Guidelines and Criteria for Appraising Diagnostic Test Studies

Sackett *et al.* (1991) criteria

The best articles evaluating diagnostic tests will meet most or all of the following 8 criteria:

1. Was there an independent or 'blind' comparison with a 'gold standard' or diagnosis?

2. Was the setting for the study, as well as the filter through which study patients passed, adequately described?

3. Did the patient sample include an appropriate spectrum of mild and severe, treated and untreated disease, plus individuals with different but commonly confused disorders?

4. Were the tactics for carrying out the test described in sufficient detail to permit their exact replication?

5. Was the reproducibility of the test result (precision) and its interpretation (observer variation) determined?

6. Was the term 'normal' defined sensibly? (Gaussian, percentile, risk factor, culturally desirable, diagnostic, or therapeutic?)

7. If the test is advocated as part of a cluster or sequence of tests, was its contribution to the overall validity of the cluster or sequence determined?

8. Was the 'utility' of the test determined? (Were the patients really better off for it?)

Greenhalgh (1997) guidelines

1. Is this test potentially relevant to my practice?

2. Has the test been compared with a true gold standard?

3. Did this validation study include an appropriate spectrum of subjects?

4. Has workup bias been avoided? (I.e. was the test involved in the identification of refence standard cases?)

5. Has expectation bias been avoided? (I.e. were observations of test and reference standard blind to each other?)

6. Was the test shown to be reproducible?

7. What are the features of the test as derived from this validation study?

8. Were confidence intervals given?

9. Has a sensible 'normal range' been derived?

10. Has this test been placed in the context of other potential tests in the diagnostic sequence?

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. (1991) *Clinical Epidemiology: a Basic Science for Clinical Medicine*, Little Brown, Chicago

Greenhalgh, T. (1997) How to read a paper: Papers that report diagnostic or screening tests. *BMJ* **315**, 540-543.

QUADAS (**Qu**ality **A**ssessment of **D**iagnostic Accuracy **S**tudies) **Tool**

(Whiting *et al.*, 2003)

1. Was the spectrum of patients representative of the patients who will receive the test in practice?

2. Were selection criteria clearly described?

3. Is the reference standard likely to correctly classify the target condition?

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?

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

6. Did patients receive the same reference standard regardless of the index test result?

7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?

8. Was the execution of the index test described in sufficient detail to permit replication of the test?

9. Was the execution of the reference standard described in sufficient detail to permit its replication?

10. Were the index test results interpreted without knowledge of the results of the reference standard?

11. Were the reference standard results interpreted without knowledge of the results of the index test?

12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?

13. Were uninterpretable/ intermediate test results reported?

14. Were withdrawals from the study explained?

Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* **3**, 25, http://www.biomedcentral.com/1471-2288/3/25

The Bland criterion

Were the cut-off points for the test determined using data different from those used for evaluation?