Suggested answers to exercise: critical appraisal of a diagnostic test study

There are no 'right answers' to most of these questions. It involves both judgement and the reader's prior knowledge of the subject area of the paper being appraised.

Greenhalgh criteria

1: *Is this test potentially relevant to my practice?* Only you can answer this one. Not to mine. There are two new tests here: culture in liquid media and cell count. Direct culture is part of the existing diagnostic process.

2: *Has the test been compared with a true gold standard?* The reference standard is an independent panel of senior clinicians, who were supplied with all clinical and laboratory data. I do not know whether this is a 'true gold standard', though it must be subject to some error. We should be concerned that the index tests, liquid culture and cell count, appear to part of the reference standard.

3: *Did this validation study include an appropriate spectrum of subjects?* Yes, in that they were derived from a consecutive series of A&E patients. However, is this how patients are usually acquired? Do they go to their GP and arrive via outpatients? Such patients may represent a different part of the spectrum.

4: *Has work-up bias been avoided?* No, it has not. The test results were included in the reference standard, as all laboratory data were available to the clinical panel.

5: *Has expectation bias been avoided?* No, as for 4.

6: *Was the test shown to be reproducible?* No mention of this.

7: *What are the features of the test as derived from this validation study?* Liquid culture has greater sensitivity than direct culture, but may also have lower specificity (no test is given, though data are given in the table and the difference is not significant). The cell count test appears in this small sample to be more sensitive and less specific than the direct culture, and to be less sensitive and less specific than the liquid medium. No significance tests are given for the cell count test. Can we assume that they were not significant? No comment is made on the combination of these tests.

8: *Were confidence intervals given?* Yes they were. But they are clearly wrong. These percentages cannot be greater than 100% in the population, so the confidence interval should not include values greater than 100%.

9: *Has a sensible 'normal range' been derived?* We do not know whether the cut-off for the cell count was pre-specified or derived from the data. If it was derived from the data, we do not know how. We do not know how changing the cut-off might alter the properties of the test.

10: *Has this test been placed in the context of other potential tests in the diagnostic sequence?* I think that the last paragraph of the paper does this.

QUADAS tool

1. *Was the spectrum of patients representative of the patients who will receive the test in practice?* Yes, in that they were derived from a consecutive series of A&E patients.

2. *Were selection criteria clearly described?* The source of the patients was described but no definition of bursitis was given.

3. *Is the reference standard likely to correctly classify the target condition?* The reference standard is an independent panel of senior clinicians, who were supplied with all clinical and

laboratory data. I do not know whether this is a 'true gold standard', though it must be subject to some error.

4. *Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?* Yes, in that all the data including the booking information was available to the panel.

5. *Did the whole sample or a random selection of the sample, receive verification using a reference standard?* Yes, they did.

6. *Did patients receive the same reference standard regardless of the index test result?* Yes, they did.

7. *Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard*)? Definitely not. The paper suggests that the test values were available to the panel as part of the information used to make their diagnosis.

8. *Was the execution of the index test described in sufficient detail to permit replication of the test?* It is difficult to say. It implies some knowledge on the part of the reader which I don't have, but to A&E personnel it may be sufficient. This is only a brief report and space is limited.

9. *Was the execution of the reference standard described in sufficient detail to permit its replication?* No. We do not know how many were on the panel nor their method of working.

10. *Were the index test results interpreted without knowledge of the results of the reference standard?* Yes, in that they came first. But the source of the cut-off for cell count is unknown.

11. *Were the reference standard results interpreted without knowledge of the results of the index test?* No, the panel appear to have had them.

12. *Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?* This is difficult to say from the paper. This question is difficult to answer unless you are familiar with the clinical area.

13. *Were uninterpretable/intermediate test results reported?* No, but there may have been none.

14. *Were withdrawals from the study explained?* None were mentioned. However, we do not know why 36 out of 54 consecutive patients were recruited. Did the others refuse? If so, why? If not, what happened to them? Note that we are told that the direct method correctly identified 10 of the 17 septic bursitis cases, 63%. But 10/17 = 59%. 10/16 = 63%. Did they lose one and not tell us about it?

What glaring statistical error is in this paper? The confidence intervals given are for percentages which cannot be greater than 100% in the population, but the confidence interval include values greater than 100%, e.g. sensitivity of liquid medium 100% (95% confidence interval 92% to 108%). The upper limit should not be greater than 100%. They have used a large sample method when the sample is far too small for it. (The correct 95% confidence interval is 80% to 100% using the direct Binomial probability method.) See Altman and Deeks (1999).

Reference

Altman DG, Deeks JJ. (1999). Sensitivity and specificity and their confidence intervals cannot exceed 100%. *BMJ* **318**, 193.