.30)
Thompson Test					
Maffulli [22]	161 subjects	.96 (.91-.99)	.93 (.75-.99)	13.47 (3.54-51.25)	.04 (.02-.10)
Matles Test					
Maffulli [22]	105 subjects	.88 (.78-.94)	.86 (.66-.95)	6.18 (2.49-15.37)	.14 (.07-.25)
Palpable Gap in Achilles Tendon					
Maffulli [22]	161 subjects	.73 (.64-.80)	.89 (.71-.97)	6.81 (2.32-19.93)	.30 (.23-.40)
Copeland Test					
Maffulli [22]	14 subjects	.79 (.49-.94)	NA	NA	NA

LR+ = positive likelihood ratio, LR- = negative likelihood ratio, CI = confidence interval, INF = infinity, * = denotes confidence intervals which cross 1.00, NA = not applicable.

Tarsal Tunnel Syndrome

Three articles[16,20,24] assessed the physical examination tests for the diagnosis of tarsal tunnel syndrome. Using the QUADAS-2, all three articles[16,20,24] were found to be at risk for bias and having concerns about applicability. From these studies, three physical examination tests were investigated: Tinel's sign, triple compression stress test, and dorsiflexion-eversion test. Only the sensitivity of Tinel's sign could be calculated, which was found to be .58. LRs could not be calculated to due specificity measures not being reported. Diagnostic accuracy measures for the triple compression stress test were reported to be .86 and 1.00 (sensitivity and specificity, respectively). The LR+ for this physical examination test was INF and the LR- was 0.14. For the dorsiflexion-eversion test, an increase in numbness, increase in pain, and increase in tenderness yielded a sensitivity of .25, .57, and .98 respectively, while specificity was calculated to be 1.00 for all components of the physical examination test. LR+ was found to be Infinite for all components. LRs- for each component are as follows: increased numbness (0.75), increased pain (.43), and increased tenderness (0.02). See Table 3 for these values.

Syndesmotic ankle sprain

Two studies[17,18] addressed the diagnostic accuracy of physical examination tests for syndesmotic ankle sprains. One of the two studies[18] had low risk of bias and low concern regarding applicability, while the other study[17] was judged to be at risk of bias and had concerns regarding applicability. Among these articles, four tests were investigated: Cotton test, external rotation test, fibular translation test, and syndesmosis squeeze test. Both of these studies[17,18] examined the external rotation test. Beumer et al[17] additionally investigated the fibular translation test and Cotton test, whereas de Cesar et al[18] reviewed the syndesmosis squeeze test. For the Cotton test, the authors[17] only reported a sensitivity value (.25), but for the fibular translation test they reported a sensitivity of .75 and specificity of .88. LRs for the Cotton test could not be determined due to insufficient data, but the authors of this systematic review calculated a LR+ of 6.30 and a LR- of 0.28 for the fibular translation test. For the external rotation test, Beumer et al[17] reported a specificity of .99, while

de Cesar et al[18] reported a sensitivity of .20 and a specificity of .85. LRs for the external rotation test as investigated by Beumer et[17] al could not be determined, but the authors of this systematic review calculated a LR+ of 1.31 and a LR- 0.94 from the de Cesar et al[18] study. For the syndesmosis squeeze test, de Cesar[18] reported the sensitivity and specificity to be .30 and .94 respectively. LRs were calculated to be 4.60 (LR+) and 0.75 (LR-). See Table 3 for these values.

Posterior tibial tendon dysfunction

One article[19] assessed the posterior tibial edema sign for posterior tibial tendon dysfunction. This study[19] had low concerns regarding applicability, but was found to be at risk of bias on the QUADAS-2. The authors of this study reported sensitivity to be .86 and specificity to be 1.00. LRs were calculated to be INF (LR+) and 0.14 (LR-). See Table 3 for these values.

Achilles tendon rupture

One study[22] investigated physical examination tests for Achilles tendon rupture and it was judged to be at risk of bias, but had low concerns regarding applicability. A total of four physical examination tests were reviewed: Thompson test, Matles test, Palpable gap in Achilles tendon, and the Copeland test. For the Thompson test, Maffuli[23] reported a sensitivity of .96 and a specificity value of .93. This led to a LR+ of 13.47 and a LR- of 0.04. For Matles test, a sensitivity of .88 and specificity of .86 were reported. The LR+ and LR- were calculated by the author of this systematic review to be 6.18 and 0.14 respectively. The sensitivity for the palpable gap in Achilles tendon was reported as .73, whereas the specificity value was .89. For the palpable gap in Achilles tendon, the LR+ was calculated as 6.81 and the LR- was 0.30. Finally, the Copeland test was reported as having a sensitivity of .78, while a specificity value was not reported. Due to insufficient data, the LRs could not be calculated for the Copeland test.

DISCUSSION

The purpose of this systematic review was to investigate the diagnostic accuracy of special tests of the ankle/foot complex. Nine[16-24] diagnostic accuracy studies were reviewed to determine the effective-

ness of sixteen physical examination tests, all of which were designed for use in identifying one of the following pathologies: lateral ankle sprain, anterior ankle impingement, tarsal tunnel syndrome, syndesmotic ankle sprain, posterior tibial tendon dysfunction, and Achilles tendon rupture. Remarkably, there was only one instance in which more than one study investigated a single test (external rotation test).[17,18]

To the authors' knowledge, this is the first systematic review which investigates the utility and quality of study scoring in clinical specific tests used in identifying ankle/foot complex pathologies. At face value, the findings suggest that with the exception of the external rotation test (low diagnostic accuracy) and the cotton test (only reliability data present) all tests used to identify ankle/foot complex pathology had the diagnostic utility for use in clinical practice. However, it is imperative to note that nearly all studies included had methodological concerns when evaluating the quantitative findings from the studies. Multiple physical examination tests included in this systematic review demonstrated 1.00 specificity values and subsequent INF LR+. This may represent an over estimation of diagnostic power of these tests secondary to the high risk of bias and may be reflective of the small sample sizes in each study. The confidence intervals for the sensitivity, specificity and likelihood ratios were very wide in nearly all studies suggesting a lack of precision in the findings. It is also worth noting that when reported, the 95% confidence interval for some of the test findings crossed 1.00, which indicates that the result of the given physical examination test is no better than chance. Lastly, there were no consistencies regarding reference standards across studies. The authors of the current systematic review feel that there is substantial opportunity to improve upon the current evidence on diagnostic accuracy of the ankle/foot complex through improvements in methodological principles.

It is well recognized that the quality of special clinical tests can influence the outcome of a diagnostic accuracy study.[25] In most cases, bias inflates the diagnostic accuracy finding, suggesting that the test is more diagnostic than it truly seems. Both prospective and retrospective guidelines[25] (including the QUADAS-2)

are designed to allow the clinician/researcher to differentiate the quality of study designs thus further refining which tests are proper for use in clinical practice. Because all the tests in this study had questionable quality, one must consider the currently advocated utility of each test and interpret results with caution.

With respect to use of clinical special tests in clinical practice, it is rare that a test has the capacity to function as a "stand-alone" finding, when attempting to make a diagnosis of any pathology. The findings of the current study should be considered with respect to this concept as well. Any clinical finding should be used in conjunction with a detailed history and other physical examination components (palpation, active/passive range of motion, manual muscle testing, functional testing, etc.) to further enhance the probability of an accurate clinical hypothesis or diagnosis. Clustering examination findings is a practice that has been used by a number of recent diagnostic accuracy studies.[26-30] In these studies, clinical findings that were not "stand alone" provided useful diagnostic utility when clustered with other clinical findings. The authors anticipate that there is great promise when diagnosing ankle/foot pathology using this concept as well.

Limitations
One limitation of this systematic review is that the search strategy was limited to those articles which were published in English. Additionally, two studies[17,25] did not investigate specificity values of the physical examination test of interest which disenabled the authors' of the current study to calculate likelihood ratios for these tests.

CONCLUSION
Traditionally, clinical special tests have been an essential component of the examination process when diagnosing ankle/lower leg pathologies. Such tests are performed either at the beginning of the physical examination to screen out a pathology or at the end of the examination to confirm a diagnosis. The results of this systematic review reveal that the majority of clinical special tests for the ankle/foot complex are confirmatory in nature and may be best utilized at the end of the physical examination process. However, recommending these tests for use in

clinical practice must be tempered by the quality of the included studies findings. Currently, there are a limited number of studies which investigate the diagnostic accuracy of ankle/lower leg pathologies, and of these studies, most demonstrate notable biases. Accordingly, there is a need for future research with more stringent study design criteria so that a more precise estimate of diagnostic power of ankle/lower leg special tests can be determined.

REFERENCES

1. Surve I, Schwellnus MP, Noakes T, et al. A fivefold reduction in the incidence of recurrent ankle sprains in soccer players using the sport-stirrup orthosis. *Am J Sports Med*. 1994;22:601-606.

2. Tropp AC, Gillquist J. Prevention of ankle sprains. Am J Sports Med. 1985;13:259-261.

3. Sizer PS, Phelps V, Dedrick G, et al. Diagnosis and management of the painful ankle/foot part 2: examination, interpretation, and management. *World Institute of Pain*. 2003;3(4):343-373.

4. Davies H, Blundell C. Clinical examination of the foot and ankle. *Orthopaedics and Trauma*. 2011;25(4):287-292.

5. 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;42:80-92.

6. McCarthy CL, Wilson DJ, Coltman TP. Anterolateral ankle impingement: findings and diagnostic accuracy with ultrasound imaging. *Skeletal Radiol*. 2008;37:209–216.

7. Preidler KW, Peicha G, Lajtai G, et al. Conventional radiography, CT, and MR imaging in patients with hyperflexion injuries of the foot: Diagnostic accuracy in the detection of bony and ligamentous changes. *Am Roentgen Ray Soc*. 1999;173:1673-1677.

8. Cook CE, Hegedus EJ. Orthopedic Physical Examination Tests: An Evidence-Based Approach. Upper Saddle River, NJ: Prentice Hall; 2008.

9. Hegedus EJ. Introduction to diagnostic accuracy. In: Cook CE, Hegedus EJ. Orthopedic physical examination tests: an evidence-based approach. Upper Saddle River, NJ:Prentice Hall; 2008:1-11.

10. Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA group Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.

11. Swartz MK. THE PRISMA statement: a guideline for systematic reviews and meta-analyses. *J Pediatr Health Care*. 2011;25(1):1-2.

12. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*. 2000;284:79-84.

13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1997;33:159-174.

14. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536.

15. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. The Evidence-Based Medicine Working Group. *JAMA*. 1994;7(271):703-707.

16. Abouelela AA, Zohiery AK. The triple compression stress test for diagnosis of taral tunnel syndrome. *The Foot*. 2012;22:146-149.

17. Beumer A, Swierstra BA, Mulder PG. Clinical diagnosis of syndesmotic ankle instability: evaluation of stress tests behind the curtains. *Acta Orthopaedica Scandinavica*. 2002;73(6):667-669.

18. de Cesar PC, Avila EM, de Abreu MR. Comparison of magnetic resonance imaging to physical examination for syndesmotic injury after lateral ankle sprain. *Foot Ankle Int*. 2011;32(12):1110-1114.

19. DeOrio JK, Shapiro SA, McNeil RB, et al. Validity of the posterior tibial edema sign in posterior tibial tendon dysfunction. Foot Ankle Int. 2011;32(2):189-192.

20. Kinoshita M, Okuda R, Morikawa J, et al. The dorsiflexion-eversion test for diagnosis of tarsal tunnel syndrome. *J Bone Joint Surg Am*. 2001;83(12):1835-1839.

21. Hertel J, Denegar CR, Monroe MM, et al. Talocrural and subtalar joint instability after lateral ankle sprain. *Med Sci Sports Exerc*.1999;31(11):1501-1508.

22. Maffulli N. The clinical diagnosis of subcutaneous tear of the Achilles tendon. *Am J Sports Med*. 1998;26(2):266-270.

23. Molloy S, Solan MC, Bendall SP. Synovial impingement in the ankle. *J Bone Joint Surg Am*. 2003;85(3):330-333.

24. Oloff LM, Schulhofer SD. Flexor hallucis longus dysfunction. *J Foot Ankle Surg*.1998;37(2):101-109.

25. Cook C, Cleland J, Huijbrets P. Creation and critique of studies of diagnostic accuracy: use of the STARD and QUADAS tools. *J Man Manipulative Ther*. 2007;15:93-102.

26. Cook C, Brown C, Keith M, Isaacs R, Roman M, Howes C, Richardson W, Hegedus E. The Clinical Value of a Cluster of Patient History and Observational Findings

as a Diagnostic Support Tool for Lumbar Spine Stenosis. *Physiother Res International*. 2010 Nov 11. doi: 10.1002/pri.500. [Epub ahead of print]

27. Roman M, Brown C, Richardson W, Isaacs R, Howes C, Cook C. The Development of a Clinical Decision Making Algorithm for Detection of Osteoporotic Vertebral Compression Fracture or Wedge Deformity. *J Man Manip Ther*. 2010;18:44–49.

28. Wainner RS, Fritz JM, Irrgang JJ, et al. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine* (Phila Pa 1976). 2003;28(1):52-62.

29. Cook C, Brown C, Isaacs R, Roman M, Davis S, Richardson W. Clustered clinical findings for diagnosis of cervical spine myelopathy. *J Man Manip Ther*. 2010;18:175-180.

30. Wainner RS, Fritz JM, Irrgang JJ, et al. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. *Arch Phys Med Rehabil*. 2005;86(4):609-18.

APPENDIX 1

Database: PubMed

#1 MeSH descriptor Foot, not exploded
#2 MeSH descriptor Ankle, explode all trees
#3 MeSH descriptor Heel, explode all trees
#4 MeSH descriptor Foot Joints, not exploded
#5 MeSH descriptor Ankle Joint, explode all trees
#6 MeSH descriptor Tarsal Joints, explode all trees
#7 MeSH descriptor Achilles Tendon, explode all trees
#8 MeSH descriptor Lateral Ligament, Ankle, explode all trees
#9 MeSH descriptor Tarsal Bones, explode all trees
#10 ankle
#11 heel
#12 syndesmo*
#13 [OR #1 - #12]
#14 MeSH descriptor Ankle Injuries, explode all trees
#15 MeSH descriptor Foot Injuries, explode all trees
#16 MeSH descriptor Sprains and Strains, explode all trees
#17 MeSH descriptor Joint Instability, explode all trees
#18 MeSH descriptor Foot Diseases, explode all trees
#19 MeSH descriptor Heel Spur, explode all trees
#20 MeSH descriptor Fasciitis, Plantar, explode all trees
#21 MeSH descriptor Metatarsalgia, explode all trees

#22 MeSH descriptor Posterior Tibial Tendon Dysfunction, explode all trees
#23 MeSH descriptor Tarsal Tunnel Syndrome, explode all trees
#24 laxity
#25 impingement
#26 [OR 14# - #25]
#27 MeSH descriptor Physical Examination/ classification, explode all trees
#28 MeSH descriptor Physical Examination/ methods, explode all trees
#29 MeSH descriptor Physical Examination/ standards, explode all trees
#30 MeSH descriptor Physical Examination/ statistics and numerical data, explode all trees
#31 diagnostic accuracy
#32 fibular translation test
#33 external rotation test
#34 cotton test
#35 clunk test
#36 syndesmosis squeeze test
#37 distal tibiofibular compression test
#38 squeeze test of the leg
#39 heel thump
#40 point test
#41 dorsiflexion maneuver
#42 one legged hop test
#43 crossed leg test
#44 anterior drawer test
#45 dimple sign
#46 prone anterior drawer test
#47 medial subtalar glide test

#48   lateral talar tilt stress test
#49   talar tilt test
#50   medial talar tilt stress test
#51   kleiger test
#52   external rotation test
#53   posterior drawer test
#54   inversion stress maneuver
#55   posterior talofibular
#56   thompson test
#57   simmonds test
#58   tinel's sign
#59   forced dorsiflexion test
#60   patla test
#61   morton's neuroma
#62   morton's test
#63   windlass sign
#64   peroneal instability test

#65   peroneal dislocation test
#66   test for peroneal tendon dislocation
#67   [OR #27 - #66]
#68   Clinical Trial [Publication Type]
#69   Comparative Study [Publication Type]
#70   Controlled Clinical Trial [Publication Type]
#71   Evaluation Studies [Publication Type]
#72   Meta-Analysis [Publication Type]
#73   Practice Guideline [Publication Type]
#74   Randomized Controlled Trial [Publication Type]
#75   Review [Publication Type]
#76   Validation Studies [Publication Type]
#77   [OR #68 - #76]]
#78   [#13 AND #26 AND #67 AND #77]