How to evaluate a medical publication

Kenneth Alonso, MD, FACP

How to evaluate an article about a diagnostic test

- Has there been an independent (“blind”) comparison a criterion standard of diagnosis?
- Has the diagnostic test been evaluated in a patient sample that includes a appropriate spectrum (prevalence) of mild and severe, treated and untreated disease, as well as individuals with different but commonly confused disorders?
- Was the study setting as well as the filter through which the patients passed adequately described?
- Has the reproducibility of the test result (precision) and its interpretation (observer variation) been determined?

How to evaluate an article about a diagnostic test

- Has the term “normal” been defined sensibly as it applies to this test?
- If the test is advocated as part of a cluster or sequence of tests, had its individual contribution to the overall validity of the cluster and sequence been determined?
- Have the tactics for carrying out the test been described in sufficient detail to permit their exact replication?
- Has the utility of the test been determined?

- Simel, DL, Drummond,R, The rational clinical examination. Evidence based clinical diagnosis. McGraw-Hill (New York) for the American Medical Association (Chicago). 2009. p3

Observer variation

- For an item to be accurate, it must be reproducible (precision).
- Is it repeatable (intra-observer) or do two or more observers agree on the presence or absence of symptom or sign (inter-observer).
- How do we evaluate the reliability of agreement?
- How do we determine agreement is not by chance?
- Utilizing a 2x2 table, agreement between two observers would be, for true positives, [(a+b)x(a+c)]/(a+b+c+d)
- For true negatives, [(c+d)x(b+d)]/(a+b+c+d)

Observer variation

- The expected agreement is true positive agreement and true negative agreement divided by the number studied
- {[(a+b)x(a+c)]/(a+b+c+d) + [(c+d)x(b+d)] / (a+b+c+d)}/(a+b+c+d)
- Agreement beyond chance, κ, is:

  (observed agreement – expected agreement) / (1 – expected agreement)
- The higher the level of κ, the better the agreement.

Observer variation

- A value of -1.0 indicates complete disagreement, while a value of 1.0 indicates complete agreement; a value of 0.0 is chance.
- A κ of >0.6 indicates substantial agreement, reproducibility.
- Precision, however, does not mean accuracy.