第二军医大学学报
Acad J Sec Mil Med Univ 2006 May;27(5)

·短篇论著·

应用 ROC 曲线下面积对两相关诊断试验进行评价和比较

Area under ROC curves in evaluation and comparison of two correlated diagnostic tests

宋花玲，贺　佳*，虞慧婷，李　玲
（第二军医大学卫生勤务学系卫生统计学教研室，上海 200433）

[摘要]　**目的**：阐明两相关诊断试验的评价及比较的方法，为新的诊断试验的临床应用提供科学依据。**方法**：通过 ROC 曲线下面积来对两相关诊断试验进行评价和比较，用拟合双正态模型的参数法来估计 ROC 曲线下面积，并通过对前列腺癌的两种诊断试剂的分析来说明方法的应用。**结果**：采用 ROC 曲线的方法，得到的对照试剂两指标 Free-PSA、Total-PSA ROC 曲线下面积分别为 0.92、0.90，检测试剂相应指标的 ROC 曲线下面积分别为 0.91、0.89，2 个试剂相应指标 ROC 曲线下面积的差异无统计学意义，经等效性检验 Z 值分别为 2.73 、2.78，相应的 P 值为 0.006 4、0.005 4，有统计学意义。**结论**：2 个诊断试剂的分析表明均具有较好的诊断性能，其诊断准确度相同，同时也表明采用 ROC 曲线下面积可方便地对两诊断试验进行评价和比较。

[关键词]　诊断试验，常规；ROC 曲线下面积；治疗等效
[中图分类号]　R 446　　[文献标识码]　B　　[文章编号]　0258-879X(2006)05-0562-02

随着科学技术的不断发展，新的医学诊断方法也不断出现，而新的诊断方法的诊断性能如何、临床医生应当如何选择诊断试验主要依靠于诊断试验的评价结果。诊断试验的基本评价指标有灵敏度、特异度、一致率等，综合评价指标有 youden 指数、似然比、ROC(receiver operating characteristic)曲线下面积等，但除了 ROC 曲线外，它们的大小均受诊断界值的影响，即随着诊断界值的改变而改变，而 ROC 曲线下面积不仅综合了灵敏度和特异度 2 个指标，而且考虑了每一个可能的界值，因而能够更客观的评价诊断试验的诊断价值，目前也已作为诊断试验公认的标准评价指标。同时，为了准确地比较 2 种诊断试验的诊断价值，常常采用配对设计的方法来进行研究，即随机选择一些患者和非患者作为研究对象，对每一个研究对象同时用 2 种诊断试验进行诊断，此时两种试验的诊断结果具有一定的相关性，用 ROC 曲线下面积来比较两诊断试验时要考虑到两面积间的相关性[1]。本文主要介绍应用 ROC 曲线下面积对两相关诊断试验进行评价和比较[2]的方法，差异无统计学意义时进行等效性检验[3]，并通过实例来具体阐明该方法的应用，从而为新的诊断方法的应用提供科学依据。

1　材料和方法

ROC 曲线是以每一个检测结果作为可能的诊断界值，以计算得到相应的真阳性率(即灵敏度)为纵座标，以假阳性率(即 1－特异度)为横坐标绘制曲线，其曲线下面积的大小表明了诊断试验准确度的大小。ROC 曲线下面积的估计有参数法和非参数法，均适用于结果为连续性资料或等级资料的诊断试验的评价。非参数法是根据实验结果直接计算绘制 ROC 曲线所需的工作点，由此绘制的 ROC 曲线为经验 ROC 曲线，其曲线下面积与患者和非患者实验结果秩和检验的 Mann-Whitney 统计量相等，但其结果常小于真实的面积值[2]。参数法是根据试验结果拟合双正态模型，可利用最大似然法估计其 2 个参数，由 2 个参数可得到光滑的 ROC 曲线及曲线下面积的估计值。参数法的应用条件为：患者与非患者的试验结果服从双正态分布，但这是指 ROC 曲线的函数形式，而不是指试验结果的基本分布，因为变量变换几乎可使任何试验结果转换为双正态分布，而且在样本量较大、相同值较少时参数法与非参数法估计的 ROC 曲线下面积常常近似相等[2]。本文采用双正态模型法计算曲线下面积。

参数法估计 ROC 曲线下面积的公式为：

$$A=\Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

其中 A 为 ROC 曲线下的面积，a、b 分别为双正态模型的 2 个参数，Φ 表示标准正态分布函数。当 2 个 ROC 曲线无交叉时相关 ROC 曲线下面积的比较可根据曲线下面积、面积的方差及面积间的协方差由以下公式[2]计算检验统计量得出结论：

$$Z=\frac{A_1-A_2}{\sqrt{Var(A_1)+Var(A_2)-2Cov(A_1,A_2)}}$$

其中的 Z 值服从或近似服从标准正态分布，查正态分布表可得相应的 P 值，A_1、A_2 为两诊断试验 ROC 曲线下的面积，$Var(A_1)$、$Var(A_2)$ 为两曲线下面积的方差，$Cov(A_1,A_2)$ 为两曲线下面积的协方差，当两诊断试验独立时，此协方差项为 0。两面积的等效性检验根据上述公式，以对照试验 ROC 曲线下面积的 5% 为参照，面积的差值与之相比进行统计学检验[3]。当 2 个 ROC 曲线交叉时，两诊断试验的比较应比较部分 ROC 曲线下的面积或固定假阳性率时的灵敏度。以上的参数及相关指标可利用最大似然法进行估计，但

[作者简介]　宋花玲，硕士，讲师.
* Corresponding author. E-mail：hejia63@yahoo.com

计算复杂,可通过软件 ROCKIT0.9β 来实现。

以杭州、武汉的某 2 个医院为研究单位,随机从门诊和住院人群及体检人群中选择 239 例男性研究对象,包括正常人、良性疾病患者和前列腺癌(prostate cancer,PCa)患者,其中的良性疾病包括前列腺炎、前列腺增生及其他良性疾病。正常人为 104 人,平均年龄(49.35±16.87)岁;其他研究对象由临床病理检查结果分为 PCa 患者组 77 人,平均年龄(64.55±12.22)岁;良性疾病组 58 人,平均年龄(56.90±17.31)岁,采用配对设计的方法,由经过培训的检验医师分别用某国际知名公司的 PCa 诊断试剂(以下简称为对照试剂)和上海某公司新开发的 PCa 诊断试剂(以下简称为检测试剂),在严格遵循操作规程的条件下同时对各研究对象血清中的游离前列腺特异抗原(Free-PSA)和总前列腺特异抗原(Total-PSA)进行定量检测。用软件 ROCKIT0.9β 对检测结果进行分析。

2 结 果

由各研究对象的检测结果分析得到检测试剂(x)和对照试剂(y)各指标的 ROC 曲线见图 1A、1B,相应的 ROC 曲线下面积及其他相关指标见表 1。

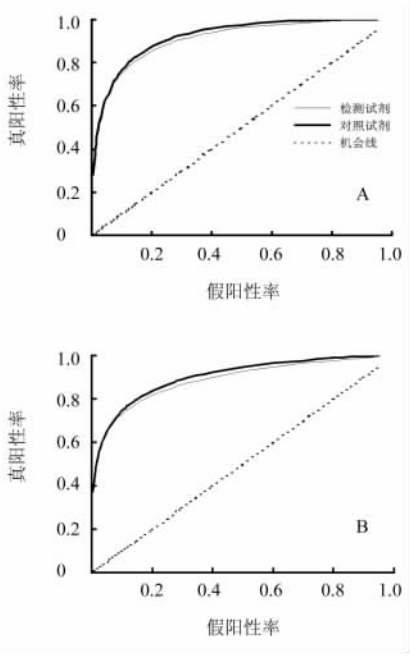

图 1 检测试剂和对照试剂 Free-PSA(A)和 Total-PSA(B)的 ROC 曲线

表 1 检测试剂和对照试剂 ROC 曲线下面积的比较

指标	试剂	A	Z_1	P_1	A_x-A_y	95%CI	Z_2	P_2	Z_3	P_3
Free-PSA	x	0.91	19.93	<0.000 1	0.01	−0.04～0.02	0.80	0.421 2	2.73	0.006 4
	y	0.92	17.63	<0.000 1						
Total-PSA	x	0.89	15.31	<0.000 1	0.01	−0.04～0.01	1.33	0.183 8	2.78	0.005 4
	y	0.90	17.77	<0.000 1						

A_x、A_y分别代表检测试剂和对照试剂的曲线下面积,Z_1、P_1分别为曲线下面积与 0.5(为完全无诊断价值的诊断试验 ROC 曲线即机会线下的面积)相比较的统计学检验结果,Z_2、P_2为检测试剂和对照试剂曲线下面积比较的结果,Z_3、P_3为检测试剂和对照试剂曲线下面积等效性检验的结果

由图 1 可直观地看出,2 个试剂各指标的 ROC 曲线均在机会线以上,并远离机会线,同时 2 个试剂相应指标的 ROC 曲线无交叉且非常接近。由表 1 可进一步得知,检测试剂和对照试剂 2 个检测指标的 ROC 曲线下面积与 0.5 相比均有统计学意义(P<0.000 1),说明 2 个试剂对 PCa 均有较好的诊断价值;2 个试剂 ROC 曲线下面积的差异比较,P 值均>0.05,差异无统计学意义,表明从真实性方面来看 2 个试剂对于 PCa 的诊断未见差异;等效性检验的结果 P 值均<0.01,进一步说明 2 个试剂对 PCa 的诊断准确度相同。

3 讨 论

一项诊断试验的应用价值主要从实用性、真实性和精确性 3 个方面来评价[4],而其中以试验本身的真实性评价最为重要。ROC 曲线下面积作为诊断试验真实性评价的固有准确度指标已被普遍认可,完全无价值的诊断试验曲线下面积为 0.5,理想的诊断试验曲线下面积为 1,而一般认为对于一个诊断试验,ROC 曲线下面积在 0.5～0.7 之间时诊断价值较低,在 0.7～0.9 之间时诊断价值中等,在 0.9 以上时诊断价值较高[5]。本研究通过 ROC 曲线下面积的估计值,表明 2 个试剂各指标的 ROC 曲线下面积都在 0.9 左右,表明诊断价值都较高;面积的差异比较无统计学意义,进一步的等效性检验表明其诊断准确度相同。综上所述,在临床应用中可以进一步从 2 个试剂的成本等方面来考虑试剂的选择。

[参 考 文 献]

[1] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach[J]. Biometrics, 1988, 44: 837-845.

[2] Zhou X, Obuchowski NA, Mcclish DK. Statistical methods in diagnostic medicine[M]. New York: Wiley, 2002: 111-136, 180-187.

[3] 孙振球. 医学统计学[M]. 北京:人民卫生出版社,2002:36-37.

[4] 林果为. 诊断试验的研究与评价[J]. 诊断学理论与实践,2003,2:附 1-附 4

[5] 余松林. 医学统计学[M]. 北京:人民卫生出版社,2002:164-178.

[收稿日期] 2006-01-23 [修回日期] 2006-03-21

[本文编辑] 尹 荼