

Right choice of a method for determination of cut-off values: A statistical tool for a diagnostic test

Balkishan Sharma¹, Ravikant Jain²

¹Associate Professor (Biostatistics), Department of Community Medicine, Sri Aurobindo Medical College and P. G. Institute, Indore (M. P.), India, ²Professor, Department of Orthopedics, Sri Aurobindo Medical College and P. G. Institute, Indore (M. P.), India

Submitted: 06-12-2013

Revised: 07-12-2013

Published: 10-03-2014

ABSTRACT

Objective: The clinical diagnostic tests are generally used to identify the presence of a disease. The cutoff value of a diagnostic test should be chosen to maximize the advantage that accrues from testing a population of human and others. When a diagnostic test is to be used in a clinical condition, there may be an opportunity to improve the test by changing the cutoff value. To enhance the accuracy of diagnosis is to develop new tests by using a proper statistical technique with optimum sensitivity and specificity. **Method:** Mean \pm 2SD method, Logistic Regression Analysis, Receivers Operating Characteristics (ROC) curve analysis and Discriminant Analysis (DA) have been discussed with their respective applications. **Results:** The study highlighted some important methods to determine the cutoff points for a diagnostic test. The traditional method is to identify the cut-off values is Mean \pm 2SD method. Logistic Regression Analysis, Receivers Operating Characteristics (ROC) curve analysis and Discriminant Analysis (DA) have been proved to be beneficial statistical tools for determination of cut-off points. **Conclusion:** There may be an opportunity to improve the test by changing the cut-off value with the help of a correctly identified statistical technique in a clinical condition when a diagnostic test is to be used. The traditional method is to identify the cut-off values is Mean \pm 2SD method. It was evidenced in certain conditions that logistic regression is found to be a good predictor and the validity of the same can be confirmed by identifying the area under the ROC curve.

Key words: Logistic Regression, Confidence Interval, Cut-off value, Discriminant analysis, ROC curve, Sensitivity and Specificity

Access this article online

Website:

<http://nepjol.info/index.php/AJMS>

INTRODUCTION

Clinical classification of persons involves categorizations of diseased and non-diseased (positive or negative) may be reported utilizing some statistical techniques. The clinical diagnostic tests are generally used to identify the presence of a disease. Therefore, such tests with optimal cut-off values may correctly identify all patients with and without the disease. However, a perfect test is never positive in a patient without disease and is never negative in a patient who is in fact diseased. Most clinical tests fall short of this ideal because of incorrect optimal cut-off values predicted from a specified test.

When a diagnostic test is to be used in a clinical condition, there may be an opportunity to improve the test by changing the

cutoff value. The test's cutoff value may require re-assessment of the determination with respect to clinical, demographical and biological conditions which may be different from the one for which the test was developed. The methodology of cut-off value has been adapted to several clinical areas heavily dependent on screening and diagnostic tests^{1,2} such as in epidemiology.³ However, some statistically valid techniques for suggesting and determining more accurate and valid cutoff value based on analytical and empirical evidences for the identification of subjects as diseased or non-diseased may also be used in addition to the clinical experiences. Logistic regression, ROC curves and DA has been proved to be beneficial statistical tools for such situation.

The aim of the study is to identify about an appropriate selection of the method use to predict the cutoff points for

Address for Correspondence:

Dr. Balkishan Sharma, Department of Community Medicine, Sri Aurobindo Medical College and P. G. Institute, Indore (M. P.), India.

E-mail: bksnew@rediffmail.com, bksnew@gmail.com; **Mobile:** 99260-49820; **Fax:** (0731) 4231010.

© Copyright AJMS

a diagnostic test. The various popular statistical techniques for the identification of the cut-off values such as Mean \pm 2SD (95% CI for mean), Logistic regression, ROC curves and DA has been discussed with their respective applications.

METHODS AND RESULTS

The ability to make a diagnosis or screen for a condition depends both on the discriminatory value of the test and on the prevalence of the disease in the population of interest.⁴ A cut-off value for a diagnostic test may be generated by using any of the suggested methods. The chosen of an appropriate method for a diagnostic test may be providing optimal cut-off points at optimal sensitivity with specificity.

Mean \pm 2SD

The conventional method to determine a cut-off is the 95% CI of mean, a crude measure for observing cut-off values. An interval, *mean \pm 2SD* obtained by subtracting 2 \times SD from mean and by adding 2 \times SD predicts that the chance of a test value coming outside this interval will be less than 5%. The lower limit, *mean-2SD* of this interval may be considered as cutoff point. If a subject's test value observes less than this cutoff then may be considered as non-diseased or normal and if value measures greater than cutoff value then considered diseased or non-normal. Alternatively, the upper limit, *mean+2SD* of the 95% confidence interval also may be considered as cutoff value. Therefore, a feasible and suitable approach may be selected to control the lowery of sensitivity and specificity, depending on the amount of loss of clarity.

For illustration, elevated markers of inflammation, in particular CRP, are associated with an increased risk of future cardiovascular events in subjects and commonly used coronary heart disease (CHD) marker.⁵ Thakur et al. (2011) investigated the predictive value of Hs-CRP (high sensitivity C-reactive protein) in CHD. The Mean levels of Hs-CRP in healthy controls and CHD subjects were 0.93 \pm 0.35 mg/L and 1.7 \pm 0.75 mg/L respectively. If, we consider the mean level in CHD subjects and when a large sample had drawn randomly from a specific population it was noted that the 95% confidence interval (CI) for Hs-CRP will be in range of 0.2 mg/L to 3.2 mg/L (i.e. 1.7–2 \times 0.75 to 1.7+2 \times 0.75). Thus, the condition can be summarized as the chance of the value of Hs-CRP in such situation lying outside the interval 0.2 mg/L to 3.2 mg/L is less than 5%. It can be justified in simple words that there is a 95% chance that Hs-CRP test value for such CHD subjects may lie in ranges of 0.2 mg/L to 3.2 mg/L. Notably, the higher limit of this interval for

Hs-CRP protein level i.e. 3.2 mg/L may be taken as a cutoff point.

In addition, it is a method of obtaining cutoff values for a diagnostic test under certain varying conditions subjected to chance variation involves some drawbacks.

ROC Curve

The Receiver Operating Characteristics (ROC) curves provide a comprehensive and visually attractive way to summarize the accuracy of predictions. The ROC curve shows the tradeoff between sensitivity and specificity and is a better method to detect the performance of a developed test which classifies subjects into two categories such as diseased and non-diseased. The ROC curve is a graph of sensitivity (y-axis) versus 1–specificity (x-axis).

ROC curve⁶ may be used to judge how well the test performs. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test. The accuracy of the test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test while an area of 0.5 represents a worthless test.

First known application of ROC curve analysis⁷ took place during Second World War when it was employed for the processing of radar signals.⁸ Osada et al. (2004) used ROC analysis to assess the ability of MRI to predict fetal pulmonary hypoplasia. Time-dependent receiver operating characteristic analysis was used to evaluate the prognostic ability for incident of Alzheimer disease over various durations of follow-up.⁹

A ROC curve analysis was carried out on the basis of data reported by¹⁰ Wong et al. (2002). There were 181 patients 67 of whom were found to have cancer by the gold standard. PET produces a measure called standardized uptake value (SUV) which is an indicator of how likely the part of the body under consideration has cancer. The following Table produced the accuracy of SUV in diagnosing cancer for various thresholds.

Accuracy of SUV	Various thresholds			
	1	3	5	7
Sensitivity	97%	93%	61%	37%
Specificity	48%	65%	88%	97%

This table makes clear the wide range of sensitivity and specificity that can be obtained by varying the threshold. It also makes clear the inverse relationship between the two measures as one increases the other decreases and

vice versa. Therefore, accuracy of SUV value at 3 may be thought of optimally better cutoff for differentiating patients. However, in general high sensitivity may be more desirable than high specificity because it consist of low number of false negative cases while high specificity leads to low number of false positive cases.

Moreover, seriousness of loss of information, test conditions and requirement the optimal value of the sensitivity and specificity is decided and the corresponding test value may be used as cutoff value for classification of subjects.

Logistic Regression

Logistic regression is useful to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable. Logistic regression is also used as classifier of a case and a control and the cutoff value can be adjusted of the predicted probability to be used in classification.

For illustration, Lajoie Y and Gallagher SP (2004)¹¹ carried out a study in which forward Wald logistic regression was used to determine a cut-off score for the Berg scale, the ABC scale and simple reaction time in order to identify those individuals at risk of falling. Results from the logistic analysis revealed that three variables were associated with fall status with 91% sensitivity and 97% specificity. The results were further assessed and successfully classified those at risk of falling.

In a comparative study Cepeda SM et al. (2003)¹² found that in the logistic regression, the bias decreased as the number of events per confounder increased. The average of individual odds ratios and the total cohort odds ratio approximate each other when the incidence of the disease is low, all subjects have low risks,¹³ and both odds ratios are the same when there is no association between the exposure and the outcome.¹⁴

Finally, the logistic regression modeling has proposed as the preferred statistical method to obtain a post-test probability of disease when results from multiple tests are available.

Discriminant Function Analysis

Discriminant analysis or DA is a commonly used statistical tool which is used to classify cases into the values of a categorical dependent, usually a dichotomy. DA is a popular tool in solving classification problem. A function is generated from a sample of known positive or diseased and negative or non-diseased cases. Thereafter, the pre-defined function is used to identify the cases as positive

or negative with observed diagnostic test values. The discriminant scores with respect to observed diagnostic test values for each subject is recorded to classify them as positive or negative. The independent variables are treated as predictors while the dependent variable, also called the grouping variable.

A discriminant function is also called a canonical root. The eigenvalue of each discriminant function reflects the ratio of importance of the dimensions which classify cases of the dependent variable. Singh G. (2007)¹⁵ concluded that a sample of known negative (normal) and positive (suffering with the particular disease) cases be chosen. The diagnostic test (for which cutoff is to be determined) may now be administered and the test value observed may be recorded.

DA is an earlier alternative to logistic regression, which is now frequently used in place of DA as it usually involves fewer violations of assumptions (independent variables needn't be normally distributed, linearly related, or have equal within-group variances), is robust, handles categorical as well as continuous variables, and has coefficients which may find easier to interpret.¹⁶⁻¹⁸

Logistic and discriminant analyses will usually yield the same conclusions, except in the case when there are independents which result in predictions very close to 0 and 1 in logistic analysis reported by Press SJ and Wilson S (1978).¹⁹ Discriminant function analysis method may be used when more than one diagnostic test administered at a time to determine cutoff scores.

Finally, discriminant function analysis method may be a suitable technique when cutoff score is to be observed. Henceforth, chosen of a correct statistical technique, to predict an optimal cutoff value, improves the strength of the diagnostic test. Moreover, selection of a right method enables the investigator, the enhancement in the accuracy of classifying subjects as positive or negative (with or without disease).

DISCUSSION

This article offers an introduction to the understanding and use of an appropriate method to observe the optimal cut off points for a diagnostic test. It is worth advisable that a cut-off value for a diagnostic test may be determined by using any of the above discussed method since the selection of a right choice for a diagnostic test may be providing an optimal cut-off which has optimal level sensitivity with specificity.

This classical method (95% CI for mean) subjected to several serious objections. To determine the cut off

value is to calculate two standard deviations from the difference between mean values of two groups under the independence assumption.²⁰ However, this method is not appropriate when the assumption is violated. The classical method may be subjected to chance of measuring some false negative cases when a sample of adequate size of diseased (positive) or cases is being considered followed by administration of diagnostic test which further lowers the sensitivity of the test diagnostic. Similarly, specificity of the test diagnostic is affected when non-diseased (negative cases) is being considered which is subjected to chance of measuring some false positive cases.

Singh G (2007)¹⁵ detected the drawback of the 95% CI ($mean \pm 2SD$) method is that upper limit found for negative (healthy) subjects may not coincide with the lower limit found for positive (diseased) cases. Sometimes there may be a gap between the two and sometimes they overlap.

Sally E Ridge and Andrew L Vizard (1993)²¹ reported that ideally, the cutoff value of a test should be chosen to maximize the benefit that accrues from testing a population. This is equally true in human and veterinary medicine. Invariably, cutoff values for diagnostic tests have been determined by arbitrary methods that fail to consider these issues. A disadvantage is the impossibility of weighing the true positive and false positive rate separately.

The study conducted by Afina S et al. (2003)²² highlighted that the diagnostic odds ratio is closely linked to existing indicators which facilitates formal meta-analysis of studies on diagnostic test performance derived from logistic models and allowed for the inclusion of additional variables to correct for heterogeneity. Results by Lajoie Y and Gallagher SP (2004)¹¹ showed the importance of the cut-off value for a diagnostic test and confirmed that study would seem rather valuable as an assessment tool for health care professionals. The results successfully classified those at risk of falling and observed the sensitivity and specificity of correctly classifying fallers and non-fallers.

In a recent study Brian C Lau et al. (2012)²³ had used ROC curve analysis to determine cutoff scores in neurocognitive and Post-Concussion Symptom Scale symptom cluster scores and obtained eighty-percent sensitivity which indicated that the corresponding cutoff correctly identifies 80% of concussed athletes requiring protracted recovery. Pineda DA et al. (2007)²⁴ also used ROC curve analysis which showed modest sensitivity and low specificity, demonstrating that an important proportion of the variance in test scores was overlapping. ROC curve may not be a good measure for comparing tests¹⁵ while Press S J and Wilson S (1978)¹⁹ conducted a research for the superiority of logistic regression for situations where the assumptions

of multivariate normality are not met (ex., when dummy variables are used), though discriminant analysis is held to be better when assumptions are met.

Using the whole range of results rather than a single cut-off point for discriminant analysis did not alter the relative performance of the screening tests. Computerized neurocognitive testing has been proven sensitive and specific to diagnosing a concussion.²³ When made a use of discriminant analysis which confirmed the good performance of the rapid interviews and it also identified glutamate dehydrogenase (GDH) as the best of the laboratory tests and of comparable efficacy to the rapid interview for the group of excessive drinkers.²⁵

Lastly, it is suggested that a cut-off value for a diagnostic test may be sometime unacceptable due to unpredictability of disease for various circumstances on certain grounds. A value generated by analyzing the data using any of the above methods may be not acceptable clinically because it may be dealt as prerogative of a statistician.

SUMMARY OF THE FINDINGS

There may be an opportunity to improve the test by changing the cut-off value with the help of a correctly identified statistical technique in a clinical condition when a diagnostic test is to be used. The traditional method is to identify the cut-off values is $Mean \pm 2SD$ method. It was evidenced in certain conditions that logistic regression is found to be a good predictor and the validity of the same can be confirmed by identifying the area under the ROC curve.

IMPLICATIONS OF THE STUDY

Some statistically valid techniques for suggesting and determining more accurate and valid cutoff value based on analytical and empirical evidences for the identification of subjects as diseased or non-diseased may also be used in addition to the clinical experiences. However, Logistic regression analysis, ROC curves analysis and DA has been proved to be beneficial statistical tools for such situation.

REFERENCES

1. Lloyd CJ. Using smooth receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association* 1998;9:1356-1364.
2. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. 2003; Oxford, UK: Oxford University Press.
3. Shapiro DE. The interpretation of diagnostic tests. *Statistical Methods in Medical Research* 1999;8:113-134.
4. Abdul G Lalkhen and Anthony McCluskey. *Clinical tests:*

- Sensitivity and Specificity Continuing Education in Anaesthesia. *Critical Care & Pain* 2008;8(6):221-223.
5. Thakur S, Gupta S, Parchwani H, Shah V and Yadav V. Hs-CRP - A potential marker for coronary heart disease. *Indian Journal of Fundamental and Applied Life Sciences* 2011;1(1):1-4.
 6. Biao Z. A semiparametric hypothesis testing procedure for the ROC curve area under a density ratio model. *Computational statistics and data analysis* 2006;50(7):1855-1876.
 7. Matjaz Majnik and Zoran Bosnic. ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis* 2013;17: 531-558.
 8. Osada H, Kaku K, Masuda K, Iitsuka Y, Seki K and Sekiya S. Quantitative and qualitative evaluations of fetal lung with MR imaging. *Radiology* 2004;231:887-892.
 9. Carol A Derby, Leah C Burns, Cuiling Wang, Mindy J. Katz, Molly E. Zimmerman, Gilbert L'Italien, et al. Screening for predementia AD, Time-dependent operating characteristics of episodic memory tests. *Neurology* 2013;80(14):1307-1314.
 10. Wong RJ, Lin DT, Schoder H, Patel SG, Gonen M, Wolden S, et al. Diagnostic and prognostic value of 18F-fluorodeoxyglucose positron emission tomography for recurrent head and neck squamous cell carcinoma. *Journal of Clinical Oncology* 2002;20(20):4199-4208.
 11. Lajoie Y and Gallagher SP. Predicting falls within the elderly community: comparison of postural sway, reaction time, the Berg balance scale and the Activities-specific Balance Confidence (ABC) scale for comparing fallers and non-fallers. *Archives of Gerontology and Geriatrics* 2004;38:11-26.
 12. Cepeda SM, Boston R, Farrar JT and Brian L. Strom. Comparison of Logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American Journal of Epidemiology* 2003;158(3):280-287.
 13. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* 1987;125:761-768.
 14. Gail MH, Wieand S and Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984;71:431-444.
 15. Singh G. Determination of cutoff score for a diagnostic test. *The Internet Journal of Laboratory Medicine* 2007;2(1):1-4.
 16. McLachlan and Geoffrey J. *Discriminant analysis and statistical pattern recognition*. NY: Wiley-Interscience. 2004 (Wiley Series in Probability and Statistics).
 17. Klecka William R. *Discriminant analysis. Quantitative Applications in the Social Sciences Series, No. 19*. 1980; Thousand Oaks, CA: Sage Publications.
 18. Huberty and Carl J. *Applied Discriminant Analysis*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics Section. John Wiley & Sons. 1994; 496.
 19. Press SJ and Wilson S. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* 1978;73:699-705.
 20. Rha SY, Yang WI, Gong SJ, Kim J J, Kim BS and Chung HC. Correlation of Tissue and Blood Plasminogen Activation System in Breast Cancer. *Cancer Letters* 2000;150:137-145.
 21. Sally E. Ridge and Andrew L. Vizard. Determination of the Optimal Cutoff Value for a Serological Assay: an Example Using the John's Absorbed EIA. *Journal of clinical microbiology* 1993;31(5):1256-1261.
 22. Afina S Glasa, Jeroen G Lijmer, Martin H Prins, Gouke J Bonse and Bossuyt MMP. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology* 2003;56:1129-1135.
 23. Brian C Lau, Michael W Collins and Mark R Lovell. Sensitivity and specificity of subacute computerized neurocognitive testing and symptom evaluation in predicting outcomes after sports-related concussion. *The American Journal of Sports Medicine* 2010;20(10):1-8.
 24. Pineda DA, Puerta IC, Aguirre DC, Garcia-Barrera MA and Kamphaus RW. The role of neuropsychologic tests in the diagnosis of attention deficit hyperactivity disorder. *Pediatric Neurology* 2007;36:373-381.
 25. Bernadt MW, Mumford J and Murray RM. A Discriminant-Function Analysis of Screening Tests for Excessive Drinking and Alcoholism. *Journal of Studies on Alcohol and Drug* 1984;45(1):81-86.

Authors Contribution:

BS – Conceptualized study, literature search, statistically analyzed and interpreted, prepared first draft of manuscript and critical revision of the manuscript;
RJ – Concept and Design of the study, review the literature, manuscript preparation and critical revision of the manuscript.

Source of Support: Nil, **Conflict of Interest:** None declared.