

影像学诊断试验的双正态模型参数法 ROC 曲线分析

徐绪党¹, 刘 标¹, 杨 伟¹, 王传兵², 孙 晋¹, 李天女^{1*}

(¹南京医科大学第一附属医院核医学科,²放射科,江苏 南京 210029)

[摘要] **目的:**运用受试者工作特征(receiver operating characteristic, ROC)曲线的双正态模型参数法分析影像学诊断试验资料。**方法:**选择影像学上常见的单个诊断试验的连续性资料和两个完全相关诊断试验的有序分类资料进行分析;前者选取 49 例胃肿瘤性病变患者,进行常规 MRI 及弥散加权成像(diffusion-weighted image, DWI),分析胃良性和恶性占位之间的表观扩散系数(apparent diffusion coefficient, ADC)值差异,并确定其最佳截断点;后者选取 55 例孤立性肺结节(solitary pulmonary nodule, SPN)的 HRCT 及 PET/CT 资料,比较两成像模式的诊断效能。结合 ROC 分析软件 ROCKIT 行 ROC 分析,参考的“金标准”皆为病理结果,差异性检验的水准 $\alpha = 0.05$ 。**结果:**前者诊断试验获得 ROC 曲线下面积 $Area(A_z) = 0.963 5$, ROC 曲线相关参数 $a = 2.347 6$, 相关参数 $b = 0.844 7$, 约登指数最大值所对应的最佳截断点为 1.915;后者诊断实验 HRCT 与 PET/CT 两诊断模式所获得的 ROC 曲线下面积分别为 0.8127 和 0.959 0, 相关系数为 0.827 9;两者 ROC 曲线下面积的差异性 Z 检验,获得的单尾 P 值为 0.000 1。**结论:**运用双正态模型参数法结合 ROC 曲线分析专用软件 ROCKIT 可以对影像学诊断试验资料进行系统分析,在单个诊断试验连续性资料的最佳工作点的确定及有序分类资料的诊断效能比较上可以实现有效的 ROC 评价。

[关键词] 诊断试验;受试者操作特性曲线;ROCKI 程序;双正态模型

[中图分类号] R445.9

[文献标志码] A

[文章编号] 1007-4368(2014)02-230-04

doi:10.7655/NYDXBNS20140224

ROC curve analysis of dual-normal parameters model for imaging diagnostic tests

Xu Xudang¹, Liu Biao¹, Yang Wei¹, Wang Chuanbing², Sun Jin¹, Li Tiannv^{1*}

(¹Department of Nuclear Medicine, ²Department of Radiology, the First Affiliated Hospital of NJMU, Nanjing 210029, China)

[Abstract] **Objective:**To analyze imaging diagnostic test data using receiver operating characteristic curve (ROC) combined with dual-normal model parameter method. **Methods:**Continuous data of a single common imaging diagnostic test and ordered segment information of two completely diagnostic tests were selected for analysis. The former group consisted of 49 cases of gastric neoplastic lesions was examined by conventional MRI and diffusion-weighted imaging (DWI) to analyze the difference of apparent diffusion coefficient (ADC) between gastric benign and malignant placeholder and determine the best cut-off point;the latter group consisted of 55 cases of Solitary pulmonary nodules (SPN) with HRCT and PET/CT data. Two types of data were analyzed by ROC analysis software (ROCKIT). The "gold standard" of diagnostic data was pathological results of significance ($\alpha = 0.05$). **Results:**Area under the ROC curve(A_z) of the former diagnostic tests was 0.963 5;the parameters a was 2.347 6, and the relevant parameter b was 0.844 7;Youden index corresponding to the maximum cut-off point was 1.915. The area under the ROC curve obtained by the latter diagnostic experiments HRCT and PET/CT were 0.812 7 and 0.959 0, respectively. The correlation coefficient was 0.827 9. Z test of the area under the curves of both group got the one-tailed P value of 0.000 1. **Conclusion:**The dual-normal parameters model combined with ROC curve analysis software ROCKIT could systematically analyze the imaging diagnostic data. ROC evaluation can be achieved in a single diagnostic test for continuous data to determine the optimum operating point and ordered categorical data on the diagnostic performance comparison.

[Key words] diagnostic tests; receiver operating characteristic curve; ROCKIT program; binormal model

[Acta Univ Med Nanjing, 2014, 34(02): 230-233]

[基金项目] 国家自然科学基金青年基金资助(81202032)

*通信作者(Corresponding author), E-mail: litiannv@126.com

ROC 评价方法在影像学的应用价值已得到很多影像学工作者的认可,而且该方法还在不断的研究发展中。高培毅等^[1]在 2000 年首次应用 ROC 分析方法在影像医学上比较 MRI 与 CT 在颅内恶性胶质瘤术后残留组织的检出效能,并陆续发表了多篇涉及 ROC 分析的文章^[2-4]。但笔者发现很多影像学工作者在应用这一方法时往往仅是提及或只给出 ROC 曲线下面积或差异性检验的 *P* 值,较少阐述具体分析方法,而对于双正态参数 ROC 评价方法更是少有涉及,这在一定程度上限制了 ROC 分析在影像学上的应用。考虑到 RCOKIT 软件属于参数法 ROC 分析专用软件,且分析功能强大,影像学工作者应当熟悉它的应用方法以便更加便捷有效地分析影像资料。笔者选择影像学常见的两种类型资料运用 ROCKIT 软件加以具体分析,现予以总结。

1 资料和方法

1.1 资料

资料 1: 回顾性分析 49 例胃肿瘤性病变患者 MR 原始资料,男 39 例,女 10 例,年龄在 24~81 岁,平均年龄 53 岁。根据良恶性分为两组:恶性肿瘤组 35 例(胃窦癌 22 例,胃底贲门癌 5 例,胃体癌 4 例,全胃癌 2 例,幽门管癌 1 例,胃底癌 1 例);良性肿瘤组 14 例(包括胃间质瘤 8 例,胃平滑肌瘤 3 例,胃管状腺瘤 3 例)。

资料 2: 回顾性收集 55 例经病理证实的肺结节(SPN)患者,其中男 34 例,女 21 例,年龄 34~78 岁,平均年龄 58 岁。其中恶性肿瘤 30 例(其中鳞癌 17 例,腺癌 8 例,肺泡细胞癌 3 例,肺淋巴瘤 2 例);良性肿瘤 25 例(其中结核瘤 14 例,炎性假瘤 5 例,支气管肺炎 3 例,错构瘤 2 例,机化性肺炎 1 例)。所有患者均调阅同时期的 HRCT 及 ¹⁸F-FDG PET/CT 影像资料。

1.2 方法

1.2.1 DWI 检查

采用单次激发 EPI(single shot,EPI)序列,扩散敏感系数 *b* 值为 800 s/mm²。其余成像参数相同,TR 1 000 ms; TE 73.4 ms;NEX 4 次;层厚 5 mm;层间距 1 mm;FOV 38 cm × 38 cm;矩阵 128 × 128;同时在 *x*、*y*、*z* 轴 3 个方向施加敏感梯度脉冲。根据患者耐受屏气的情况,分 2~3 次屏气完成扫描,尽量使屏气基线一致。

1.2.2 PET/CT 检查

正电子药物 ¹⁸F-氟脱氧葡萄糖(¹⁸F-FDG)由美

国 GE 公司研制的回旋加速器 PETtrace 通过 ¹⁸O(*p*, *n*)¹⁸F 核反应生产 ¹⁸F 同位素,然后利用自动合成模块 TracerlabFX-FDG 自动控制合成,药物的放射化学纯度 > 99%。显像仪器为德国 Siemens 公司生产的 Biograph 16 HRPET/CT 扫描仪。所有患者均禁食 6 h 以上,注药前空腹血糖 ≤ 7.0 mmol/L,按 3.70~5.55 MBq/kg 静脉注射 ¹⁸F-FDG,安静状态下药物在体内分布 60 min 后,进行全身 PET/CT 检查。应用 CT 数据进行衰减校正,迭代法重建,层厚 5 mm,最终获得横断、矢状、冠状面 CT、PET 及 PET/CT 融合图像。PET/CT 检查时嘱患者平静呼吸,检查后行常规胸部 CT 屏气扫描。

1.2.3 HRCT 检查

使用德国 SIEMENS 公司 SOMATOM Sensation 64 多层螺旋 CT 机。患者取仰卧位头先进扫描。先进进行常规扫描,确定兴趣区,再运用 HRCT 扫描技术对兴趣区进行扫描。HRCT 扫描参数:电压 120 kV,电流 200 mA,层厚 1.2 mm,螺距为 1.0,高分辨率重建算法重建。所有采集的数据经工作站进行薄层重建;重建层厚 1 mm,重建间隔 0.7 mm。

1.2.4 资料的分组

依据上述检查方法获得的影像学资料,均由本专业 2 位高年资医师联合诊断定性,具体定性方法参考相关文献^[5-6]。将定性结果对照“金标准”,即病理学结论分组统计,以便运行 ROC 分析软件 RCOKIT。对于 MR 资料根据获得的 ADC 值只需对应“金标准”顺序列出即可(表 1)。对于 PET/CT 与 HRCT 资料的分组实行分级统计,即按肯定良性、可能良性、不确定、可能恶性、肯定恶性进行五级分类,级数越高代表恶性可能越大;分组统计时将良性结节与恶性结节分别置于两诊断模式下比较,依据定性判断结果逐级统计(表 2)。

表 1 胃肿瘤 ADC 值与病理对照

Table 1 Comparison of stomach neoplasms' ADC value and pathology

病理结果	ADC 值
良性肿瘤	2.12, 1.98, 2.41, 1.95, 2.63, 2.48, 2.36, 1.91, 1.96, 2.54, 2.35, 1.90, 2.30, 2.20
恶性肿瘤	1.64, 1.65, 1.90, 1.87, 2.01, 1.95, 1.78, 1.69, 1.90, 1.85, 1.75, 1.76, 2.00, 1.68, 1.69, 1.72, 1.78, 1.64, 1.92, 1.89, 1.93, 1.78, 1.85, 1.67, 1.84, 1.86, 1.89, 1.90, 1.87, 1.86, 1.79, 1.88, 1.67, 1.69, 1.62

1.2.5 分组资料的 ROC 分析

将上述两种已经分组处理的资料行 ROC 分析,此处利用 ROC 分析软件 ROCKIT1.1 B2, 该软件的

表2 PET/CT与HRCT对SPN的定性结果
Table 2 Qualitative results of SPN PET/CT and HRCT (n)

PET/CT 分类	HRCT 分类											
	良性结节						恶性结节					
	良性	可能良性	不确定	可能恶性	肯定恶性	合计	良性	可能良性	不确定	可能恶性	肯定恶性	合计
良性	12	4	2	0	0	18	1	0	0	0	0	1
可能良性	0	0	1	4	0	5	1	0	0	0	0	1
不确定	0	0	0	0	0	0	0	2	0	0	0	2
可能恶性	0	0	0	0	1	1	0	1	4	2	0	7
肯定恶性	0	0	0	0	1	1	0	0	0	4	15	19
合计	12	4	3	4	2	25	2	3	4	6	15	30

分析方法属于拟合的双正态模型参数法。由于该软件的编程界面使用的是FORTRAN语言,不利于修改,使用时最好先用记事本编写文本文件再启动程序调用。对于表1资料,分析目的是运用ADC值鉴别胃部肿瘤良恶性时的最佳截断点;对于表2资料,分析目的是确定两种诊断模式对SPN的诊断效能。两资料需将相应频数逐一录入文件,正确格式录入是顺利运行软件的关键步骤(表3)。

表3 ROCKIT软件处理两类资料的格式

Table 3 Two kinds of data formats for ROCKIT software

表1 资料/连续性资料	表2 资料/有序分类资料	
ADC值鉴别胃部肿瘤	PET/CT与HRCT鉴别SPN	
KIT	KIT	
“ADC”	“PET/CT”	“HRCT”
CSL	DNL	DNL
2.12	1	1
1.98	1	1
...
2.30	4	5
2.20	5	5
*良性组结束	*良性结节结束	
1.64	1	1
1.65	2	1
...
1.69	5	5
1.62	5	5
*恶性组结束	*恶性结节结束	

两类资料的程序首行是对该资料的简要描述;第二行KIT确定分析该资料运用ROCKIT;表1资料录入文本的第三行用来说明需要具体分析指标名称;“CSL”中的“C”表示连续型数据,“S”表示数值越小提示阳性,“L”表示以列方式输入资料;其后的数据录入依照表一的顺序依次录入,良性组与恶性组结束后的尾行首字符必须是“*”,否则程序无法运算。表2资料录入文本的第三行说明需要比较的两种诊断模式或方法;“DNL”中的“D”表示离散型资料,“N”表示分组的级数越大恶性(阳性)可能越大,“L”表示以列方式输入资料;自第五行起连续输入12个“11”,再连续输入4个“12”,接着输入1个“23”,输完良性结节组再以同样的方法输入恶性结节组,良性组与恶性组结束后的尾行首字符同样是“*”。

根据表3建立的备用文件,启动ROCKIT软件,进入初始界面后依据系统提示调用即可完成分析。

2 结果

2.1 连续性诊断试验资料(资料1)的分析结果

资料1根据MRI资料的ADC值来鉴别胃部肿瘤的良恶性,主要目的是找出诊断的最佳界值,一般取约登指数最大时的对应点作为最佳判断界值。相应的参数估计包括患者组与非患者组试验结果的标准化均数之差a;非患者组与患者组试验结果的标准差之比b;ROC曲线下面积Az;以及各参数估计的标准误差及95%可信区间(表4)。

表4 资料1的ROC分析结果

Table 4 ROC analysis results of Data 1			
指标	a	b	Az
估计值	2.347 6	0.844 7	0.963 5
95%CI	(1.082 8, 3.612 5)	(-0.076 7, 1.766 1)	(0.882 1, 0.991 8)
标准误差 SE	0.645 3	0.470 1	0.024 8

根据Az值(> 0.9)可知依靠ADC值评价胃部肿瘤的良恶性价值较高。结果显示当约登指数最大时所对应的灵敏度为0.887,特异度为0.911,此时ADC值为1.915,即ADC评价胃部良恶性肿瘤的最佳界值为1.915。

2.2 有序分类资料(资料2)的分析结果

PET/CT的ROC分析相关参数a = 2.865 5, b = 1.309 6, Area(Az) = 0.959 0;HRCT的ROC分析相关参数a = 1.277 7, b = 1.034 6, Area(Az) = 0.812 7。两曲线下面积的一致性检验的相关系数为0.827 9,说明两者的一致性较好;差异性Z检验统计量Z = -3.645 8,单尾P = 0.000 1(由于该资料主要目的是评价PET/CT与HRCT在SPN的诊断效能,最直观的比较方法就是比较两者ROC曲线下的面积)。可以认为在SPN的定性诊断效能上PET/CT较HRCT有明显优势。

3 讨论

ROC 曲线分析分非参数法和参数法两种。非参数法根据实验结果直接计算绘制 ROC 曲线所需的工作点(假阳性率、真阳性率),并借此绘制 ROC 曲线,此曲线称为经验 ROC 曲线,曲线下面积可由梯形规则计算得到;非参数法对诊断试验结果(患者与非患组)的资料分布没有特殊要求,因此,使用的范围比较广泛。目前在临床科研中使用统计软件 SPSS 进行的 ROC 曲线分析就属于非参数估计法。而参数法是假设诊断试验结果资料均服从正态分布,再根据试验结果拟合双正态模型,进而获得拟合 ROC 曲线^[7-9](光滑 ROC 曲线)。对于影像学诊断试验资料的 ROC 曲线分析目前应用较多的也是使用统计学软件 SPSS,但是有研究证明在诊断试验结果相同值较多及样本量较小时,ROC 曲线下面积计算结果常小于真实的面积^[10]。虽然参数法可以避免上述情况但同时也要注意其适用条件,即在使用时避免样本量过小或不符合正态分布的情况,以避免分析结果出现过大的偏倚。

资料 1 的分析显示运用 MRI 常规结合 DWI 可以对胃部肿瘤样病变作定性诊断,至于诊断的截断点可能会因样本含量及扫描参数的差异有所变化,但对于总体而言,这种差异不具备统计学意义;资料 2 的分析显示 ¹⁸F-FDG PET/CT 显像在 SPN 的定性诊断上的确优于单纯 HRCT 检查,因此,HRCT 检查可疑的 SPN 患者可以考虑进一步的 PET/CT 检查。对于这种诊断效能比较的影像资料人们往往只注重两者差异性而忽略了两者的 consistency,通常情况下需要对两者进行一致性检验,进而再进行差异性检验,故资料 2 给出了两者曲线下面积的相关系数,对于一致性不佳的两种诊断试验进行效能比较是没有什么临床意义的。

在应用参数法 ROC 分析影像资料时首先要考虑样本含量,最好是待分析的样本大于 30 例,具体分析时先确定资料类型,再根据资料类型进行资料整理和分组,对于定量影像资料多为本文资料 1 的类型,有时也涉及 2 个及以上诊断试验比较;对于有序分类资料,除本文介绍的以外也会涉及单个诊断试验分析。具体方法可参见 ROCKIT 使用说明。需要注意的是在涉及 2 个及以上诊断试验比较时,应首先明确资料设计属于分组资料(独立诊断试验)还是配对资料(相关诊断试验),因为这两种设计资料在

编辑 ROCKIT 程序时稍有差异,如果混淆必定出现错误的结果。在实践分析工作中如果样本含量较大且相同值不多,可以选择参数法或者非参数法进行 ROC 分析,样本量较小时选择非参数法。

值得提及的是在运用双正态模型参数法分析时,对于连续性资料的正态分布属性一般没有异议,但对于等级资料或有序分类资料的统计描述上明确界定其分布属性且套用正态模型来分析似乎不妥。鉴于此,有学者提出对于等级资料人们可以将其视作是来自某个潜在连续分布资料的粗分类^[11]。在这个意义上讲,使用正态模型参数估计也是完全可行的。

[参考文献]

- [1] 张红梅,高培毅. 颅内恶性胶质瘤术后早期 MR 影像学表现[J]. 中华放射学杂志,2000,34(2):90-94
- [2] 卢传友,柳澄,秦维昌. ROC 评价法在临床诊断中的应用价值[J]. 医学影像学杂志,2004,14(9):757-759
- [3] 郑敏,蔡卫民,翁红雷,等. ROC 曲线评价血清纤维化指标对诊断肝纤维化的价值 [J]. 中华传染病杂志,2002,20(4):225-228
- [4] 王晓华,陈卉,马大庆,等. 人工神经网络在孤立性肺结节 CT 诊断研究中的应用 [J]. 中华放射学杂志,2006,40(4):377-382
- [5] Christensen JA, Nathan MA, Mullan BP, et al. Characterization of the solitary pulmonary nodule: 18F-FDG PET versus nodule-enhancement CT [J]. Am J Roentgenol, 2006, 187 (5):1361-1367
- [6] Avcu S, Arslan H, Unall O. The role of diffusion-weighted MR imaging and ADC values in the diagnosis of gastric tumors[J]. JBR-BTR, 2012, 95(1):1-5
- [7] Erkanli A, Sung M, Jane Costello E. Bayesian semi-parametric ROC analysis[J]. Stat Med, 2006, 25 (22):3905-3928
- [8] Wang MC, Li SS. Bivariate Marker Measurements and ROC Analysis[J]. Biometrics, 2012, 68(4):1207-1218
- [9] Foucher Y, Giral M, Souillou JP, et al. Time-dependent ROC analysis for a three-class prognostic with application to kidney transplantation [J]. Stat Med, 2010, 29 (30):3079-3087
- [10] Cai T, Moskowitz CS. Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test[J]. Biostatistics, 2004, 5 (4):573-586
- [11] Huang Y, Pepe M. A Parametric ROC model-based approach for evaluating the predictiveness of continuous markers in case-control studies [J]. Biometrics, 2009, 65 (4):1133-1144

[收稿日期] 2013-05-29