



## Evaluating Diagnostic Tests: Challenges with Measures of Test Function

<p><b>Key Points</b></p>	<ul style="list-style-type: none"> <li>A goal of diagnostic testing is to reduce diagnostic uncertainty. Yet there is usually <b>uncertainty associated with diagnostic testing itself</b>.</li> <li>Evaluating a diagnostic test usually entails making a comparison when one is available. Yet, there are often <b>problems with making test comparisons</b>. And there are problems when there is no comparator.</li> <li>"Measures of Test Function" evaluate the accuracy and prediction capabilities of diagnostic tests. There are <b>often problems with Measures of Test Function</b>.</li> </ul> <p>Consequently there are special considerations for critically appraising studies of diagnostic tests <b>in addition to usual considerations of study validity, clinical relevance, applicability and usability</b>.</p>
<p><b>How much uncertainty is associated with this test?</b></p>	<p><b>Uncertainty can rarely be eliminated</b> due to –</p> <ol style="list-style-type: none"> <li>Uncertainty about what equates with a meaningful result since <b>assignment of normal and abnormal values</b> is usually arbitrary when dealing with a range, and "normal/abnormal" may not equate with being disease-free or having a disease;</li> <li><b>Trade-offs between sensitivity and specificity</b> – Setting the cut-off to identify more patients with the disorder will almost always yield more patients with false positives. For example, if you set the cut-off for an abnormal fasting blood sugar at a low level to identify more diabetics (higher sensitivity), you will pay the price of including more non-diabetics (lower specificity and high false positive rate) as well.</li> <li><b>Variations in the test's</b> accuracy and precision, its application, its predictive capabilities, and/or its interpretation;</li> <li><b>Variations in how values might vary</b> within an individual, a population or within different populations – including assumptions about who is and who is not disease-free, and variations in disease spectrum such as early to late disease, or mild or severe disease, or rate of disease progression;</li> <li>Sometimes having a test for an <b>intermediate outcome</b>, but not having good information available about whether the outcome is actually associated with meaningful clinical outcomes (e.g., PVCs following an MI indicate higher risk for cardiac mortality)</li> <li>Frequently needing to choose a <b>less accurate method</b> due to cost or risk (e.g., chest x-ray vs lung biopsy).</li> </ol> <p>Consequently, the uncertainty surrounding the diagnostic test must be evaluated.</p>
<p><b>Is the comparison valid?</b></p>	<ul style="list-style-type: none"> <li>Frequently there is <b>no single, accurate test</b> for diagnosis. For example the diagnosis of rheumatoid arthritis involves history, physical exam plus laboratory testing.</li> <li>Often there is <b>no way that is 100% accurate</b> to establish a diagnosis. Comparing a new diagnostic test or procedure to an inaccurate "standard" may make it seem that the new method is in error even if it is actually better than the current "standard."</li> <li>It is <b>rare that a test is both highly sensitive and highly specific</b>, which can make it difficult to find a perfect gold standard.</li> <li>Often good <b>information about negative tests is lacking</b> since patients with negative tests are generally not subsequently exposed to further invasive or uncomfortable testing. Therefore, it is more likely to be unknown if the negative results are valid or invalid.</li> </ul> <p>Consequently, the comparison method must be evaluated.</p>
<p><b>Measures of Test Function – Is the test accurate and predictive?</b></p>	<p>Results for measures of test function may be <b>misleading</b> depending on the population used to make those calculations.</p> <ul style="list-style-type: none"> <li>Calculations are often based not on <b>prevalence</b> within a community, but in the pool tested.</li> <li>The test's reported <b>sensitivity can be misleading if the sensitivity is determined in a patient population that is different</b> from where the test is applied. For example, the sensitivity of CPK for an MI may be overstated if it is determined using CCU patients, but the test is used in general hospital admissions. This is due to the greater severity of disease in the CCU, which may result in a greater likelihood that those with the disease test positive. Conversely, a general population includes more people presenting earlier in the course of the disease. These people with early disease are more likely to have lower CPK levels, resulting in a lower sensitivity for the test.</li> <li>Prevalence and severity of disease is often higher in <b>academic settings</b> than in the general population, and sensitivity may be higher than if the study or test had been performed in a "usual care" setting.</li> </ul> <p>Consequently, variables like age and gender in the study subjects must be evaluated, along with severity, stage and duration of disease to ensure that various stages of disease have been studied, to determine if the measures of test function are appropriate for your population.</p>

