BMJ

BMJ 2013;346:f2493 doi: 10.1136/bmj.f2493 (Published 19 April 2013)

ENDGAMES

STATISTICAL QUESTION

Receiver operating characteristic curves

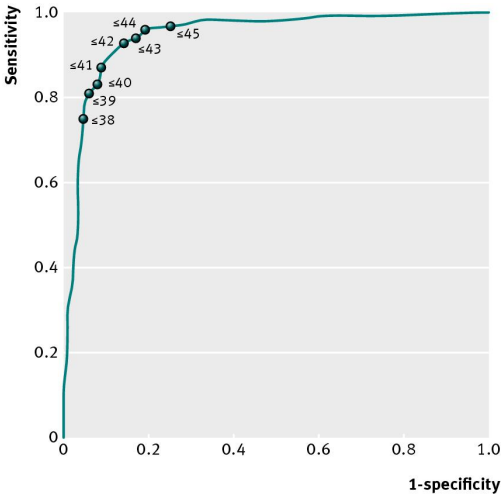
Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers evaluated the performance of a cognitive test called "test your memory" as a screening test for Alzheimer's disease. The screening test is designed to use minimal operator time and to be suitable for non-specialist use. It is self administered under medical supervision. The test has a minimum score of zero and a maximum score of 50; lower scores indicate greater cognitive impairment.[1]

The study was based in hospital outpatient departments. Participants included 94 patients diagnosed as having Alzheimer's disease. For each patient, three age matched healthy controls without Alzheimer's disease (n=282) were recruited from accompanying relatives. All patients and controls completed the screening test.

The optimal test score for discriminating between patients with Alzheimer's disease and controls was investigated. Each score from 50 down to zero was taken successively as the cut-off point between a "negative" and "positive" screening test result; all scores less than or equal to the cut-off score were considered positive and others were considered negative. For each cut-off score the sensitivity and specificity of the screening test was calculated, and these values were used to derive a receiver operating characteristic curve (figure). The area under the receiver operating characteristic curve was 0.95.



Receiver operating characteristic curve for "test your memory" scores differentiating patients with Alzheimer's disease (n=94) and age matched controls (n=282). Numbers on curve refer to a range of selected cut-off scores between "negative" and "positive" results

Which of the following statements, if any, are true?

a) For successive cut-off scores, as the sensitivity of the screening test decreases in value the specificity increases

b) The value of (1 minus specificity) represents the proportion of controls identified as positive (high risk) on screening

c) A screening test with an area under the curve equal to one half (0.5) would discriminate perfectly between patients and controls

Answers

Statements *a* and *b* are true, whereas *c* is false.

The performance of the test your memory cognitive test as a screening tool for Alzheimer's disease was investigated. In

p.sedgwick@sgul.ac.uk

particular, the aim was to establish the optimal cut-off score between a negative (low risk) and positive (high risk) result. By definition the screening test will never be 100% accurate—some patients with Alzheimer's disease will be classified as negative (false negatives) and some controls as positive (false positives). The optimal cut-off score would be the one that best discriminates patients with Alzheimer's disease and controls—that is, identifies the greatest proportion of patients with Alzheimer's disease as positive and the greatest proportion of controls as negative. The test could subsequently be used to screen patients in memory clinics; those patients identified as positive on the basis of the optimal cut-off score would be referred for further diagnostic clinical assessment because they would be considered to be at high risk of Alzheimer's disease. Patients identified as negative would be considered to be at low risk and would not need further assessment. The screening test would therefore prevent patients from undergoing unnecessary clinical assessment.

The minimum score on the test your memory tool is zero and the maximum is 50; lower scores indicate greater cognitive impairment. Each score from 50 down to zero was taken successively as the cut off between a negative and positive screening test result; all scores less than or equal to the cut-off score were considered positive results and all others were considered negative. For each cut-off score the sensitivity and specificity values were derived. The sensitivity and specificity values for a series of cut-off scores between 38 and 45 were presented by the researchers (table⇓) and indicated on the receiver operating characteristic curve.

Sensitivity is the proportion of patients with Alzheimer's disease correctly identified by the screening test with a positive result. Specificity is the proportion of control participants correctly identified by the screening test with a negative result.

As the cut-off score decreases in value from 50 down to zero, a smaller proportion of patients with Alzheimer's disease would be correctly identified as positive (high risk), resulting in a smaller sensitivity. At the same time, as the cut-off point decreases in value, the proportion of control participants correctly identified as negative (low risk) becomes larger, resulting in a larger specificity (*a* is true). Therefore, the optimal cut-off score is a trade-off between sensitivity and specificity based on the implications of correctly identifying patients with Alzheimer's disease and incorrectly identifying controls as positive (false positives).

The receiver operating characteristic curve was constructed as the plot of sensitivity against (1 minus specificity) for each cut-off score. The value of (1 minus specificity) is the proportion of controls incorrectly identified by the screening test with a positive result (*b* is true), and therefore as false positives. The value of (1 minus specificity) is referred to as the false positive rate. The receiver operating characteristic curve therefore provides a graphical representation of the proportion of patients with Alzheimer's disease correctly identified as positive against the proportion of control patients incorrectly identified as positive for each cut-off score. Intuitively, sensitivity should be as large as possible and the false positive rate as small as possible; a small false positive rate equates to a large specificity.

If a cut-off score produced a perfect screening test—if it predicted patients with diagnosed Alzheimer's disease and controls with 100% accuracy—both sensitivity and specificity would equal 1.0 and the false positive rate would equal zero. In such a case, the receiver operating characteristic curve would pass through the top left hand corner of the figure. In effect, the curve would start at the origin, go vertically up the y axis to a sensitivity of 1.0 and then horizontally across to a false positive rate of 1.0. Therefore, the closer the receiver operating characteristic curve is to the upper left corner, the higher the overall accuracy of the screening test across all potential cut-off scores. For that reason, the cut-off score closest to the top left hand corner is typically chosen as the optimal cut-off score. However, this does assume that the implications of a false positive and a false negative result are similar. The researchers commented that the optimal cut-off score for screening for Alzheimer's disease was 42 or less (sensitivity 93%, specificity 86%).

The overall accuracy of the screening test can be quantified by the area beneath the receiver operating characteristic curve, typically expressed as a proportion. Therefore, the closer the area under the curve is to one—that is, the closer the curve is to the top left hand corner—the better the screening test is at discriminating between patients with diagnosed Alzheimer's disease and controls.

A diagonal line is shown on the receiver operating characteristic curve from (0,0), where sensitivity and false positive rate both equal zero, to (1,1), where sensitivity and false positive rate both equal one. Points on this line represent a sensitivity and false positive rate of equal value at each cut off. This line would indicate a screening test that had no ability to discriminate between patients with Alzheimer's disease and controls; it identifies exactly the same proportion of each patient group as high risk at a particular cut-off point. Any such screening test would have a receiver operating characteristic curve with an area under the curve of 0.5 (*c* is false).

The receiver operating characteristic curve is probably most useful when comparing the ability of two or more screening tests to discriminate patients with and without a particular diagnosis. Although it may be argued that a receiver operating characteristic curve has little value for a single screening test, it is a useful graphical representation if there are many possible cut-off points.

Competing interests: None declared.

1 Brown J, Pengas G, Dawson K, Brown LA, Clatworthy P. Self administered cognitive screening test (TYM) for detection of Alzheimer's disease: cross sectional study. *BMJ* 2009;338:b2030.

Cite this as: *BMJ* 2013;346:f2493

Table

Table 1| Sensitivity and specificity for screening of Alzheimer's disease with the "test your memory" cognitive test in 94 patients with Alzheimer's disease and 282 age matched controls

Cut-off score	Sensitivity (%)	Specificity (%)
≤38	75	96
≤39	81	95
≤40	83	92
≤41	87	91
≤42	93	86
≤43	94	83
≤44	96	81
≤45	97	75