INFORMATION POINT:

Receiver Operating Characteristic (ROC) curves

A receiver operating characteristic (ROC) curve allows us to explore the relationship between the sensitivity and specificity of a clinical test for a variety of different cut points, thus allowing the determination of an optimal cut point.

In order to determine the presence or absence of a disease, we often have to carry out a test which provides a result on a continuous measure, for example a blood glucose measurement or a score on the Morse fall scale. From this we need to decide if the disease is present or absent, so a cut point is selected. To one side of this cut point, say above, we claim the disease is present and below this cut point we claim the disease is absent.

Using any test we will make diagnostic errors. Commonly used measures of the performance of a test are the sensitivity and specificity. Sensitivity is the probability that we diagnose the disease when it is actually present (the true positive rate). Specificity is the probability that we identify that the disease is absent when it is truly absent (the true negative rate). Ideally we want both sensitivity and specificity to be one. Unfortunately, changing the cut point to try and increase either sensitivity or specificity will usually result in a decrease in the other measure.

The ROC curve is a graphical technique to try and establish the optimal cut point and is a procedure derived from the early days of radar and sonar detection used in the Second World War, hence the name receiver-operating characteristic. In order to construct a ROC curve we need to calculate the sensitivity and specificity of the test for each possible cut point value. So, for example, suppose the test scale permits scores varying between 10 and 20 and we have measured this score for 500 people with known disease status. Then we could explore how these 500 people would be diagnosed if the cut point was 10 and we could calculate the sensitivity and specificity. Then we could repeat this exercise using a cut point of 11 and so on.

To make the ROC graph, the X-axis is 1 minus the specificity (the false positive rate) and the Y-axis is the sensitivity (the true positive rate). We draw a diagonal line on the graph from (0,0) in the lower left hand corner to (1,1) in the upper right hand corner. This line reflects the characteristics of a test with no discriminating power. The closer the graph gets to the upper left hand corner (0,1), the better the test is at discriminating between cases and noncases. An index of the goodness of the test is the area under the curve – a perfect test has area 1.0, whilst a nondiscriminating test (one which falls on the diagonal) has area 0.5. Streiner & Norman (1995) discuss this in more detail and provide examples.

The ROC is most helpful when comparing two or more methods. Altman (1998) argues that for a single test the ROC does not really add anything to a table of sensitivity and specificity values for different cut points. However, the graph might be a preferable presentation if there are many possible cut points. Altman (1998) also notes that if the cost of a false positive and a false negative are not equal then the best cut point is not necessarily the point nearest the top left hand corner.

Further reading

Altman D.G. (1998) Confidence intervals for the number needed to treat. ***British Medical Journal*** **317**, 1309–1312.

Streiner D.L. & Norman G.R. (1995) *Health Measurement Scales*, 2nd edn. Oxford University Press, Oxford.

NICHOLA CRICHTON