

医学论文中 ROC 曲线应用错误例析

王曼 徐春燕 施学忠

郑州大学学报编辑部,450001,郑州

摘要 诊断试验是医学研究中重要的试验方法。ROC 曲线是评价诊断试验价值的常用分析方法。目前医学研究中 ROC 曲线应用越来越广泛,涉及 ROC 曲线的医学论文也越来越多。文章对已发表的医学文献中 ROC 曲线应用方面,如有无必要绘制 ROC 曲线、ROC 曲线形态是否错误、ROC 曲线结果报告和解读等常常出现的错误,进行了举例,并分析错误原因,提出应对方法。

关键词 医学论文;诊断试验;ROC 曲线

Analysis of errors in application of ROC curves in medical papers/WANG Man, XU Chunyan, SHI Xuezhong

Abstract Diagnostic test is an important test method in medical research. ROC curve is a common analytical method used to evaluate the value of diagnostic tests. At present, ROC curve is more and more widely used in medical research, and more and more medical papers are concerning ROC curve. This paper summarized the application of ROC curve in published medical literatures, such as whether it is necessary to use ROC curves, the shape of ROC curve is right or not, whether errors in report and interpretation of ROC curve are illustrated, and the causes of errors are analyzed. Based on these analysis, we also proposed related countermeasures.

Keywords medical paper; diagnostic test; ROC curve

Authors' address Editorial Department of Journal of Zhengzhou University, 450001, Zhengzhou, China

DOI:10.16811/j.cnki.1001-4314.2019.02.011

诊断试验是临床上常用的一类试验方法。临床上诊断试验的目的不仅包括利用各种诊断指标区分病患和非病患,还扩展至利用诊断指标评估预后、判断疗效等;因此准确评价诊断试验的价值意义重大。陈平雁^[1]曾对诊断试验方法进行了详细的总结。张媛等^[2]也提出编辑应重视医学论文中诊断试验分析方法的审核。敏感度和特异度是常用的诊断试验评价指标。敏感度也称真阳性率,反映某试验方法发现阳性的能力。特异度也称真阴性率,反映某试验方法排除阳性的能力。由于阳性和阴性对象的某指标检测结果可能重叠;因此敏感度和特异度有可能随着诊断界值的改变而改变,而且两者变化方向相反,一个指标增大则另一个必然减小。

内容受试者工作特征曲线(ROC 曲线)是以敏感度为纵坐标,以(1-特异度)为横坐标绘制而成,曲线上的任意一点表示观察指标的某个阳性标准值对应的

一对敏感度和特异度;曲线下面积(AUC)反映试验准确度。ROC 曲线是目前最常用的诊断试验评价工具。笔者发现,涉及 ROC 曲线的医学论文越来越多,但存在很多问题,对这类文章的编辑加工要点少有总结。现对已发表文献中 ROC 曲线的错误应用举例分析,与编辑们交流探讨。

1 ROC 曲线绘制的必要性

传统的诊断试验评价方法要求将试验结果分为 2 类,再进行统计分析。而 ROC 曲线无此限制,可以根据实际情况划定试验结果,允许有中间状态,试验结果可以呈现为多分类(如正常、大致正常、可疑、大致异常和异常等),这种特性对于鉴别诊断更有意义。利用 ROC 曲线寻找某项诊断指标的最佳阳性标准值(即诊断界值或截断值)是绘制 ROC 曲线的主要目的。当诊断指标是多分类或连续型变量时,每设定一次阳性标准值,就得出—对敏感度和特异度。通常曲线左上角那一点对应最佳界点,在此界点上,可同时满足诊断试验的敏感度和特异度相对最优;但是对于二分类资料(比如某诊断方法的临床诊断标准已经确定),只有一个诊断界点,ROC 曲线表现为有一个顶点的折线。这种情况下绘制 ROC 曲线并求取 AUC 的意义不大。同样的,当观察指标为分类变量,分类较少时,有无必要绘制 ROC 曲线值得商榷。通常,我们认为对于诊断指标为 5 个以上分类或连续型数据才有绘制 ROC 曲线的必要。

案例 1 某研究^[3]收集了经临床证实的微小肝癌(MHCC)患者 40 例(52 灶)和再生结节患者 21 例(35 灶)的超声、CT 和 MRI 等影像学数据,运用 ROC 曲线比较 3 种影像学检查方法对 MHCC 的诊断效能,见图 1。

解析 该研究的诊断指标为超声、CT 或 MRI 判定 MHCC 的结局(即判定是和不是),诊断指标为二分类变量,取值只有一个,无需绘制 ROC 曲线。

2 ROC 曲线形态错误

ROC 曲线的绘制有非参数法和参数法 2 种方法。非参数法绘制的 ROC 曲线也称为经验 ROC 曲线,一般 SPSS、SAS 等统计软件均采用非参数法绘制,曲线

表现为顶点较多的折线。一些专用的 ROC 分析软件(如 ROCKIT)采用参数法绘制 ROC 曲线,曲线光滑,故也称为光滑 ROC 曲线。有些文章中 ROC 曲线的光滑度明显与描述的统计软件不符。还有些文章中 ROC 曲线表现为向右下凸起,此时曲线在机会线以下, AUC 小于 0.5。出现这种错误有可能是在绘制 ROC 曲线时阳性、阴性标准的设置反了。ROC 曲线纵横坐标的刻度单位是一样的,因此纵横坐标比例为 1:1,全域图片应该呈正方形。另外,笔者建议可把多个相同指标的多条 ROC 曲线绘制在一张图上,以便于对比。

案例 2 某研究^[4]应用 ROC 曲线分析 miR-195 表达对诊断 CIN、宫颈癌的意义,分析结果表明外周血

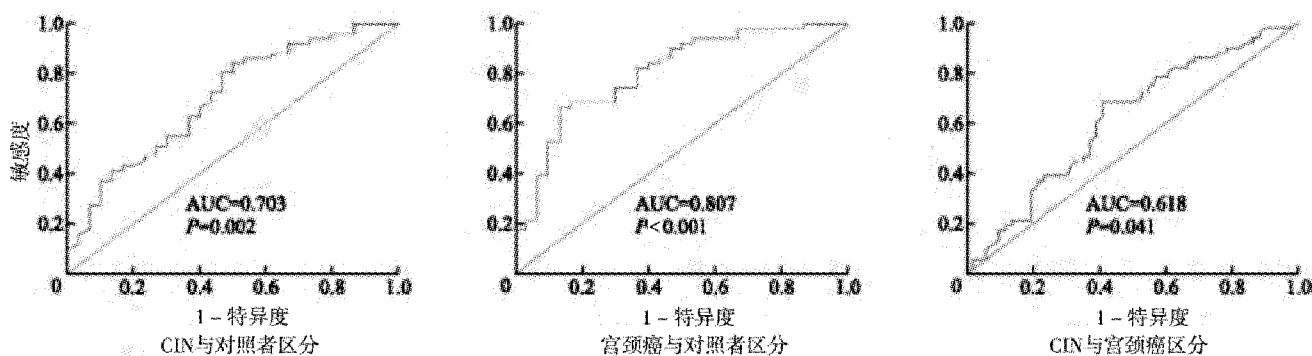


图2 案例2中 miR-195 表达诊断 CIN、宫颈癌的 ROC 曲线

解析 该研究 3 条 ROC 曲线可绘制在一张图中,对比效果更强,提示作用更大;3 张图的纵横坐标比例均明显小于 1,横长纵短。作者或编辑可能将 ROC 曲线当做一般线图,按照纵横坐标比例 5:7 绘制,忽略了 ROC 曲线的原理和纵横坐标的含义。

3 ROC 曲线结果报告和解读错误

对于诊断指标是多分类或连续型变量的研究,除了图片,还应根据实际情况报告敏感度、特异度、约登指数、阳性和阴性似然比等指标;在报告这些指标时必须说明所取的界值。一般情况下,AUC 及其 95% 可信区间(CI)也是必须报告的内容。应注意 ROC 曲线与表格数据、文字叙述的一致性^[5]。有些文章表格中的敏感度、特异度等数据与曲线不符。有文章^[6]的 AUC 竟然大于 1。AUC 的比较用检验统计量 Z , Z 值符合渐进标准正态分布。当遇到多于 2 个 AUC 比较时,涉及多重比较的问题,需要调整检验水准 α 。有些文章在进行多个 AUC 比较时, α 仍为 0.05,此时会增加 I 型错误的概率。

案例 3 某项研究^[7]纳入了首次新发急性缺血性

miR-195 能够很好地鉴别宫颈癌和 CIN,ROC 曲线见图 2。

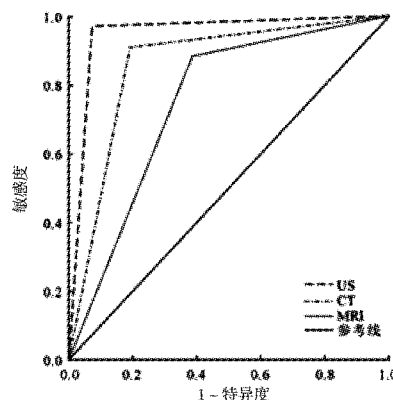


图1 案例1的 ROC 曲线

脑卒中(AIS)存活患者 240 例,随访 6 个月,依据神经功能恢复情况分为预后良好组 136 例和预后不良组 104 例。检测患者血浆巨噬细胞移动抑制因子(MIF)水平,采用 ROC 曲线(图 3)评估血浆 MIF 水平判断 AIS 患者不良预后的价值。作者对 ROC 曲线进行了文字总结(原文 1014 页 2.3 部分),叙述为“MIF 水平判断 AIS 患者不良预后的 ROC 曲线下面积为 0.722 ($P < 0.01$),且判断不良预后的最佳界点为 107.8 $\mu\text{g/L}$,判断不良预后的敏感性为 89.6%,特异性为 67.4%。”

解析

1)对于某项指标的检测,必须说明检测时间,是在治疗或干预前,治疗或干预过程中,还是治疗或干预后?对于预测研究来说,用就诊时或治疗前的检测结果进行预测才有应用价值。随访中的检测结果不适宜用于预测。该研究随访 6 个月,文内却没有交代 MIF 的检测时间。

2)文中叙述“最佳界点为 107.8 $\mu\text{g/L}$,判断不良预后的敏感性为 89.6%,特异性为 67.4%”。由图 3 中的虚线可见,ROC 曲线上纵坐标为 89.6% (即敏感

度为 89.6%) 的点(即虚线与曲线相交的点)所对应的横坐标为 67.4%, 而横坐标为 1 - 特异度, 所以由图得出 1 - 特异度 = 67.4%, 那么由图得出的特异度为 1 - 67.4% = 32.6%。可见, 图 3 反映出的特异度为 32.6%, 而原文 2.3 文字叙述则为“特异性为 67.4%”。因此, 图文数据不符, 存在错误。进一步分析可以发现, 敏感度为 89.6%、特异度为 32.6% 并不能达到相对最优, 提示最佳界点选取有问题。

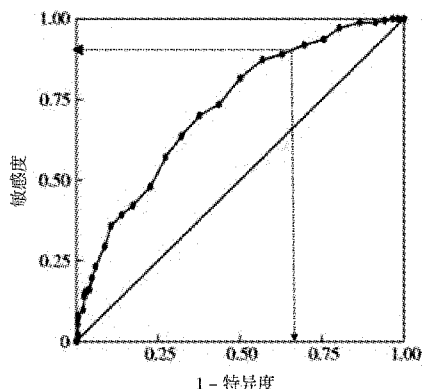


图3 案例3中MIF判断AIS不良预后的ROC曲线

案例4 某研究^[8]回顾性分析了168例创伤继发ARDS患者的临床资料, 受伤后28d内死亡65例, 存活103例, 利用SPSS绘制ROC曲线, 评价患者入院24h内APACHE II评分、ISS评分和SOFA评分对死亡的预测价值。ROC曲线见图4, 分析结果见表1, 认为ISS的预测价值大于APACHE II、APACHE II大于SOFA。

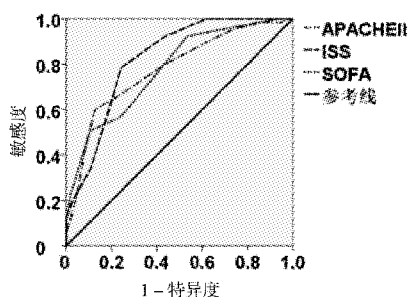


图4 案例4中3种评分预测死亡的ROC曲线

表1 案例4中3种评分预测死亡的ROC曲线分析结果

评分	AUC(95% CI)	P
APACHE II	0.778(0.705 ~ 0.851)	<0.001
ISS	0.823(0.762 ~ 0.885)	<0.001
SOFA	0.777(0.706 ~ 0.847)	<0.001

解析 该研究属配对设计, 文内未对3条ROC曲线AUC的比较进行说明, 表1中未报告Z值。3条曲线AUC(95% CI)区域重叠严重, P值可疑。以此结果

下结论(ISS的预测价值大于APACHE II、APACHE II大于SOFA), 严重影响结论的客观性和科学性。

3个评分对应的ROC曲线有交叉, 因此不能简单地比较3条ROC曲线的总AUC。胡良平^[9]认为, 当2条ROC曲线交叉时, 应比较2个诊断试验部分AUC或固定假阳性率时的敏感度。鉴于该研究诊断指标的特点, 建议比较固定假阳性率时的敏感度。

4 结束语

ROC曲线最早应用在军事上, 后来逐渐应用到医学领域, 目前在医学研究中广泛应用。ROC分析其实质是对分类方法的评价, 不仅仅应用于医学的诊断试验, 也大量应用于各种分类器、分类算法的评价。正确应用ROC曲线才能得到科学、客观的医学研究结果。本文所列举的案例均来自于已发表文献。医学期刊编辑应注重对医学统计学学习^[10-11], 了解ROC曲线的原理和应用时的注意事项, 以增强对这类文章的审阅、编辑加工的能力, 从而有助于期刊内容质量的提升。

5 参考文献

- [1] 陈平雁. 关于诊断试验方法的若干问题[J]. 北京大学学报(医学版), 2010, 42(6): 764
- [2] 张媛, 韩银, 石朝云. 编辑应重视审核医学论文中的诊断试验分析方法[J]. 编辑学报, 2013, 25(5): 431
- [3] 宋琦炜, 张雷. 微小肝癌早期诊断的临床常规影像学检查方法对照分析[J]. 实用放射学杂志, 2017, 33(12): 1943
- [4] 赵静. MicroRNA-195在宫颈癌及子宫上皮内瘤变患者外周血中的表达及临床意义[J]. 江苏大学学报(医学版), 2017, 27(6): 496
- [5] 陈先军. 科技期刊论文的图表审读处理方法探讨[J]. 编辑学报, 2018, 30(3): 266
- [6] 解欣欣, 郑慧, 万颖. ASQ技术评价慢乙型肝炎纤维化的应用价值[J]. 安徽医科大学学报, 2016, 51(10): 1529
- [7] 李素萍, 何保明, 梁益, 等. 巨噬细胞移动抑制因子对急性缺血性脑卒中患者神经功能缺损的预测价值[J]. 中华老年心脑血管病杂志, 2018, 20(10): 1012
- [8] 江浩, 郑贺根, 朱超男, 等. APACHE II、ISS、SOFA评分对创伤继发ARDS严重程度及预后的预测价值比较[J]. 临床肺科杂志, 2018, 23(11): 2074
- [9] 胡良平. 医学统计学: 运用三型理论分析定量与定性资料[M]. 北京: 人民军医出版社, 2009: 255
- [10] 韩磊. 医学期刊编辑提升统计学素养的难点及策略[J]. 科技与出版, 2016, 35(1): 50
- [11] 周英智. 医学论文“统计学处理”常见问题分析及建议[J]. 中国科技期刊研究, 2016, 27(5): 480

(2018-11-20收稿; 2019-01-20修回)