

# 参数法与非参数法计算 AUC 评价诊断试验准确性的样本量对比

谷红梅<sup>1</sup> 高金霞<sup>2</sup> 牛莹莹<sup>3</sup> 范兴君<sup>4</sup> 杨晓炜<sup>5</sup>  
牡丹江医学院, 中国·黑龙江 牡丹江 157011

**【摘要】**目的: 比较不同参数下参数法与非参数计算 AUC 时所需样本量, 为评价诊断试验准确性估算样本量提供参考。方法: 据 Obuchowski NA 等人提出的公式估计初始值, Monte Carlo 模拟调整参数法、非参数法计算 AUC 时样本量。结果: 参数法计算 AUC 时, 所需样本量较非参数法少, 预期  $\theta > 0.900$  时, 所需样本量少于 50。结论: 评价双正态模型时, 可据参数法计算 AUC 进行样本量估计。

**【关键词】**参数法; 非参数法; 样本量

**【基金项目】**本文系黑龙江省教育厅科研基金项目《双正态 ROC 参数比较的样本量确定》(11541383)。

## 引言

ROC 曲线下面积(AUC)是评价诊断试验准确性的公认指标<sup>[1-2]</sup>, 目前常用的 AUC 计算方法主要有参数法、非参数法、贝叶斯法<sup>[3]</sup>。有报道显示: 样本量较小时, 参数法计算 AUC 结果可能严重偏离真实值<sup>[4]</sup>。这可能会对使用 AUC 评价诊断试验准确性的样本量估计产生影响, 故本次研究采用参数法、非参数法计算 AUC, 进行 Monte Carlo 模拟试验, 对比 AUC 两种计算方法评价诊断试验所需样本量, 为实际评价诊断试验准确性时估算样本量提供参考。

## 1 研究方法

据  $N = 4 \times Z_{\alpha/2}^2 \cdot V(\hat{\theta}) / \Delta^2$  公式<sup>[5]</sup>对双正态模型的诊断试验进行病例组初始值样本量估计, Monte Carlo 模拟调整样本量<sup>[6]</sup>, 模拟试验参数设定: 正常组与病例组样本量比  $R=1:1$ ; 正常组与病例组试验测量值标准差比  $B$  为  $1/1$ 、 $1/2$ 、 $1/3$ ; 预设中等以上诊断价值<sup>[7]</sup>AUC 为  $0.700 \sim 0.975$ ; (4) 预期  $\alpha = 0.05$ 、 $0.01$ ; (5) 允许误差  $\Delta = 0.10$ 。

## 2 结果

### 2.1 $\alpha = 0.05$ 时两种 AUC 计算方法病例组样本量

$\alpha = 0.05$  时, 除  $B=1/1$  且预期 AUC=0.975 时两种 AUC 计算方法所需样本量相同外, 其他预设  $B$  与预期 AUC 组合时, 两种方法所需样本量不同(见表 1), 参数法所需样本量  $N_p$  相对非参数法所需样本量  $N_{nonp}$  少。在预期 AUC > 0.900 且  $B=1$  时, 出现所需样本量 < 50 的情况。

表 1  $\alpha = 0.05$  时参数法与非参数法计算 AUC 样本量

预期 AUC	预期 B=1/1		预期 B=1/2		预期 B=1/3	
	$N_p$	$N_{nonp}$	$N_p$	$N_{nonp}$	$N_p$	$N_{nonp}$
0.700	199	205	203	223	207	246
0.725	188	193	193	210	196	231
0.750	172	178	179	196	184	216
0.775	158	162	165	181	170	201
0.800	141	145	149	162	155	181
0.825	124	127	131	144	137	162
0.850	105	108	113	124	119	140
0.875	86	89	93	104	99	118
0.900	66	69	72	82	77	95
0.925	46	49	51	59	54	70
0.950	28	29	31	35	34	42
0.975	15	15	17	19	18	24

### 2.2 $\alpha = 0.01$ 时两种 AUC 计算方法病例组所需样本量

$\alpha = 0.01$  时, 除标准差比  $B=1/1$  且预期 AUC=0.950、0.975 时参数法、非参数法计算 AUC 评价诊断试验准确性所需样本量相同外, 其他预设的  $B$  与预期 AUC 组合情况下, 两种计算 AUC 方法所需样本量不同(见表 2), 参数法计算 AUC 时所需样本量  $N_p$  相对较非参数法所需样本量  $N_{nonp}$  少, 在预期 AUC > 0.950 时, 三种预设的  $B$  值, 均出现所需样本量低于 50 的情况。

表 2  $\alpha = 0.01$  时两种方法计算 AUC 评价诊断试验准确性所需样本量

预期 AUC	预期 B=1/1		预期 B=1/2		预期 B=1/3	
	$N_p$	$N_{nonp}$	$N_p$	$N_{nonp}$	$N_p$	$N_{nonp}$
0.700	346	360	355	389	359	420
0.725	321	336	334	364	341	398
0.750	300	309	310	340	317	377
0.775	277	281	286	312	294	346
0.800	248	254	260	280	268	314
0.825	216	224	228	251	239	280
0.850	186	190	197	215	205	242
0.875	153	155	164	183	173	203
0.900	118	123	128	144	137	168
0.925	87	89	97	105	101	126
0.950	58	58	64	72	68	86
0.975	31	31	35	41	38	50

## 3 讨论

ROC 曲线下面积 AUC 作为诊断试验准确性的评价指标, 因其诸多优点而被广泛应用。目前常用的 AUC 计算方法中, 非参数法没有使用的条件限制, 但所计算的 AUC 会出现较真实值偏小的情况。而双正态模型的参数法计算 AUC 较准确, 但也存在样本量较小时偏离真实值的情况<sup>[7]</sup>。故在评价诊断试验准确性时, 样本量大小合适与否, 对最终评价结果有着至关重要的影响。本次研究发现, 在预设诊断试验结果为双正态模型情况下, 无论预期预期错误概率  $\alpha$  是 0.05 还是 0.01, 参数法计算 AUC 所需样本量多呈现为较非参数法计算 AUC 所需样本量少, 仅在预期 AUC 较大时出现两种计算 AUC 所需样本量相等的情况。通常将样本量  $N \geq 50$  视为大样本<sup>[8]</sup>, 本次研究结果显示, 仅在  $\alpha = 0.05$  且预期  $B=1/1$ 、预期 AUC > 0.900 时, 和  $\alpha = 0.01$  同时三个预期  $B$  值及预期 AUC > 0.950 出现样本量小于 50 的小样本量情况。结合 Feng Dai 等人<sup>[9]</sup>报道 AUC  $\geq 0.85$  且样本量  $N \leq 15$  时, AUC 可信区间估计参数法优于非参数法, 故建议在评价双正态模型诊断试验时, 可据参数法计算 AUC 所需的样本量进行诊断试验评价的设计。

### 参考文献:

- [1] Krzanowski WJ and Hand DJ. ROC curves for continuous data. Boca Raton, FL: CRC Press, 2009.
- [2] Zhou XH, Obuchowski NA and McClish DK. Statistical methods in diagnostic medicine. Vol. 712, New York: John Wiley & Sons, 2011.
- [3] Feng Dai, Cortese Giuliana, Baumgartner Richard. A comparison of confidence credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size[J]. Statistical Methods in Medical Research, 2017, 26(6): 1-23.
- [4] 宋花玲, 贺佳, 黄品贤, 李素云. ROC曲线下面积估计的参数法与非参数法的应用研究[J]. 第二军医大学学报, 2006, 27(07): 726-728.
- [5] Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices[J]. Stat. Med., 1997, 16(13): 1529-1542.
- [6] 谷红梅, 李康. 诊断试验ROC参数估计双正态样本量估计方法探讨[J]. 中国卫生统计, 2006, 23(1): 2-4.
- [7] 余松林. 主编. 医学统计学. 北京: 人民卫生出版社, 2002: 164-178.

[8] 李康. 主编. 医学统计学[M]. 7版. 北京: 人民卫生出版社, 2018: 72.

### 作者简介:

谷红梅 (1969.12-), 女, 汉, 籍贯: 黑龙江牡丹江, 研究生, 副教授, 哈尔滨医科大学流行病学与卫生统计, 人群健康管理, 牡丹江医学院公共卫生学院。

高金霞 (1978.09-), 女, 满族, 籍贯: 黑龙江富锦市, 研究生, 讲师, 哈尔滨医科大学流行病学与卫生统计, 人群健康促进, 牡丹江医学院公共卫生学院。

牛莹莹 (1978.09-), 女, 汉, 籍贯: 黑龙江绥化市, 研究生, 讲师, 佳木斯大学流行病学与卫生统计, 职业人群健康促进, 牡丹江医学院公共卫生学院。

范兴君 (1981.09-), 女, 汉, 籍贯: 陕西省临汾市, 研究生, 讲师, 山西医科大学卫生毒理学, 职业人群健康促进, 牡丹江医学院公共卫生学院。

杨晓炜 (1981.06-), 男, 汉, 籍贯: 内蒙古赤峰市, 本科, 讲师, 哈尔滨医科大学公共卫生事业管理, 环境卫生检验, 牡丹江医学院公共卫生学院。