Physical Examination of the Shoulder: Considerations of Sensitivity and Specificity

LAUREN C. OLMSTED, MEd, ATC, and CRAIG DENEGAR, PhD, ATC, PT
The Pennsylvania State University

Evidence-based sports medicine is becoming a reality for allied health educators, as well as practitioners. For health-care professionals, examining current literature related to pathology, diagnosis, treatment, and rehabilitation of sport-related injuries is essential. What works, why does it work, and how does it work are questions that must always be considered. Clinicians need to know whether a particular strategy is effective, whether a special test is valuable in making a diagnosis, or whether a specific treatment improves the outcome of care. An important point to remember is that just because something might seem commonplace in sports medicine, that does not mean that it is actually effective. For example, the McMurray test for meniscal damage is of little value, especially in this age of diagnostic imaging (C. R. Denegar and M. A. Fraser, unpublished data, 2003).

One area of sports medicine that needs to be examined further is the efficacy of diagnostic or special tests at the shoulder. In order to determine whether a test is accurate in identifying a pathology, the test must be compared to a gold standard that has already been determined to be accurate.1,2 For example, when performing a shoulder examination for labral pathology, how does a special test compare to the gold standard of arthroscopy? The purpose of this article is to examine the diagnostic usefulness of the clinical assessment of glenoid labral pathology, or SLAP (superior labrum anterior–posterior) lesions, in the shoulder.

The most common clinical symptoms of labral pathology are pain, clicking, or popping deep in the glenohumeral joint.3 Special tests at the shoulder to assess labral pathology should therefore try to reproduce these symptoms. The compression-rotation test; anterior-slide test; active-compression, or O’Brien, test; crank test; pain-provocation test; and diagnostic magnetic resonance imaging (MRI) have all been proposed to assist in diagnosis of labral tears. This article focuses on the sensitivity, specificity, and positive and negative likelihood ratios associated with these tests.

Definitions

The study of diagnostic procedures requires a working understanding of key terms. In order for a special test to be effective it must be accurate. Accuracy, also known as validity, means that the test is measuring what it is supposed to measure.4 In order to determine whether a test is accurate, test results must be compared to an established gold standard with a known accuracy. For example, arthroscopic surgery performed by a skilled surgeon is 100% accurate for diagnosing labral tears. If the O’Brien test is positive in
200 patients and all 200 patients are arthroscopically diagnosed with a labral tear, then the O’Brien test is also 100% accurate.

When examining the effectiveness of a special test, the clinician must ask whether the gold standard is “gold enough.” The gold standard used as the criterion should be highly accurate. Second, it must be applied to all patients that undergo a special test. Third, the patients being tested should not have symptoms too severe or not severe enough for the test to appropriately measure them. The patients should be representative of those in a typical practice setting. Finally, the examiner performing the special test should be blinded to the medical history of the patients.

Sensitivity, the ability to identify everyone with a particular condition and miss nobody, is defined as the proportion of patients with a condition who have a positive test. It is also known as the true positive. Tests with high sensitivity have fewer false negatives, so a negative result might help rule out the condition. The formula for calculating sensitivity (see Table 1) is

\[
\text{Sensitivity} = \frac{A}{A + C}
\]

For example, if 100 people with shoulder pain were tested for a labral tear via the O’Brien test and 90 people both had a positive O’Brien test and were diagnosed with a tear after arthroscopy, the sensitivity of the O’Brien test would be 90%.

Specificity is the proportion of patients without a particular condition who have a negative test. It is also known as the true negative. Tests with high specificity have few false positives, so a positive result helps rule in the condition. The formula for calculating specificity (see Table 1) is

\[
\text{Speciﬁcity} = \frac{D}{B + D}
\]

For example, if 100 people with shoulder pain were tested for a labral tear via the O’Brien test and 80 both had a negative O’Brien test and were diagnosed without a labral tear after arthroscopy, the specificity of the O’Brien test would be 80%.

There are no special tests for labral pathology that are both 100% sensitive and 100% specific. Although sensitivity and specificity are useful to rule in or rule out a condition, most tests with high sensitivity will have low specificity, and vice versa. In order to better understand how to apply sensitivity and specificity values to make clinical decisions, likelihood ratios can be calculated. A useful test will produce a large shift in the probability that a condition is present (after a positive test) or absent (after a negative test).

A positive likelihood ratio (+LR) helps explain the impact that a positive test has on raising the suspicion that a condition really exists. A positive test with a high +LR further increases confidence in a working diagnosis. The formula for calculating a +LR is

\[
\text{A positive likelihood ratio (+LR)} = \frac{\text{sensitivity}}{1 - \text{specificity}}
\]

For example, an athlete reports to the athletic training room or clinic with shoulder pain, and based on history and observation you suspect that there is a 50% chance that he or she has a biceps-tendon rupture. You then perform a test to examine biceps-tendon integrity that has a high +LR. If the test is positive, the probability that there really is a biceps-tendon tear is increased substantially. A positive test for a biceps-tendon tear that has a 3.5 +LR, for example, means that the odds of having a tear are 3.5 times greater than the odds based on history alone. In other words, you are increasing the odds of a condition existing 3.5 times over the pretest estimates.

The higher the +LR, the greater the likelihood of concluding that a condition exists after a positive test. LR greater than 10 generate large and often conclusive shifts in probability, LR between 5 and 10 generate moderate shifts, LR between 2 and 5

<table>
<thead>
<tr>
<th>Table 1. Calculation of Sensitivity and Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Standard (+)</td>
</tr>
<tr>
<td>Special test (+)</td>
</tr>
<tr>
<td>Special test (–)</td>
</tr>
</tbody>
</table>

Note: Sensitivity = A/(A + C); Specificity = D/(B + D).