Austin Biometrics and Biostatistics

Open Access

Austin Publishing Group

Mini Review

Nonparametric Approaches to Comparing the Accuracy of Diagnostic Tests with Multiple Readers

Eunhee Kim*
Department of Biostatistics and Center for Statistical Sciences, Brown University, USA

***Corresponding author:** Eunhee Kim, Department of Biostatistics and Center for Statistical Sciences, Providence, Rhode Island, 02912 USA

Received: September 08, 2014; **Accepted:** October 10, 2014; **Published:** October 13, 2014

Abstract

In diagnostic imaging studies, the test results often depend on the subjective interpretation of the reader. Because of variability in readers' accuracy, studies evaluating diagnostic tests usually involve multiple readers. Receiver Operating Characteristic (ROC) analysis has been a popular method for evaluating the performance of diagnostic imaging modalities. In this mini-review, I introduce current literature on nonparametric methods to compare the accuracy of diagnostic tests with multiple readers using ROC analysis. Nonparametric approaches do not require distributional assumptions for the test results or the ROC curve, making them attractive for use when the total sample size/number of readers is small or when distributional assumptions may be problematic.

Keywords: Diagnostic test; Multi-reader; Multi-test design; Nonparametric methods; Receiver operating characteristic curve

Introduction

Early and accurate diagnosis of disease is vital for the clinical management of patients. For example, imaging modalities such as computed tomography and magnetic resonance imaging have become important tools for the diagnosis of various diseases because of their non-invasive nature. Given the recent advances in medical imaging technologies, numerous studies have been conducted to compare the performance of currently available diagnostic tests. In radiological studies, the accuracy of such tests often depends on the subjective interpretation of readers (or radiologists). Because of variability in readers' accuracy, studies comparing two or more imaging modalities usually involve multiple readers; these studies are often designed so that multiple readers interpret all test results from a sample of patients who undergo multiple diagnostic tests, referred to as the multi-reader, multi-test design. This design is efficient for comparing diagnostic tests because it requires a smaller patient population than other study designs [1]. For example, Table 1 presents a data structure in a multi-reader, multi-test design, in which each of N patients experiences two diagnostic tests that are interpreted by J different readers.

Numerous statistical methods have been developed to evaluate the performance of diagnostic tests in the framework of Receiver Operating Characteristic (ROC) analysis [2-4]. The ROC curve is a standard tool used to compare diagnostic tests when test results are continuous or ordinal. In an ROC curve, the true positive rate is plotted as a function of the false positive rate across all possible cut-points. The area under the ROC curve (AUC) is a commonly used summary measure of diagnostic accuracy, in which AUC values close to 1 indicate that a test has high diagnostic accuracy. The partial area under the ROC curve (pAUC) is another summary measure that can be used when the interest is only in a range of specificity.

Most existing methods for analyzing multi-reader ROC data have applied mixed-effects Analysis of Variance (ANOVA) models [5-9]. Among those, the Dorfman-Berbaum-Metz (DBM) and Obuchowski-Rockette (OR) methods, respectively, proposed by Dorfman et al. [5] and Obuchowski and Rockette [6] are the most widely used. The DMB and OR methods have some drawbacks [1,10,11] and several approaches to overcome these drawbacks have been proposed [9,10,12,13]. On the other hand, methods that do not rely on mixed-effects ANOVA modeling have been less explored. In this mini-review, I introduce current literature on nonparametric approaches to analyzing multi-reader, multi-test ROC data.

Nonparametric approaches

Significant methodological developments can be applied in situations where each participant is examined by multiple readers using a single diagnostic test, or multiple diagnostic tests are read by a single reader. These settings constitute a special case of the multi-reader, multi-test design, and a comprehensive review of relevant statistical methods can be found in Zhou et al. [1] and Zou et al. [14]. Here, I introduce several nonparametric methods developed by using the theory of generalized U-statistics. Among those methods; DeLong et al. [15] approach is one of the most widely used for comparing diagnostic tests. Noting that the nonparametric AUC estimate is equivalent to the U-statistics, they derived its asymptotic normality and variance expression. Furthermore, they estimated the variance of the nonparametric AUC by applying Sen's [16] method of structural components. Gallas [17] proposed a new variance estimation technique using the idea of Barrett et al. [18]. If ROC

Table 1: Presentation of data from Multi-Reader, Multi-Test Study.

	Reader 1		…	Reader j		…	Reader J	
	Test 1	Test 2	…	Test 1	Test 2	…	Test 1	Test 2
1	T_{111}	T_{112}	…	T_{1j1}	T_{1j2}	…	T_{1J1}	T_{1J2}
2	T_{211}	T_{212}	…	T_{2j1}	T_{2j2}	…	T_{2J1}	T_{2J2}
…	…	…	…	…	…	…	…	…
k	T_{k11}	T_{k12}	…	T_{kj1}	T_{kj2}	…	T_{kJ1}	T_{kJ2}
…	…	…	…	…	…	…	…	…
N	T_{N11}	T_{N12}	…	T_{Nj1}	T_{Nj2}	…	T_{NJ1}	T_{NJ2}

T_{kj1}(T_{kj2}) denotes the result of test 1 (test 2) interpreted by reader j from patient k

Citation: Kim E. Nonparametric Approaches to Comparing the Accuracy of Diagnostic Tests with Multiple Readers. Austin Biom and Biostat. 2014;1(1): 3.

curves of different diagnostic tests cross, the use of AUCs may not be appropriate. In this case, pAUC can serve as an alternative. Several papers [19-21] have compared nonparametric pAUCs by extending DeLong et al. [15] approach.

A common nonparametric approach to analyzing multi-reader, multi-test ROC data is to compare correlated AUCs. Typically, a reader-specific AUC for a diagnostic test is calculated first to describe the diagnostic performance of a specific reader, and the diagnostic accuracy of the test is summarized as the average accuracy of all readers. Finally, the performance of different diagnostic tests is compared by testing the differences in reader-averaged AUCs [22-26].

Song [22] proposed a nonparametric method to analyze such ROC data by generalizing DeLong et al.'s [15] approach. She used the jackknife methods to estimate the variance of the nonparametric AUC, which can be computationally demanding. Kaufmann et al. [24] used the method of rankings for the nonparametric Behrens-Fisher problem and derived the AUC estimates as well as their covariance matrix. Bandos et al. [25] introduced a permutation test for comparing nonparametric AUCs. The latter two methods have been used to compare two diagnostic modalities only. Recently, Kim et al. [26] presented the closed form expression of the nonparametric AUC and estimated it using the method of structural components similar to Delong et al. [15]. They also developed a power formula to compare the correlated AUCs for any two diagnostic tests in a multi-reader, multi-test study design under the asymptotic normality of the nonparametric AUC differences. They showed that their power formula is especially useful when a study is expected to have a relatively small number of readers (e.g., less than 4) and that it can serve as an alternative to the conventional power calculations developed based on the mixed-effects ANOVA models [27-29].

While several nonparametric methods have been developed based on the theory of U-statistics, Li and Zhou [30] took a different approach by treating nonparametric ROC curves as stochastic processes and derived their asymptotic distribution theory. They used a Monte Carlo re sampling method to approximate the empirical ROC processes and compared correlated AUCs. Instead of relying on the reader-averaged AUCs presented in previous papers, Tang et al. [31] proposed using a weighted linear combination of the reader-specific AUC differences to possibly achieve a higher power for comparing two diagnostic modalities.

Conclusion

Current nonparametric approaches in multi-reader, multi-test studies mainly focus on comparisons of correlated AUCs. Nonparametric approaches do not require distributional assumptions for the test results or the ROC curve, making them attractive for use when the total sample size/number of readers is small or when distributional assumptions may be problematic. As a final remark, readers should note that the nonparametric approaches treat readers as fixed effects; they are appropriate for use in phase II studies in which readers are selected from a specific institution and the interest is in making inferences only about those readers. However, in phase III studies in which readers should represent a general population of readers, they should be selected in a representative manner and be treated as random to account for variability across readers. In this case, nonparametric approaches can still be applied but they may result in power loss when reader effects are actually random [8,14,26].

References

1. Jin H, Lu Y. A non-inferiority test of areas under two parametric ROC curves. Contemp Clin Trials. 2009; 30: 375-379.
2. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978; 8: 283-298.
3. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143: 29-36.
4. Swets JA, Pickett RM. Evaluation of Diagnostic Systems: Methods from Signal Detection theory. New York: Academic. 1982.
5. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol. 1992; 27: 723-731.
6. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: an ANOVA approach with dependent observations. Communications in Statistics - Simulation and Computation. 1995; 24: 285-308.
7. Beiden SV, Wagner RF, Campbell G. Components-of-variance models and multiple-bootstrap experiments: an alternative method for random-effects, receiver operating characteristic analysis. Acad Radiol. 2000; 7: 341-349.
8. Obuchowski NA, Beiden SV, Berbaum KS, Hillis SL, Ishwaran H, Song HH, et al. Multireader, multicase receiver operating characteristic analysis: an empirical comparison of five methods. Acad Radiol. 2004; 11: 980-995.
9. Hillis SL. A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. Stat Med. 2007; 26: 596-619.
10. Hillis SL, Berbaum KS. Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. Acad Radiol. 2005; 12: 1534-1541.
11. Song X, Zhou XH. A marginal model approach for analysis of multi-reader multi-test receiver operating characteristic (ROC) data. Biostatistics. 2005; 6: 303-312.
12. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. Acad Radiol. 2008; 15: 647-661.
13. Hillis SL, Obuchowski NA, Berbaum KS. Power estimation for multireader ROC methods an updated and unified approach. Acad Radiol. 2011; 18: 129-142.
14. Zou KH, Liu A, Bandos AI, Ohno-Machado L, Rockette HE. Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis. New York: Chapman & Hall/CRC Press. 2011.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44: 837-845.
16. Sen PK. On some convergence properties of U-statistics. Calcutta Statistical Association Bulletin 1960; 10: 1-18.
17. Gallas BD. One-shot estimate of MRMC variance: AUC. Acad Radiol. 2006; 13: 353-362.
18. Barrett HH, Kupinski MA, Clarkson E. Probabilistic foundations of the MRMC method. MP Eckstein, Y Jiang, editors. In: Medical imaging: image perception, observer performance, and technology assessment. Proc SPIE. 2005; 5749: 21–31.
19. Zhang DD, Zhou XH, Freeman DH Jr, Freeman JL. A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. Stat Med. 2002; 21: 701-715.
20. Dodd LE, Pepe MS. Partial AUC estimation and regression. Biometrics. 2003; 59: 614-623.

21. He Y, Escobar M. Nonparametric statistical inference method for partial areas under receiver operating characteristic curves, with application to genomic studies. Stat Med. 2008; 27: 5291-5308.

22. Song HH. Analysis of correlated ROC areas in diagnostic testing. Biometrics. 1997; 53: 370-382.

23. Lee MLT, Rosner BA. The average area under correlated receiver operating characteristic curves: a nonparametric approach based on generalized two-sample Wilcoxon statistics. Applied Statistics. 2001; 50: 337–344.

24. Kaufmann J, Werner C, Brunner E. Nonparametric methods for analysing the accuracy of diagnostic tests with multiple readers. Stat Methods Med Res. 2005; 14: 129-146.

25. Bandos AI, Rockette HE, Gur D. A permutation test for comparing ROC curves in multireader studies a multi-reader ROC, permutation test. Acad Radiol. 2006; 13: 414-420.

26. Kim E, Zhang Z, Wang Y, Zeng D. Power calculation for comparing diagnostic accuracies in a multi-reader, multi-test design. Biometrics. Forthcoming.

27. Obuchowski NA. Multireader, multimodality receiver operating characteristic curve studies: hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. Acad Radiol. 1995; 2 Suppl 1: S22-29.

28. Obuchowski NA. Multireader receiver operating characteristic studies: a comparison of study designs. Acad Radiol. 1995; 2: 709-716.

29. Obuchowski NA. Sample size calculations in studies of test accuracy. Stat Methods Med Res. 1998; 7: 371-392.

30. Li G, Zhou K. A Unified Approach to Nonparametric Comparison of Receiver Operating Characteristic Curves for Longitudinal and Clustered Data. J Am Stat Assoc. 2008; 103: 705-713.

31. Tang LL, Liu A, Chen Z, Schisterman EF, Zhang B, Miao Z. Nonparametric ROC summary statistics for correlated diagnostic marker data. Stat Med. 2013; 32: 2209-2220.

Austin Biom and Biostat - Volume 1 Issue 1 - 2014
Submit your Manuscript | www.austinpublishinggroup.com

Citation: Kim E. Nonparametric Approaches to Comparing the Accuracy of Diagnostic Tests with Multiple Readers. Austin Biom and Biostat. 2014;1(1): 3.