MakroCare
New Possibilities, Everyday

WHITE PAPER

IVDR – When Diagnostic Accuracy Gold Standard Procedure is not Available

Diagnostic tests are very common and used by health professionals for detecting the disease condition in human beings. Generally, most of the tests are giving accurate results in comparison with a gold standard method or a well-established standard procedure. In this situation, all standard diagnostic accuracy estimates such as sensitivity, specificity, false positive rate, false negative rate, positive predictive value, positive predictive value adjusted for known prevalence of the disease, negative predictive value, negative predictive value adjusted for known prevalence of the disease, false omission rate, false discovery rate, prevalence of the disease based on the data, proportion correctly classified, proportion incorrectly classified, confidence intervals for various rates, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio. After all, the diagnostic accuracy measures the ability of a test to detect a condition when it is present and detect the absence of a condition when it is absent.

However, certain situations, the gold standard or a well-established standard procedure may not exist for diagnosis of a particular disease. In such situations, an imperfect reference standard will be used and hence the above mentioned diagnostic accuracy estimates cannot be computed since the subjects' disease condition is unknown. The objective of this paper is to review the statistical methods for the estimation of diagnostic accuracy when gold standard procedure is not available.

POSITIVE AND NEGATIVE PERCENT AGREEMENT

In the absence of gold standard, Positive Percent Agreement and Negative Percent Agreement can be estimated in the place of sensitivity and specificity. Positive percent agreement is the proportion of individuals with the target condition by the imperfect reference standard who test positive. Similarly, negative percent agreement is the proportion of the individuals free of the target condition by imperfect reference standard who test negative.

OVERALL PERCENT AGREEMENT

In situations where there is no gold standard procedure is available, it has been opined that the concepts of sensitivity, specificity, etc., should not be used. Instead, the 2 × 2 table of results comparing the new test with the imperfect reference test and measures of overall percent agreement between tests may be better used with discussions and investigations. Overall percent agreement is the sum of true positives plus true negatives divided by the total number of individuals tested.

COHEN'S KAPPA

Suppose two observers or physicians are asked to rate the same subjects for the presence or absence of a disease. Cohen's kappa coefficient can be used to assess the agreement between the two raters. However an important assumption underlying the use of the kappa coefficient is that errors associated with the two sets of ratings are independent. This requires the subjects to be independent and Rater 1's ratings to be independent of Rater 2's. The kappa coefficient, therefore, is not appropriate for a situation in which one observer is required to either confirm or disconfirm a known previous rating from another observer. For diagnostic tests, this concept may be used to estimate the agreement between two diagnostic procedures.

The above mentioned diagnostic accuracy estimates can be computed only for the dichotomous results with outcomes of either positive or negative test results. When a new test gives continuous values against a gold standard procedure with dichotomous results, ROC method can be used to identify the optimum cutoff point for the classification of subjects into positives or negatives based on new test. In the absence of gold standard, ROC method is not the appropriate method for identification of optimum cutoff point for classification of subjects into positives or negatives.

CONCLUSION

To summarize, when gold standard exists, then all standard diagnostic accuracy estimates can be computed confidently. On the other hand, if a new test is being compared to an imperfect reference test, then the above mentioned diagnostic accuracy estimates such as positive percent agreement and negative percent agreement etc. along with 95% confidence intervals may be computed and reported with special mention about the imperfect reference test.

*****