Evaluation of New Medical Devices

To the Editor—Anesthesiologists still lack consistently reliable means to monitor depth of anesthesia. Among the physiologic manifestations purport to vary with anesthetic depth is spontaneous lower esophageal contractility. A device is now available to detect spontaneous lower esophageal contractions (SLEC) by an esophageal probe.1 Watcha and White2 attempted to validate the clinical usefulness of this device by determining if patient movement following skin incision could be predicted by the frequency of SLEC immediately prior to incision; the study group was children anesthetized with halothane. No correlation between SLEC rate and patient movement was observed.8

The use of such a device to sort patients into categories (movement vs. no movement) can be viewed as a diagnostic test. An extensive repertoire of decision analytic and statistical tools can provide additional insights into the usefulness of a diagnostic test.3 One of the merits of their report is the inclusion of original data (Watcha and White,4 fig. 2) which allows construction of some additional statistics.

A diagnostic test is useful if it correctly predicts that sufficiently anesthetized patients will not move and that insufficiently anesthetized patients will move. These two criteria can be defined algebraically and have been labelled to test sensitivity (true-positive rate) and specificity (true-negative rate), respectively; sensitivity and specificity are expressed as decimal fractions ranging from 0–1.

Like many tests, the SLEC rate is not an either/or variable; the average SLEC rate per min is a continuous variable. Watcha and White reported a SLEC rate ranging from 0 to >4 counts per min immediately prior to incision. To calculate specificity and sensitivity values from the data of Watcha and White requires the choice of a diagnostic threshold. However, using any particular value of the SLEC rate as a cut point for classifying patients into sufficiently and insufficiently anesthetized groups would be arbitrary and inadequately describes the diagnostic properties of the test.

A Receiver Operating Characteristic (ROC) curve graphically shows the usefulness of a diagnostic test by plotting all possible cutpoints present in the data.4 The ROC curve is a plot of the true-positive rate (sensitivity) versus the false-positive rate (1–specificity) for all classifications. An accurate test will have a curve that quickly approaches a true-positive rate of 1. A useless test will lie close to the line of identity (true-positive rate = false-positive rate). The area under the ROC curve is the probability of a correct diagnosis. If the ROC curve is close to the 45° line, the area under the curve is about 0.5; such a test is no better than flipping a coin.

I have constructed an ROC curve (fig. 1) from the data of Watcha and White. The area 8 ± the curve (nonparametric estimation) is 0.4880 ± 0.0892. The use of SLEC is clearly demonstrated to be of no value in predicting movement in this patient group, regardless of the choice of SLEC frequency used to classify patients.

The inventors and proponents of new devices and tests often enthusiastically propose their adoption and use by the community of anesthesiologists. One of the purposes of anesthesia research is to perform a critical and sceptical assessment of new devices; this technology assessment should help in the rational allocation of resources. I would urge that reports of diagnostic tests and validations of new monitors include ROC curves and other decision analytic statistics as appropriate. Such statistical methods can provide graphically clear demonstrations of their usefulness or uselessness. This can provide additional help to the practicing anesthesiologist in deciding to invest time and money in new technology.

NATHAN LEON PACE, M.D.
Professor of Anesthesiology
Adjunct Professor of Bioengineering
Department of Anesthesiology
The University of Utah
50 North Medical Drive
Salt Lake City, Utah 84132

REFERENCES


(Accepted for publication February 3, 1990.)