Evidence in Action

How to Evaluate Diagnostic Tests: The Thessaly Test

By Christopher B. Roecker, DC, MS, FACO

A colleague tells you about a new orthopedic test. How do you know if this test is worth integrating into your practice?

You are attending a continuing education seminar and a colleague tells you about a new orthopedic test for evaluating meniscal tears in the knee. The colleague explains that this orthopedic test, called the Thessaly test, is "great" at evaluating meniscal tears and suggests that you begin using it. At this point, you may wonder just how great the Thessaly Test is for diagnosing meniscal tears and whether it is worth incorporating into your physical exam process.

Doctors of chiropractic are commonly faced with challenges when determining a clinical diagnosis, and diagnostic tests are tools used to inform accurate diagnosis. Clinicians are charged with the difficult task of having to navigate an ever changing environment of diagnostic tests; critical appraisal of such tests helps with this process. This article will discuss how to critically appraise studies describing diagnostic tests and will use the example of an article related to the Thessaly test to review this process.

So, where to begin when evaluating diagnostic tests?

Evaluating a diagnostic test requires critically appraising the original article and should focus on 3 primary questions:[1]

1. Are the results valid?
2. What are the results?
3. Can these results help me care for my patient?

Are the results valid?

Assessing the validity of a diagnostic test begins by examining how the study was performed. This information is located in the Methods section of the article. Assessing validity should evaluate how the authors selected the patients and whether a reference standard was used to make comparison with the diagnostic test being studied.

Which patients were included in the study affect the diagnostic test validity. The patients in the study should be representative of a wide spectrum of patients likely to receive the diagnostic test in the real world. It is important that the research study include patients with varying degrees of severity, ranging from mild to severely affected.

The accuracy of a diagnostic test is established by comparing results with a reference or criterion standard, which is occasionally called a gold standard. This reference standard should unequivocally demonstrate the presence of the target condition and should be applied to all patients in the study. Also, the comparisons between the results of the diagnostic test being studied and the reference standard should be blinded. In addition, those individuals making comparisons should be unaware of each test result. This safeguard against bias is commonly referred to as an "independent and blinded comparison."[2]

If the results are valid, proceed to critical appraisal of the results.

What are the results?

Assessing the properties of diagnostic test results requires a dedicated search of the Results section. Diagnostic tests will commonly be reported in terms of *sensitivity* and *specificity*.

- *Sensitivity* is the ability of a test to correctly identify people who have the target disorder. Therefore, a negative result from tests with high sensitivity is good for ruling out a particular disorder.
- *Specificity* is the ability of a test to correctly identify people who do not have

the target disorder. Therefore, a positive result from tests with high specificity is good for ruling in a particular disorder.

More recently, *likelihood ratios* are being reported in diagnostic test results. Likelihood ratios combine specificity and sensitivity statistics and are used to assess the value of performing a diagnostic test. A likelihood ratio is a test statistic that links the pre-test probability with the post-test probability.[3] It is important to remember that a likelihood ratio of 1.0 will have no influence on the post-test probability and performance of such a test is of no diagnostic value.[3] A guide for interpreting likelihood ratios is provided in Table 1.

If the results of the diagnostic test are meaningful, continue to evaluate the diagnostic test for your individual needs.

Can these results help me care for my patient?

Clinicians should evaluate the reproducibility of the diagnostic test, the applicability within their unique patient population, and whether using this test will have an impact on patient management.

The performance of a diagnostic test should be described with sufficient detail for you to accurately reproduce the test. Diagnostic studies should describe the performance and interpretation of the diagnostic test. Tests that are highly reproducible reduce the likelihood of erroneous test results, such as false positives or false negatives, and increase the likelihood of an accurate diagnosis.

Clinicians should also assess the applicability of the test to their own clinical practices. This is accomplished by comparing the study population with the patients you commonly encounter in clinical practice. Also, evaluating whether your clinical setting is sufficiently similar to the research setting will help inform whether using a diagnostic test is appropriate in your practice. Finally, you should consider the costs of performing a diagnostic test when assessing its appropriateness. It is easy to limit your consideration of costs to a test's affordability, but considerations should also be made for factors such as the discomfort associated with a test or the length of time required. All such factors should be weighed against the diagnostic utility (usefulness) of a diagnostic test when determining appropriateness.

Testing should not be performed indiscriminately. A diagnostic test's utility should always outweigh the costs associated with the test. Whether a particular diagnostic test should be applied is a decision best made by a clinician with a thorough understanding of the validity, clinical importance, and appropriate application of the test in question.

Returning to our scenario:

Your colleague explained a new orthopedic test, called the Thessaly Test. You would like to know how useful this test is for diagnosing meniscal tears and whether this is worth incorporating into your physical exam process.

You locate the following study regarding the Thessaly test:

Harrison BK, Abell BE, Gibson TW. The Thessaly test for detection of meniscal tears: validation of a new physical examination technique for primary care medicine. *Clin J Sport Med*. 2009 Jan;19(1):9-12.

Abstract

OBJECTIVE:

To assess the validity of a new clinical test (Thessaly) as a means of detecting meniscal tears of the knee by comparing arthroscopic findings to a clinical examination finding.

DESIGN:

Retrospective cohort study.

SETTING:

All preoperative examinations were performed in the Department of Orthopedic Surgery, a secondary care center, Dwight David Eisenhower Army Medical Center, Fort Gordon, Georgia.

PARTICIPANTS:

116 consecutive patients undergoing knee arthroscopy for suspected meniscal pathology.

INTERVENTION:

The Thessaly test was performed during the preoperative examination as previously described. The clinician supports the patient by holding the patient's outstretched hands while the patient stands flatfooted. The patient then rotates the knee and body, internally and externally, three times, keeping the knee flexed at 20 degrees. Patients with suspected meniscal tears will experience joint-line discomfort.

MAIN OUTCOME MEASURES:

The Thessaly test performed at 20 degrees of flexion and arthroscopic surgical diagnosis.

RESULTS:

Of the 66 patients with a positive Thessaly test, 65 had an arthroscopically verified meniscal tear. The Thessaly test revealed a sensitivity of 90.3%, specificity of 97.7%, positive predictive value of 98.5%, negative predictive value of 86.0%, likelihood ratio for a positive test of 39.3, likelihood ratio for a negative test of 0.09, and diagnostic accuracy of 88.8%. The resulting kappa coefficient revealed a statistically significant level of agreement ($P < 0.001$) for the surgical diagnosis and the Thessaly test.

CONCLUSIONS:

The Thessaly test is a valid and reproducible physical examination technique for predicting meniscal tears. The Thessaly test shows promise as an easily performed maneuver that may have better diagnostic accuracy than traditional tests. However, this study was performed at a referral center; therefore, the diagnostic relevance cannot be appropriately applied to a more generalized population.

Critical appraisal of this diagnostic study:

Focusing on the 3 primary questions listed above allows for a structured and meaningful assessment of the article.

1. Are the results valid?
   The article states that patients were recruited via consecutive referral to an orthopedic surgical center with symptoms suggestive of meniscal pathology. Recruitment was open to patients of all ages. Ages ranged from 11 through 67. This study compared the results of the Thessaly test with the findings from diagnostic arthroscopy for all 116 patients. The article failed to report whether the reference standard was independently performed and whether those performing and/or interpreting the results were blinded to the results of the Thessaly test.
2. What are the results?
   Several test statistics were summarized and reported in the article, including sensitivity, specificity, positive likelihood ratio (LR+), and negative likelihood ratio (LR-). The Thessaly test was reported to have 90% sensitivity, 98% specificity, an LR+ of 39, and an LR- of 0.09. These results indicate that the results of Thessaly are likely to produce a large and often conclusive shift in diagnostic probability, which means it is highly useful for ruling in and ruling out meniscal tears and has high diagnostic utility.
3. Can I apply the results to patient care?
   The article provides a detailed description of the performance and interpretation of the Thessaly test, which supports the accurate reproducibility of this test. Conveniently, this article also includes images of examiners who were performing various stages of the Thessaly test.
   The patient population used for this study involved 116 patients referred to an orthopedic surgical center with knee pain (suggestive of meniscal tear) that was nonresponsive to 6 weeks of conservative management. The average patient was 36 years old. Ages ranged from 11 through

67, and approximately 59% of all patients were male.

Costs associated with the Thessaly test appear to be limited to the exacerbation of knee pain (positive test result). It is important to note that the Thessaly test does not require any equipment, there are no financial costs, and it can be performed within minutes.

How do these findings relate to your clinical practice?

These findings from the critical appraisal of the Thessaly test should be compared with other diagnostic tests used to evaluate meniscal tears. It is ultimately up to individual clinicians to select the diagnostic tests that are most appropriate and useful to their own practices. The use of any diagnostic test should inform clinical decision-making, and the Thessaly test is an option that may be appropriate to integrate into your clinical practice.

Note on terms used:

Target condition: the pathological condition of importance (to be ruled in or ruled out)

Pre-test probability: the likelihood that a patient exhibits a specific disorder prior to the performance of a diagnostic test.

Post-test probability: the likelihood that a patient has a specific disorder after a diagnostic test has been performed.

Christopher B. Roecker, DC, MS, FACO, is an instructor in the Life Sciences Department at Palmer College of Chiropractic, Davenport, Iowa.

References

1. Furukawa TA, Strauss S, Bucher HC, Guyatt G. Diagnostic tests. In: Guyatt G, Rennie D, Mead MO, Cook DJ, editors. *Users' guide to the medical literature: a manual for evidence-based clinical practice*. 2nd edition. New York: McGraw-Hill; 2008. p. 399-406.
2. Bhandari M, Guyatt GH. How to appraise a diagnostic test. *World J Surg*. 2005 May;29(5):561-566.
3. Jaeschke R, Guyatt GH, Sackett DL. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA.* 1994;271:703-707.

Tables

Table 1. Interpreting likelihood ratios*

LR+	LR-	Interpretation
>10	<0.1	Large and often conclusive shift in probability
5 – 10	01 – 0.2	Moderate shift in probability
2 – 5	0.2 – 0.5	Small, but sometimes important, shift in probability
1 – 2	0.5 – 1.0	Small, and rarely important, shift in probability

LR+, positive likelihood ratio; LR-, negative likelihood ratio

* Adapted from Jaeschke R, Guyatt GH, Sackett DL. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-707.

Table 2. Summary of Thessaly test (20° flexion) results*

	Sensitivity	Specificity	LR+	LR-
Injury of medial meniscus	89%	97%	29.6	0.11
Injury to lateral meniscus	92%	96%	23.0	0.08

LR+, positive likelihood ratio; LR-, negative likelihood ratio

* Adapted from Karachalios T, Hantes M, Zibis AH, et al. Diagnostic accuracy of a new clinical test (the Thessaly test) for early detection of meniscal tears. *J Bone Joint Surg Am*. 2005 May;87(5):955-962.