Diagnosis
Sample CAT worksheet for a diagnosis paper, from McMaster University (go ahead and download the paper version, which has answers and sample calculations):

Diagnosis Paper Worksheet

Key Concepts:
General
The critical concepts in diagnostic testing are:
1. Diagnostic certainty and treatment threshold
2. Morbidity / mortality risk of the illness
3. Morbidity / mortality risk of the treatment
4. Pre-test probability of disease
5. Positive and negative likelihood ratios
6. Post-test probability of disease
7. Tests are not always dichotomous variables

- Diagnostic certainty is closely related to treatment threshold
  ♠ If the illness carries low morbidity or mortality risk (viral URI), and the proposed treatment is also benign (fluids or acetaminophen), then the diagnostic certainty does not need to be very high for action to be taken, and thus no diagnostic test needs to be performed.
  ♦ If the illness carries high morbidity or mortality risk (for example PE), the treatment threshold may not be very high. You probably would treat for PE if there was a 15% chance that PE is the correct diagnosis.
  ♠ If the treatment is not benign (bone marrow transplant, thrombolytics), then the diagnostic certainty needs to be fairly high before initiating treatment, and a diagnostic test needs to be performed.

- Pre-test probability of disease
  ♥ If the prevalence of disease in a certain patient population is known, it can be used as the pre-test probability.
  ♠ If a clinical prediction rule exists, it may be used to determine the pre-test probability (e.g. Wells criteria for DVT or PE).
  Example: Link to CPRs
  ♠ Your clinical acumen and experience may be used to determine the pre-test probability.

- Likelihood ratios. (link to LR nomogram)
  ♦ For any given test, each result has its own likelihood ratio. It can be argued that there is no such thing as a “positive” or “negative” test result; tests only make you either more or less certain of disease than you were before the test.
    - The LR is the numerical value for how much more or less certain you are of disease after obtaining a test result. If the LR is >1.0, then you are more certain of disease than you were before the test was done. If the LR is <1.0, then you are less certain of disease than you were before the test.
  ♥ Consider the diagnosis of iron deficiency. Let’s say a serum ferritin level is performed on 260 elderly patients who also receive bone marrow biopsies (gold standard for diagnosing iron deficiency). Every single ferritin result (from 1 ng/ml to 2000 ng/ml) has a likelihood ratio associated with it. At some point, say
around 40 ng/ml, the likelihood ratio equals one. At that point – where the likelihood ratio is 1 – as many people with that particular test result (ferritin = 40) have the disease (iron deficiency) as do not.

Pearl: A likelihood ratio of 1 is the only test result which is completely unhelpful to you; all other values will either make you more or less certain of disease than you were before testing.

Pearl: A strong + LR is typically 10 or greater. A strong − LR is typically 0.10 or lower.

→ Test results are not always dichotomous variables.

♠ The UNC lab uses 3 ng/ml as the cutoff for a “normal” ferritin. Let’s examine why assigning “normal” or “abnormal” to lab results is often misleading.

Example: Let’s say the pre-test probability of iron deficiency in your elderly patient is 20%. Oral iron is somewhat unpleasant to take – and iron overload can be harmful - so you do want to perform a test to help decide whether the patient is truly iron deficient before starting iron replacement therapy. The serum ferritin value returns as 8 ng/ml, which the lab reports as “normal”. Ferritin of 8 has a + LR for iron deficiency of around 40. Thus, the probability of iron deficiency after performing this test is 90%! The test moved you from 20% probability of disease to 90% probability of disease.

♣ This logic can be applied to almost all tests: 3 mm of ST elevation after 2 minutes of exercise on a treadmill is a “positive” test result, as is failure of HR to recover after 12 minutes of exercise. Despite both being “positive” test results, they would clearly have different positive likelihood ratios for diagnosis of CAD.

→ Link to Likelihood Ratios for common physical exam findings based on the JAMA rational clinical examination series

→ Link to Likelihood Ratios for commonly ordered tests, developed by Mike Pignone MD, UNC Dept of General Medicine

Are the results valid?

→ Did clinicians face diagnostic uncertainty? We don’t need a test that rules out MI in 20 year old women with rib pain; likewise we don’t need a test that tells us that grouped painful vesicles in an elderly man are herpes zoster. We need help in the grey area in-between these extremes.

♦ Sometimes pre-test probability of disease is so low that it does not even reach your test threshold; that is to say, if the test result were positive, you wouldn’t believe it (you would consider it a false positive).

♥ Likewise, there are situations where you are so certain of disease that you would not believe a negative test result (you would consider it a false negative).

♠ In these cases – where you are already below your test threshold or above your treatment threshold – a test is not helpful, because in these situations there is no diagnostic uncertainty.

→ The way this applies to the validity of your Journal Club paper is summed up by asking the following questions: did the investigators test appropriate patients, and are my patients similar to the patients in the study?
The Gold Standard: the accuracy of a test is determined by comparison with the reference standard (representing “the truth”). No gold standard is perfect (not even biopsy or autopsy), so you have to ask yourself whether you accept the investigators’ gold standard.

Pearl: The gold standard is unacceptable if it includes the test being studied.

If you accept the gold standard, next assess whether the test in question and the gold standard test were interpreted blindly (not being blinded introduces bias; for example, we can all hear the murmur after we’ve seen the echo; we can all find the spot on CXR after we’ve seen the CT).

Long term follow up is usually an acceptable gold standard, but look to see if the investigators performing the follow up were blinded to test results.

What are the results?
The 2x2 table. You use a two by two table when results are reported in dichotomous fashion. You can also use 2x3, 2x4 or 2 by whatever for non-dichotomous results.

“The truth goes on top”. Thus, the gold standard (representing truth), goes on top of your table.

Consider the previous example of the diagnosis of iron deficiency anemia:

**The test**: ferritin **The gold standard**: bone marrow biopsy

2x2 table

<table>
<thead>
<tr>
<th></th>
<th>BMBx positive</th>
<th>BMBx negative</th>
<th>+ LR</th>
<th>- LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferritin &lt;19</td>
<td>47 (a) (TP)</td>
<td>2 (b) (FP)</td>
<td>41.5</td>
<td></td>
</tr>
<tr>
<td>Ferritin ≥ 19</td>
<td>38 (c) (FN)</td>
<td>148 (d) (TN)</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Formulae

\[
LR^+ = \frac{TP}{TP+FN} / \frac{FP}{FP+TN}
\]

\[
LR^- = \frac{FN}{FN+TP} / \frac{TN}{TN+FP}
\]

\[
LR^+ = \frac{a}{(a+c)} / \frac{b}{(b+d)}
\]

\[
LR^- = \frac{c}{(a+c)} / \frac{d}{(d+b)}
\]

\[
Sens = TP / (TP + FN)
\]

\[
Spec = TN / (TN + FP)
\]

\[
PPV = TP / (TP + FP)
\]

\[
NPV = TN / (TN + FN)
\]

2x4 table

<table>
<thead>
<tr>
<th></th>
<th>BMBx positive</th>
<th>BMBx negative</th>
<th>+ LR</th>
<th>- LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferritin &lt;19</td>
<td>47</td>
<td>2</td>
<td>41.5</td>
<td></td>
</tr>
<tr>
<td>19-45</td>
<td>23</td>
<td>13</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>46-100</td>
<td>7</td>
<td>27</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>&gt;100</td>
<td>8</td>
<td>108</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Likelihood ratio in words: a serum ferritin value of less than 19 ng/ml is 41 times more likely to occur in a patient with iron deficiency than in an iron replete patient.
Converting pre-test probability to post-test probability

Step 1. Convert pre-test probability to odds
Step 2. Multiply odds by LR
Step 3. Convert post-test odds back to probability

Example:
Pre-test probability of iron deficiency: 20%, or 20/100, or 1/5
Convert probability to odds: odds = 20:80 (or, 1/4)
Ferritin = 15 ng/ml, LR = 41 (see table above)
multiply the odds by the LR: (1/4) x 41 = 10
Post test odds: 10:1 (10/1)
Convert to probability = 10/11, or 91%

Why bother converting probability to odds?
What if I said I watched the news last night, and they said the probability of rain on Tuesday is 60%. Today I hear the weatherman say it’s now twice as likely to rain on Tuesday. So what is the probability of rain on Tuesday now? Is it 120%? No! You can’t multiply probabilities, you can only multiply odds!
Initial probability of rain: 60%, or 60/100
Initial odds: 60:40, or 3:2, or 3/2
LR = 2 (twice as likely)
Multiple odds by LR: 3/2 x 2 = 6/2 = 3
Revised odds = 3:1, or 3/1
Revised probability of rain on Tuesday = 3/4, or 75%

Why is the positive predictive value less helpful than the positive likelihood ratio?
Because the PPV is entirely dependent on the underlying prevalence of disease in the population being studied, something we don’t often know.

Consider the example of using the ANA test for diagnosis of lupus. No matter what patient you are testing, the LR+ for the ANA test is 19 (hypothetical), and does not change.

♦ So, if you perform the ANA on a 70 year old man with joint aches, the LR+ is 19; and if you use perform the ANA on a 40 year old female with joint pain, photosensitive rash, and pleurisy, the LR+ is 19.

♦ However, in the first patient the PPV for ANA is 1.9%; and in the second patient, the PPV for ANA is 95%. A huge difference in the property of the ANA test from patient to patient!

♥ What is the difference between the 70 year old with joint aches and the 40 year old with joint aches, photosensitive rash and pleurisy? The difference is the prevalence of lupus in these two very different patient populations. In 70 year olds with joint aches, the prevalence of lupus is less than 1% (say 0.10%). In 40 year olds with three of the most common lupus symptoms, the prevalence of lupus is much higher (say 50%).

♦ Link for ANA worksheet
While the LR doesn’t change from patient to patient, the predictive values may vary widely. Thus PPVs are largely unhelpful to us, while LRs are very useful.

For conditions like iron deficiency and hypothyroidism, there may be multiple diagnostic tests available. For iron deficiency, there are the ferritin, mean cell volume (MCV), transferrin saturation (T Sat), red cell volume distribution, and others. Likewise for hypothyroidism there are the TSH, total T4 level, free T4 level, thyroid binding index, and others. Why have ferritin and TSH been chosen as the tests of choice?

The receiver operator characteristic curve (ROC curve) plots the TP rate on the y axis and the FP rate on the x axis. The plot with the largest area under the curve is the best test. The point at which the slope is such that one sees the most TPs for the fewest FPs is often chosen as the test “cut-off” between normal and abnormal. Every point on the curve represents a positive likelihood ratio (TP rate over FP rate).