



· 专题论著 ·



陈卫国，主任医师，教授，博士研究生导师，南方医科大学南方医院放射科主任。中华医学会放射学分会乳腺学组委员、广东省肝脏病学会影像诊断专业委员会主任委员、广东省医院协会医学影像中心管理专业委员会副主任委员、广东省医学会放射医学分会委员及乳腺学组副组长、中国乳腺癌风险防控及早诊早治联盟常委、中国医学影像整合联盟第一届理事会理事兼乳腺与女性生殖专委会常委、广东省医师协会放射科医师分会常委及乳腺专业组组长、中国医药教育协会理事及专家委员会委员。目前研究方向为数字化放射诊断和乳腺疾病的综合影像学诊断，对泌尿生殖系统和乳腺X线诊断有丰富的临床经验。

基于临床信息的logistic回归模型在乳腺影像报告和数据库4类中对病灶良恶性的鉴别价值

林晓佳，马 乐，蔡裕兴，陈卫国

南方医科大学南方医院放射科，广东 广州 510515

[摘要] 目的：探讨基于患者临床信息的logistic回归模型在乳腺影像报告和数据库（Breast Imaging Reporting and Data System, BI-RADS）4类中鉴别病灶良恶性的价值。方法：回顾并收集经过病理学检查证实的BI-RADS 4类乳腺病灶患者221例（良性133例，恶性88例）的临床信息。采用logistic回归分析筛选能够鉴别病灶良恶性的临床信息特征，建立回归模型。比较BI-RADS联合模型与单独采用BI-RADS分类在鉴别乳腺良恶性病灶上的区别。结果：经logistic回归分析，发现9个临床信息特征与乳腺良恶性病灶相关，其中是否触及病灶（OR=7.196）、病灶是否固定（OR=10.150）、病灶最大径是否>2 cm（OR=4.208）等3个特征有较高的危险度（ $P<0.05$ ）。单独采用BI-RADS分类，其诊断灵敏度为86.3%、特异度为69.9%、准确率为76.5%；将BI-RADS分类联合回归模型，其灵敏度为88.6%、特异度为73.7%、准确率为79.6%。结论：BI-RADS分类联合基于患者临床信息的logistic回归模型有助于提高鉴别乳腺病灶良恶性的诊断效能，减少不必要的良性活检。

[关键词] 乳腺病灶；良恶性；logistic分析；鉴别模型

DOI: 10.19732/j.cnki.2096-6210.2020.02.004

中图分类号：R737.9；R445.4 文献标志码：A 文章编号：2096-6210(2020)02-0085-05

Logistic regression models based on clinical information in discriminating breast malignant lesions from benign lesions of Breast Imaging Reporting and Data System 4 LIN Xiaojia, MA Le, CAI Yuxing, CHEN Weiguo (Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong Province, China)

Correspondence to: CHEN Weiguo E-mail: chenweiguo1964@21cn.com

[Abstract] **Objective:** To explore the value of logistic regression model based on subjects' clinical information in discriminating breast malignant lesions from benign lesions of Breast Imaging Reporting and Data System (BI-RADS) 4. **Methods:** Retrospectively 221 subjects (133 benign and 88 malignant) confirmed by histopathology were recruited whose BI-RADS grade was 4 and the

基金项目：广东省自然科学基金（2019A1515011168）；广东省医学科学技术研究基金（B2018017）

通信作者：陈卫国 E-mail: chenweiguo1964@21cn.com

clinical information were collected. Logistic regression analysis was used to screen the clinical information features that can discriminate malignant from benign lesions and a regression model was established. The comparison was made between regression model combined with BI-RADS and BI-RADS classification alone for differential diagnosis between malignant and benign lesions.

Results: Nine clinical information features were found to be related to malignant and benign lesions. Three features of whether the lesion can be touched (OR=7.196), whether the lesion position was fixed (OR=10.150), and whether the maximum diameter of the lesion was more than 2 cm (OR=4.208) have a higher risk than other clinical information ($P<0.05$). Using BI-RADS classification alone, the diagnostic sensitivity, specificity and accuracy were 86.3%, 69.9% and 76.5%; the diagnostic sensitivity, specificity and accuracy of regression model combined with BI-RADS were 88.6%, 73.7% and 79.6%. **Conclusion:** Logistic regression model based on subjects' clinical information combined with BI-RADS classification is helpful to improve the diagnostic efficiency of malignant and benign lesions and further to reduce unnecessary benign biopsy.

[**Key words**] Breast lesions; Malignant and benign; Logistic regression analysis; Discriminating model

乳腺癌已成为全球女性死亡的主要疾病之一, 其发病率呈逐年上升的趋势^[1-2]。在中国, 乳腺癌高居女性恶性肿瘤发病率之首, 严重影响其身心健康^[3]。因此, 早发现、早诊断、早治疗, 有利于提高乳腺癌的治愈率, 降低死亡率。乳腺X线摄影作为主要的筛查手段, 在早期诊断中发挥着重要作用^[4]。为了更好地指导临床, 方便临床医师对乳腺病灶进行综合性评估, 美国放射学会 (American College of Radiology, ACR) 提出乳腺影像报告和数据系统 (Breast Imaging Reporting and Data System, BI-RADS) 的X线分类标准^[5]。尽管这种分类已广泛用于影像学诊断评估, 但在临床研究发现仍然存在大量的良性活检^[6]。本研究通过回顾性分析, 构建基于临床信息的logistic回归模型并联合BI-RADS分类鉴别乳腺病灶良恶性, 以提高诊断效能, 减少不必要的良性活检。

1 资料和方法

1.1 一般资料

本研究通过了南方医科大学南方医院伦理委员会的批准, 所有受检者均签署知情同意书。回顾并收集2018年7月—2019年12月于南方医科大学南方医院行乳腺X线摄影, 检查报告BI-RADS分类为4类 (包括4A、4B、4C) 且经病理学检查证实的女性患者共221例, 其中良性133例, 恶性88例。所有患者均有完整的临床信息。

1.2 仪器及方法

采用美国Hologic公司的Selenia Dimensions乳腺X线机, 所有患者均行头尾位和内外斜位检查。在进行检查前, 患者需填写个人信息, 包括年龄、初潮年龄、结婚年龄、生育第一胎年龄、是否绝经、哺乳史、生育史、家族史。操作技师填写触诊信息, 包括是否触及病灶、病灶是否固定、病灶最大径是否 >2 cm等。

1.3 统计学处理

采用SPSS 20.0统计分析软件, 对患者年龄、初潮年龄、结婚年龄、生育第一胎年龄等4个临床信息特征行独立样本 t 检验, 对其余7个临床信息特征行 χ^2 检验。采用logistic回归分析检验不同特征与乳腺病灶良恶性之间的关联, 筛选有统计学意义的临床信息; 入选的临床信息根据比值比 (odds ratio, OR), 建立乳腺病灶良恶性评分模型, 并根据模型的临界值调整BI-RADS分类。具体调整方案: ① BI-RADS 4A类病灶, 评分 $<$ 临界值时BI-RADS分类保持不变; 评分 $>$ 临界值时调整为BI-RADS 4B; ② BI-RADS 4B类病灶, 评分 $<$ 临界值时调整为BI-RADS 4A; 评分 $>$ 临界值时调整为BI-RADS 4C; ③ BI-RADS 4C类病灶, 评分 $<$ 临界值时调整为BI-RADS 4B; 评分 $>$ 临界值时保持不变。采用受试者工作特征 (receiver operating characteristic, ROC) 曲线评价BI-RADS联合评分模型及单独采用BI-RADS对乳腺病灶良恶性的诊断效能。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 病理学检查结果

221个乳腺病灶中, 良性病灶133个, 包括乳腺纤维腺瘤40个, 乳腺腺病32个, 纤维囊性病27个, 导管内乳头状瘤16个, 炎性反应肿块8个, 增生性病变5个, 囊肿2个, 隆突性皮肤纤维肉瘤1个, 管状腺瘤1个, 良性分叶状肿瘤1个; 恶性病灶88个, 包括导管内癌71个, 浸润性癌(含小叶及导管成分)14个, 乳头状癌2个, 肌上皮癌1个。

2.2 临床信息比较

除初潮年龄外, 恶性病灶的患者年龄、结婚年龄、生育第一胎年龄均显著高于良性病灶的患者($P < 0.05$)。经 χ^2 检验, 是否触及病灶、病灶最大径是否 > 2 cm、病灶是否固定等3个特征差异有统计学意义(表1)。

表1 患者临床信息特征分析

临床信息	良性 (N=133)	恶性 (N=88)	P值
患者年龄/岁	45.23 ± 9.80	48.32 ± 9.54	0.023
初潮年龄/岁	14.68 ± 2.11	14.73 ± 2.13	0.993
结婚年龄/岁	21.78 ± 6.45	23.70 ± 4.21	0.011
生育第一胎年龄/岁	22.31 ± 7.93	24.62 ± 6.18	0.018
是否绝经			0.176
是	35	43	
否	98	45	
哺乳史			0.467
有	116	82	
无	17	6	
生育史			0.620
有	120	85	
无	13	3	
家族史			0.550
有	5	2	
无	128	86	
是否触及病灶			<0.001
是	86	81	
否	47	7	
病灶最大径是否 > 2 cm			<0.001
是	39	56	
否	94	32	
病灶是否固定			<0.001
是	42	73	
否	91	15	

2.3 logistic回归分析

以经过病理学确诊的乳腺病灶作为因变量(恶性病灶 $Y=1$, 良性病灶 $Y=0$), 经logistic单变量分析, 发现9个临床信息特征差异有统计学意义。其中是否触及病灶($OR=7.196$, $P < 0.001$)、病灶最大径是否 > 2 cm($OR=4.208$, $P < 0.001$)、病灶是否固定($OR=10.15$, $P < 0.001$), 这3个特征有较高的危险度, 2个临床信息特征差异无统计学意义($P > 0.05$, 表2)。将9个差异有统计学意义的临床信息特征进行二元logistic回归分析(表3), 并建立评分模型。计算公式: $Z=0.037 \times$ 患者年龄 $+0.303 \times$ 是否绝经 $+0.092 \times$ 结婚年龄 $-0.006 \times$ 生育第一胎年龄 $+0.481 \times$ 哺乳史 $-0.591 \times$ 生育史 $+0.141 \times$ 是否触及病灶 $+0.743 \times$ 病灶最大径是否 > 2 cm $+1.949 \times$ 病灶是否固定 -4.756 。

表2 单变量logistic回归分析结果

临床信息	OR	95% CI	P值
患者年龄	1.069	1.05~1.09	<0.001
初潮年龄	1.071	0.98~1.17	0.130
是否绝经	3.325	2.18~5.06	<0.001
结婚年龄	1.096	1.05~1.15	<0.001
生育第一胎年龄	1.051	1.02~1.08	<0.001
哺乳史	2.503	1.22~5.13	0.012
生育史	3.049	1.30~7.12	0.010
家族史	0.379	0.10~1.14	0.137
是否触及病灶	7.196	4.26~12.17	<0.001
病灶最大径是否 > 2 cm	4.208	2.81~6.30	<0.001
病灶是否固定	10.150	6.55~15.73	<0.001

2.4 单独采用BI-RADS分类与BI-RADS分类联合评分模型的诊断效能比较

以病理学诊断结果为金标准, BI-RADS 4A判断为良性, BI-RADS 4B、4C判断为恶性。对221例患者单独采用BI-RADS分类诊断, 其诊断乳腺病灶的灵敏度为86.3%、特异度为69.9%、准确率为76.5%, 曲线下面积(area under curve, AUC)为0.781; 采用BI-RADS分类联合评分模型的灵敏度为88.6%、特异度为75.2%、准确率为80.5%, AUC为0.827(图1)。图2A和图3A为拍摄的患者患侧头尾位及内外斜位图像, 通过调整

表 3 二元logistic回归分析的结果

临床信息	回归系数	标准误	OR	95% CI	P值
患者年龄	0.037	0.019	1.038	1.00~1.07	0.047
是否绝经	0.303	0.379	1.354	0.64~2.85	0.424
结婚年龄	0.092	0.039	1.096	1.01~1.18	0.020
生育第一胎年龄	-0.006	0.046	0.994	0.91~1.09	0.898
哺乳史	0.481	0.840	1.617	0.31~8.38	0.567
生育史	-0.591	1.406	0.554	0.04~8.70	0.674
是否触及病灶	0.141	0.391	1.151	0.54~2.48	0.719
病灶最大径是否>2 cm	0.743	0.273	2.102	1.23~3.59	0.006
病灶是否固定	1.949	1.041	7.020	3.93~12.53	<0.001
常量	-4.756	1.041	0.009	-	<0.001

BI-RADS分类, 使得其BI-RADS分类与病理学检查结果 (图2B、图3B) 一致。

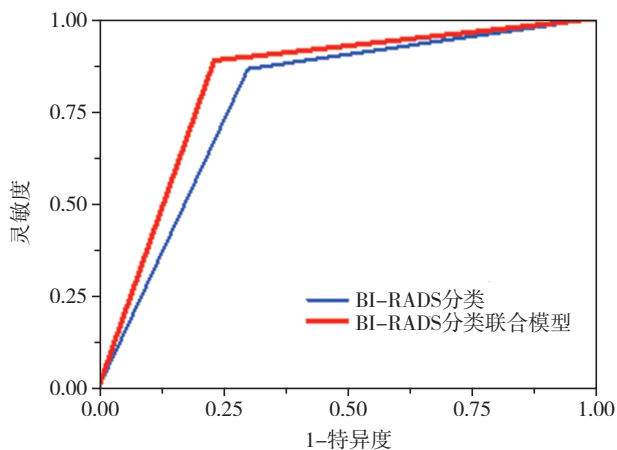


图 1 BI-RADS分类与BI-RADS分类联合模型诊断乳腺良恶性病灶的ROC曲线

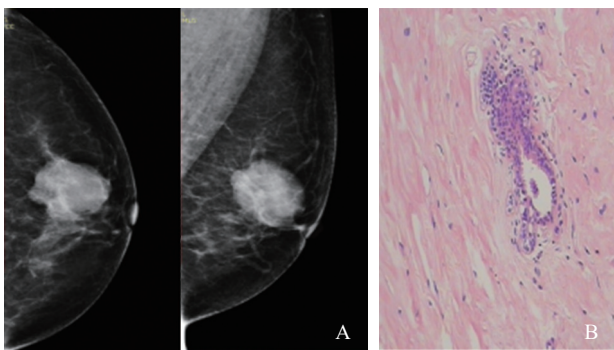


图 2 典型病例1 (女性, 54岁) 的影像学 and 病理学表现

A: X线图像示左侧乳腺肿块形态不规则, 呈分叶状, 边缘部分清晰, 部分模糊。单独应用BI-RADS分类诊断为4B, 模型评分<临界值, 调整为4A类; B: 病理学诊断为乳腺纤维腺瘤 (H-E, $\times 100$)。

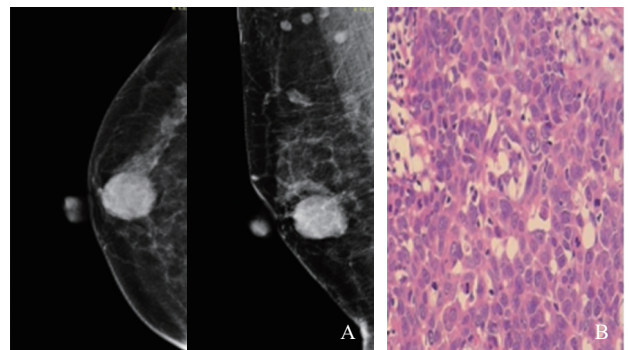


图 3 典型病例2 (女性, 62岁) 的影像学 and 病理学表现

A: X线影像示右侧乳腺肿块呈椭圆形, 形态规则, 边缘清晰。单独应用BI-RADS分类诊断为4A, 模型评分>临界值, 调整为4B类; B: 病理学诊断为乳腺浸润性癌 (H-E, $\times 100$)。

3 讨 论

BI-RADS分类标准已广泛应用于临床, 使乳腺X线摄影报告更加规范, 指导临床医师更加准确地对乳腺病灶做出处理, 同时对乳腺随访检查起到重要作用。然而, 单独应用BI-RADS分类诊断病灶的良恶性会有一些的误判, 特别是对于4类病灶^[7-8]。为避免不必要的活检, 本研究将BI-RADS分类联合基于患者临床信息的logistic回归模型, 用于提高鉴别乳腺病灶良恶性的诊断效能。

经logistic回归分析, 是否触及病灶、病灶是否固定及病灶最大径是否>2 cm这3个临床特征有较高的危险度。Mazouni等^[9]的研究发现, 是否触及肿块这个特征的危险度要高于年龄、是否

绝经、病灶大小等特征，这与本研究结果一致。另外病灶是否固定的危险度最高，无论是单变量logistic回归分析，还是二元logistic回归分析，表明恶性病灶的触诊位置相对固定，而良性病灶的触诊位置多可移动。目前，仍没有文献对这一特征的风险性进行报道，在未来有望作为乳腺癌的高风险因子以及良恶性的鉴别特征。特征病灶最大径是否 $>2\text{ cm}$ 也存在一定的危险度，表明恶性病灶的最大径要大于良性病灶。

研究表明，患者年龄可作为一个危险因子，随着年龄增高，乳腺癌患病风险逐渐增高^[10]。Zheng等^[11]把年龄作为特征用于区分乳腺癌患者与正常人，通过ROC曲线得到AUC为0.633，证明年龄具有一定的区分效果。在本研究中，恶性患者的年龄要显著高于良性患者，且在logistic回归分析中，患者年龄具有风险性。Ahern等^[12]的研究证实，有家族史的患者其乳腺癌患病风险要高于无家族史的患者。而在本研究中，家族史在良恶性患者中没有风险性，可能与恶性病灶的数据量偏少有关。

根据患者的临床信息特征，我们建立了logistic评分模型，联合BI-RADS分类结果，重新对BI-RADS 4类的病灶进行调整，结果表明，BI-RADS联合评分模型较单独应用BI-RADS诊断乳腺病灶的准确率更高。在临床工作中，单独应用BI-RADS分类的灵敏度要高于特异度，说明医师对恶性病灶的判断准确率较高，而在良性病灶上，所给出的BI-RADS分类可能偏高，因此增加了良性活检的次数。而加入本研究所提出的logistic评分模型之后，将7个定义BI-RADS 4B的良性病灶降低为BI-RADS 4A，最后归纳为良性，提高了特异度，有助于降低后续不必要的良性活检。

本研究仍存在不足之处：① logistic回归模型需进一步扩大训练样本量；② 恶性与良性患者的数量比例有待提高。

总之，构建基于临床信息的logistic回归模型并联合BI-RADS分类鉴别乳腺病灶良恶性，有利于提高诊断准确率，减少不必要的良性活检。

[参 考 文 献]

- [1] BRAY F, FERLAY J, SOERJOMATARAM I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2018, 68(6): 394-424.
- [2] SIEGEL R L, MILLER K D, JEMAL A, et al. Cancer statistics, 2017 [J]. *CA Cancer J Clin*, 2017, 67(1): 7-30.
- [3] CHEN W, ZHENG R, BAADE P D, et al. Cancer statistics in China, 2015 [J]. *CA Cancer J Clin*, 2016, 66(2): 115-132.
- [4] MASSAT N J, DIBDEN A, PARMAR D, et al. Impact of screening on breast cancer mortality: the UK program 20 years on [J]. *Cancer Epidemiol Biomarkers Prev*, 2016, 25(3): 455-462.
- [5] D'ORSI C J, MENDELSON E B, MORRIS E. ACR BI-RADS Atlas: breast imaging reporting and data system; mammography, ultrasound, magnetic resonance imaging, follow-up and outcome monitoring, data dictionary [M]. Reston: American College of Radiology, 2013.
- [6] MENDEZ A, CABANILLAS F, ECHENIQUE M, et al. Mammographic features and correlation with biopsy findings using 11-gauge stereotactic vacuum-assisted breast biopsy (SVABB) [J]. *Ann Oncol*, 2004, 15(3): 450-454.
- [7] 吕文豪, 聂芳, 朱阳阳, 等. CEUS评估不同大小BI-RADS 4类乳腺病灶恶性风险 [J]. *中国医学影像技术*, 2019, 35(11): 1673-1677.
- [8] 张松松, 张玉梅, 尹逊娣, 等. 超声弹性成像定量分析诊断BI-RADS 4类乳腺肿块良恶性 [J]. *中国医学影像技术*, 2016, 32(7): 1065-1069.
- [9] MAZOUNI C, SNEIGE N, ROUZIER R, et al. A nomogram to predict for malignant diagnosis of BI-RADS category 4 breast lesions [J]. *J Surg Oncol*, 2010, 102(3): 220-224.
- [10] AMIR E, FREEDMAN O C, SERUGA B, et al. Assessing women at high risk of breast cancer: a review of risk assessment models [J]. *J Natl Cancer Inst*, 2010, 102(10): 680-691.
- [11] ZHENG B, SUMKIN J H, ZULEY M L, et al. Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment [J]. *Eur J Radiol*, 2012, 81(11): 3222-3228.
- [12] AHERN T P, SPRAGUE B L, BISSELL M C S, et al. Family history of breast cancer, breast density, and breast cancer risk in a U.S. breast cancer screening population [J]. *Cancer Epidemiol Biomarkers Prev*, 2017, 26(6): 938-944.

(收稿日期: 2020-01-01 修回日期: 2020-02-16)