

· 计算机应用 ·

诊断试验评价的 MedCalc 统计软件实现*

杨皓斌¹ 罗 磊¹ 舒 眉^{2,3△}

【摘要】目的 探讨诊断试验相关评价指标的 MedCalc 统计软件分析方法的实际应用。**方法** 结合临床诊断试验研究实例,介绍诊断试验样本量估计以及相关评价指标的 MedCalc 统计软件实现过程和结果解读。**结果** MedCalc 统计软件可以更方便地实现诊断试验中样本量估计、ROC 曲线分析、连续性资料的截断值确定、计算灵敏度、特异度、阳性预测值、阴性预测值、阳性似然比、阴性似然比、Youden 指数及其 95% 置信区间,比较两种诊断试验的 ROC 曲线下面积。**结论** MedCalc 统计软件具有界面友好、操作简单和结果可视化清晰的优点,适合在诊断试验或筛检试验相关评价指标的实际应用。

【关键词】 诊断试验 评价 ROC 曲线 MedCalc 统计软件

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.06.030

诊断试验是指应用各种实验技术、医疗仪器及其它手段对患者进行检查,以对疾病做出诊断的试验^[1]。诊断试验研究与评价的核心要点包括选择适宜的金标准,样本量估计以及诊断试验的真实性和可靠性评价等。其中样本量估计以及灵敏度、特异度、似然比、预测值、ROC 曲线下面积等评价指标的统计描述和统计推断均需要利用统计分析软件来实现。大多数学者使用 SPSS、SAS、R 等统计软件来对诊断试验评价指标进行统计分析,然而 excel、SPSS 等软件只能分析单个诊断试验计算 ROC 曲线下面积,不适合两种诊断试验 ROC 曲线下面积差异的分析,也缺乏一站式解决诊断试验真实性和可靠性评价的解决方案。SAS、R 等软件虽然能够通过编辑代码来实现比较完整的统计分析,但非统计专业人士一般难以编辑相应的软件操作代码,进而影响软件的实际应用^[2-5]。Medcalc 软件 (MedCalc © version 20.0.27, MedCalc Software Ltd ©, Ostend, Belgium, <https://www.medcalc.org>, 2022) 是比利时的一款生物统计学软件^[6],分析功能强大,用户界面友好,一般非专业人士也能掌握。但目前国内外应用 Medcalc 软件进行诊断试验分析的文章较少^[7-9],本文将利用 MedCalc 统计软件来实现诊断试验有关指标的计算,并进行结果解读,以便医学研究者在实际工作中推广应用。

MedCalc 统计软件简介

MedCalc 是一款专门为医学和生物学科学研究设计的统计软件^[6],集统计、绘图和样本量估计等功能于一体,主要统计分析有计量资料的统计描述,分类资

料统计分析,相对数的估计与比较, t 检验、方差分析、非参数检验、相关与回归、生存分析、meta 分析、连续性决策资料的序列测量分析,医学参考值范围的制定,方法比较与评价、诊断试验和样本量估计等。MedCalc 是诊断试验 ROC 曲线分析的经典统计分析软件,同时包含了检验医学中特有的统计分析过程,如检测(测量)方法比较与评价方法中的 Bland-Altman 图、火山图、Passing-Bablok 回归和 Deming 回归等。MedCalc 统计软件采用易学易用的可视化操作界面,具有用户友好、快速和可靠的特点。该软件的统计模块包括 ROC 曲线分析与比较、预测值、似然比以及 2×2 列联表的诊断试验分析。该软件还可方便实现诊断试验的样本量估计,如以灵敏度、特异度预期值估计样本量,以 ROC 曲线的样本量估计。

MedCalc 在诊断试验样本量估计中的应用

诊断试验中研究对象包括基于金标准确定的患者和非患者。在单个诊断试验研究中,可以利用率的参数估计公式来估计样本量,通常利用诊断试验的灵敏度估计患者的样本量,用特异度时估计非患者的样本量;在两诊断试验灵敏度或特异度比较研究中,可以类似于两样本率比较的公式来估计样本量,用诊断试验的灵敏度估计患者的样本量,用特异度时估计非患者的样本量。另外,还可以比较两诊断试验 ROC 曲线下面积 (area under the curve, AUC) 来估计样本量。

1. 按照诊断试验的灵敏度、特异度估计样本量

(1) 单个诊断试验样本量估计

【例 1】 预计所评价的诊断试验的灵敏度为 90%, 特异度为 85%, $\alpha = 0.05$, 允许误差为 5%, 试计算诊断试验的患者和非患者各需要多少例样本?

软件操作: 分别选择【sample size】【confidence interval estimation & precision】【single proportion】菜单, 打开 sample size: CI for single proportion 对话框, 在【confidence level (%)】中输入 95, 在【proportion (%)】

* 基金项目: 湖南省卫生健康委 2023 年度卫生科研课题 (D202312017098)

1. 长沙市疾病预防控制中心 (410004)

2. 中南大学湘雅三医院护理部

3. 中南大学湘雅三医院消化内科

△通信作者: 舒眉, E-mail: 306517494@qq.com

中输入 90,在【confidence interval width(2-sided)】:输入置信区间总宽度,本例为 10。然后点击【calculate】按钮,即可直接获得患者数量;重复上面的步骤,在【proportion(%)】中输入 85,即可获得非患者数量。

分析结果:开展本诊断试验研究需要患者数量为 141 人,非患者数量为 196 人,共需要检查样本为 337 人。

(2)两个诊断试验比较的样本量估计

【例 2】预计某种诊断试验的灵敏度为 90%,特异度为 85%,另外一种诊断试验的灵敏度为 80%,特异度为 75%, $\alpha=0.05,\beta=0.1$,试计算诊断试验的病例组和对照组各需要多少例样本。

软件操作:分别选择【sample size】【comparison of two proportions】菜单,打开采样:comparison of proportions 对话框,在【type I error(Alpha,significance)】中输入 0.05,在【type II error(Beta,1-power)】中输入 0.10,在【proportion in group 1(%)】中输入 90,在【proportion in group 2(%)】中输入 80,在【ratio of sample size in group 1/group 2】中输入 1。然后点击【calculate】按钮,即可直接获得患者数量;重复上面的步骤,在【proportion in group 1(%)】中输入 85,在【proportion in group 2(%)】中输入 75,即可获得非患者数量。

分析结果:开展两个诊断试验研究的比较,每个诊断试验分别需要患者数量为 266 人,非患者数量为 335 人,每个诊断试验需要检查样本为 601 人。

2.按照 ROC 曲线下面积估计样本量

(1)单个诊断试验 ROC 曲线下面积的样本量估计

【例 3】某课题组研究糖化血红蛋白(HbA_{1c})含量对糖尿病的诊断价值,通过预实验,糖化血红蛋白(HbA_{1c})含量诊断糖尿病 ROC 曲线下面积为 0.8025, $\alpha=0.05,\beta=0.1$,试计算诊断试验需要多少例样本?

软件操作:分别选择【sample size】【Area under ROC curve, AUC】菜单,打开采样:Area under ROC curve, AUC 对话框,在【type I error(Alpha,significance)】中输入 0.05,在【type II error(Beta,1-power)】

中输入 0.10。在【Area under ROC curve】中输入 0.8025,在【Null hypothesis value】中输入 0.5,在【Ratio of sample sizes in negative/positive groups】中输入 2。然后点击【calculate】按钮,即可获得患者和非患者的数量。

分析结果:开展本次诊断试验研究,分别需要患者数量为 13 人,非患者数量为 26 人,每个诊断试验需要检查样本为 39 人。

(2)两独立样本诊断试验 ROC 曲线下面积比较的样本量估计

【例 4】已知某公司的某型号的 MRI 诊断某病的 ROC 曲线下面积为 0.8265,该公司发明了另外新型号的 MRI,需要评价该型号的检查效能,经过预实验,该新型号的 MRI 诊断该病的 ROC 曲线下面积为 0.9024, $\alpha=0.05,\beta=0.1$,试计算诊断试验需要多少例样本?

软件操作:分别选择【sample size】【Area under ROC curve, AUC】菜单,打开采样:Area under ROC curve, AUC 对话框,在【type I error(Alpha,significance)】中输入 0.05,在【type II error(Beta,1-power)】中输入 0.10。在【Area under ROC curve】中输入 0.9024,在【Null hypothesis value】中输入 0.8265,在【Ratio of sample sizes in negative/positive groups】中输入 2。然后点击【calculate】按钮,即可获得患者和非患者的数量。

分析结果:开展本次诊断试验研究,分别需要患者数量为 115 人,非患者数量为 230 人,每个诊断试验需要检查样本为 345 人。

MedCalc 在连续性资料的 ROC 曲线分析中的应用

【例 5】以颜艳,王彤主编的《医学统计学》第 5 版的例 33-6 为例^[10],为评价红细胞平均容积(mean corpuscular volume, MCV)对缺铁性贫血患者的诊断价值,以 100 例可疑为缺铁性贫血患者作诊断,以骨髓诊断为金标准,获得了表 1 的原始数据。试确定 MCV 截断值和绘制 ROC 曲线。

表 1 100 名受试者的红细胞平均容积

骨髓诊断	红细胞平均容积结果
正常 g=0	60 66 68 69 71 71 73 74 74 74 76 77 77 77 77 78 78 79 79 80 80 81 81 81 82 82 83 83 83 83 83 83 83 84 84 84 84 85 85 86 86 86 87 88 88 88 89 89 90 90 91 91 92 93 93 93 94 94 94 94 96 97 98 100 103
异常 g=1	52 58 62 65 67 68 69 71 72 72 73 73 74 75 76 77 77 78 79 80 80 81 81 81 82 83 84 85 85 86 88 88 90 92

软件操作如下。

(1)建立并打开数据库。数据库包括 2 列(分别为 g,MCV)100 行。

(2)依次选择【statistics】【ROC curve】【ROC curve analysis】菜单,打开 ROC 曲线分析对话框。将“红细胞平均容积”置入【variable】,将“分组 g”置入

【classification variable】。在【methodology】中选择计算 A 的标准误,推荐使用【DeLong et al】计算 A 的标准误。在【options】中分别选择【list criterion values with test characteristics】【include all observed criterion values】,在【95% confidence interval】中选择【sensitivity/specificity】【likelihood ratios】【predictive values】。在

【graphs】中选择【display ROC curve window】【mark points corresponding to criterion values】【include 95% confidence bounds】。ROC 曲线分析对话框见图 1。

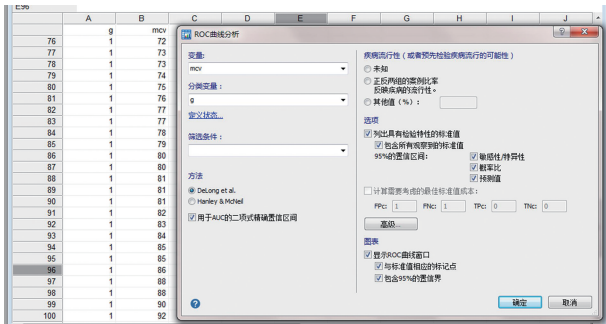


图 1 ROC 曲线分析对话框

交互点图汇制:交互点图可以用于研究诊断试验的准确性。图中的水平线表示截断值,右侧可以显示其灵敏度和特异度。交互点图软件操作如下。

(1) 打开上面的数据库。

(2) 依次选择【statistics】【ROC curve】【interactive dot diagram】菜单,打开【interactive dot diagram】对话框,将“红细胞平均容积”置入【variable】,将“分组 g”置入【classification variable】。

分析结果:①ROC 曲线下面积: $A=0.717$,95% CI: 0.618~0.803, $Z=4.126$, $P<0.001$,有统计学意义。② Youden 指数,临床上通常以 Youden 指数来确定截断值,本例 Youden 指数最大值为 0.3422,95% CI: 0.1899~0.4870,对应的 MCV 截断值为 ≤ 81 fl(95% CI: 72~86)。即当 $MCV \leq 81$ fl 时,即可考虑诊断为缺铁性贫血。③ROC 曲线,在曲线的左上角的点可以显示其灵敏度、特异度和截断值。当为 $MCV \leq 81$ fl 时,灵敏度为 70.6%(95% CI: 52.5%~84.9%)、特异度 63.6%(95% CI: 50.9%~75.1%),见图 2。④交互点图结果:图中的水平线为对应的截断值($MCV \leq 81$ fl),灵敏度为 70.6%、特异度 63.6%。水平线下面的非患者组个体为假阳性,水平线上方的患者组个体为假阴性,见图 3。

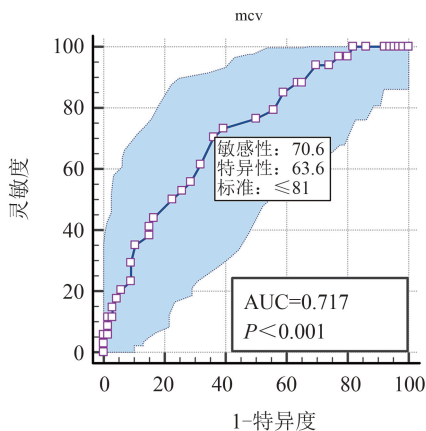


图 2 ROC 曲线

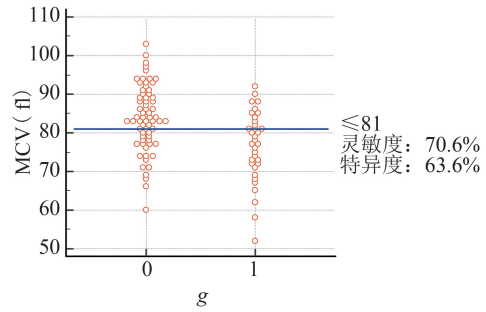


图 3 交互点图

MedCalc 在 ROC 曲线比较中的应用

ROC 曲线分析是评价指标诊断能力的重要方法。可通过 ROC 曲线下面积比较不同指标诊断疾病的价值高低。

1. 相关样本(非独立样本)ROC 曲线比较

【例 6】某课题组检测 3 个连续性诊断指标对同一批(20 人)病人进行诊断,比较 3 个指标诊断该疾病的价值,获得表 2 的数据,试比较 3 个诊断指标的准确性差异。

表 2 不同诊断指标检测结果

指标 1(X1)	指标 2(X2)	指标 3(X3)	疾病(D)
82.7	5.4	26.2	0
87.0	5.3	25.1	1
112.6	4.9	23.4	1
93.1	5.1	20.5	0
96.1	5.4	25.9	0
108.1	4.7	24.7	1
89.2	6.2	22.7	0
102.3	5.6	23.5	0
94.5	5.2	24.3	0
83.2	5.1	23.1	1
103.1	4.8	24.8	0
94.4	5.6	27.6	0
97.3	5.5	25.6	0
98.8	4.5	24.2	1
85.5	6.1	26.1	0
105.5	5.4	20.8	0
101.3	5.8	22.3	0
97.6	4.6	23.5	1
88.4	4.7	21.9	0
112.1	5.1	20.8	0

软件操作如下。

(1) 建立并打开数据库。数据库包括 4 列(分别为 X1~X3,D)20 行。

(2) 依次选择【statistics】【ROC curve】【comparison of ROC curves】菜单,打开 comparison of ROC curves 对话框。在【variable】中分别写入 X1,X2,X3,在【classification variable】中写入 D,在【methodology】

中选择【Hanley & McNeil】【binomial exact confidence interval for the AUC】,在【graphs】中选择【display ROC curve window】。

分析结果:①三条 ROC 曲线下面积:指标 1 的 AUC 为 0.571(95% CI: 0.334~0.787),指标 2 的 AUC 为 0.875(95% CI: 0.651~0.979),指标 3 的 AUC 为 0.506(95% CI: 0.277~0.733)。②三条 ROC 曲线下面积比较。指标 1 与指标 2 的 AUC 差值为 0.304,95%CI: -0.004~0.611, Z=1.937, P=0.0528;指标 2 与指标 3 的 AUC 差值为 0.369,95%CI: 0.095~0.643, Z=2.638, P=0.0083;指标 1 与指标 3 的 AUC 差值为 0.066,95%CI: -0.288~0.419, Z=0.363, P=0.7164。结果表明指标 2 与指标 3 的 AUC 差异具有统计学意义,但指标 1 与指标 2,指标 1 与指标 3 的 AUC 差异无统计学意义。③3 个指标的 ROC 曲线对比图,显示指标 2 的 AUC 最大,诊断效能最高,其次为指标 1,指标 3 最差。

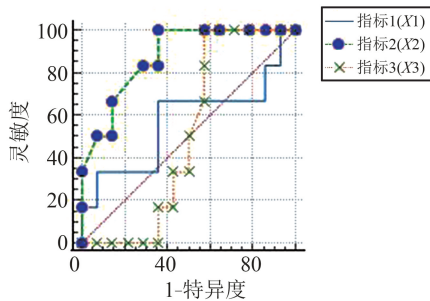


图 4 3 个指标的 ROC 曲线对比图

2. 独立 ROC 曲线比较

【例 7】某研究人员测量损伤 24 h 的脑脊液中某种同工酶含量来预测颅脑严重损伤的结局,分别在年轻患者和老年患者进行了测量,两者的 ROC 曲线下面积分别为 0.7856 和 0.9656,其标准误分别为 0.0776 和 0.0801,试比较该同工酶在年轻患者和老年患者结局的准确性有无差异。软件操作如下。

依次选择【tests】【comparison of】【areas under independent ROC curves】菜单,打开 comparison of independent ROC curves 的对话框,分别在【1st ROC curve】中输入 0.7856,在【standard error】中输入 0.0776,在【2nd ROC curve】中输入 0.9656,在【standard error】中输入 0.0801。点击【test】按钮。

分析结果:两个人群测量的 AUC 差异为 -0.180,差值的标准误为 0.112, Z = -1.614, P = 0.1065。假设检验结果显示还不能认为该同工酶在年轻患者和老年患者结局的准确性存在差异。

2×2 列联表的诊断试验评价指标分析

MedCalc 统计软件可以通过“2×2 列联表的诊断试验”的汇总资料计算灵敏度、特异度、阳性预测值、阴性预测值、阳性似然比、阴性似然比及其 95% 置信区间。

【例 8】以颜艳,王彤主编的《医学统计学》第 5 版的例 33-5 为例^[10],以 Backman 法评价血清总淀粉酶诊断急性胰腺炎的价值,以 CT 检查为金标准,获得表 3 的资料。

表 3 血清总淀粉酶诊断急性胰腺炎的评价

血清总淀粉酶	有急性胰腺炎	无急性胰腺炎	合计
>医学参考值上限	37	14	51
≤医学参考值上限	2	113	115
合计	39	127	166

软件操作如下。

依次打开【tests】【diagnostic test (2×2 table)】菜单,打开 diagnostic test 对话框,分别在四个格子里写入 37、14、2、113 等 4 个数值。然后点击【test】按钮,即可获得相关分析结果。

分析结果:可以获得各种指标的点估计值和 95% 置信区间。总体显示血清总淀粉酶 (Backman 法) 对诊断急性胰腺炎具有良好的诊断价值,见表 4。

表 4 2×2 列联表的诊断试验评价指标分析结果

评价指标	点估计值	95% 置信区间
灵敏度 (%)	94.87	82.68~99.37
特异度 (%)	88.98	82.20~93.84
AUC	0.919	0.867~0.956
阳性概率比	8.606	5.223~14.182
阴性概率比	0.058	0.015~0.223
阳性预测值 (%)	72.55	61.60~81.33
阴性预测值 (%)	98.26	93.60~99.54
准确性 (%)	90.36	84.82~94.39

参 考 文 献

[1] 杨土保,胡国清. 医学科学研究与设计(第 3 版)[M].北京:人民卫生出版社,2020:168-180.

[2] 韩晖,苏玲,潘宝骏.用 SPSS 软件包快速绘制 ROC 曲线并决定敏感性、特异性截断点的方法[J].海峡预防医学杂志,2022,28(2):86-90.

[3] 李太顺,刘沛.ROC 曲线绘制和曲线下面积比较的 SAS 宏包[J].中国卫生统计,2018,35(2):302-309.

[4] 王喜文,董柏青,刘飞鹰.两相关诊断试验的 ROC 曲线下面积比较的 SAS 程序实现[J].数理医药学杂志,2010,23(6):671-674.

[5] 吴研鹏,钱晨坚,陈平雁,等.诊断试验样本量估计及其在 nQuery 和 SAS 软件上的实现[J].中国卫生统计,2021,38(5):790-799.

[6] 李志辉,杜志成.MedCalc 统计分析方法及应用[M].北京:电子工业出版社,2018:239-276.

[7] 丁勇生,夏滢林,王汉杰.Medcalc 软件进行 ROC 曲线分析在影像学中的应用[J].临床医药实践,2011,20(4):275-277.

[8] Hastie T, Venske-Parker S, Aps JKM. Impact of viewing conditions on the performance assessment of different computer monitors used for dental diagnostics[J]. Imaging Science in Dentistry, 2021, 51(2):137-148.

[9] 贺爱民.应用 MedCalc 软件评价 D-二聚体在静脉血栓栓塞症中的作用[J].实验与检验医学,2015,33(4):482-484.

[10] 颜艳,王彤.医学统计学(第 5 版)[M].北京:人民卫生出版社,2020:569-576.

(责任编辑:林美艺)