



· 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诊断，对泌尿生殖系统和乳腺的X线诊断有丰富的临床经验。

## 基于第5版乳腺影像报告和数据系统方法评价 乳腺密度的一致性研究

徐泽园，曾辉，秦耿耿，徐维敏，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：探讨不同医师对乳腺密度乳腺影像报告和数据系统（Breast Imaging Reporting and Data System, BI-RADS）分类的观察者间一致性及原始报告对乳腺密度分类评估的可靠性。方法：回顾并分析2018年1—5月于南方医科大学南方医院行乳腺X线筛查的774例女性的图像，采用 $\chi^2$ 检验分析不同年龄组筛查女性乳腺X线密度差异。采用Kappa检验分析观察者间及与金标准的一致性水平。结果：以医师评估中占多数的BI-RADS分类结果为金标准，774幅图像中脂肪类乳腺13例，散在纤维腺体类乳腺112例，不均匀致密类乳腺526例，极度致密类乳腺123例。<60岁与≥60岁筛查妇女间乳腺X线密度差异有统计学意义；低（R1）、中（R2）、高（R3）年资医师及原始报告分类准确率分别为81.14%（628/774）、87.86%（680/774）、90.96%（704/774）、67.70%（524/774）；R1与金标准的一致性中等（Kappa=0.602），R2、R3与金标准的一致性较好（Kappa=0.766、0.817），原始报告与金标准的一致性中等（Kappa=0.430）；R1、R2、R3分类的观察者间的总体一致性中等（Kappa=0.671），两两一致性从较差到中等（Kappa=0.396~0.604， $P$ 均<0.001）。结论：乳腺癌筛查妇女年龄与乳腺X线密度有关，不同观察者采用第5版BI-RADS分类乳腺X线密度的一致性一般，原始报告对乳腺密度分类评估的可靠性有限。

**[关键词]** 乳腺密度；乳腺X线摄影；乳腺影像报告和数据系统；结果可重复性

DOI: 10.19732/j.cnki.2096-6210.2020.03.006

中图分类号：R737.9；R445.4 文献标志码：A 文章编号：2096-6210(2020)03-0235-06

**Consistency research on the evaluation of breast density based on the fifth edition Breast Imaging Reporting and Data System** XU Zeyuan, ZENG Hui, QIN Genggeng, XU Weimin, 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 inter-observer consistency in the Breast Imaging Reporting and Data System (BI-RADS) classification of breast density and reliability of classification of breast density assessed by original report. **Methods:** The

基金项目：南方医科大学南方医院临床研究专项（2019CR002）；广东省医学科研基金（B2020095）

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

retrospective study was conducted on 774 women who underwent mammography screening in Nanfang Hospital from Jan. to May. 2018. Chi-square test was used to analyze the differences in mammographic density of screening women among different age groups. The Kappa test was used to analyze the level of consistency between observers and between the observer and the gold standard. **Results:** Using the majority of classification results in radiologist evaluations as the gold standard, of the 774 cases, 13 were the fatty, 112 were the scattered areas, 526 were the heterogeneously dense, and 123 were the extremely dense. There was a statistically significant difference in mammographic density between  $<60$  years and  $\geq 60$  years old women. The accuracy rates of junior (R1), intermediate (R2), senior (R3) radiologist and original report classification were 81.14% (628/774), 87.86% (680/774), 90.96% (704/774), and 67.70% (524/774), respectively; the agreement between R1 and gold standard was moderate (Kappa=0.602); the agreement between R2 and R3 was good (Kappa=0.766, 0.817), and the consistency between the original report and the gold standard was moderate (Kappa=0.430); the overall agreement between the observers was moderate (Kappa=0.671), and the consistency between pairs was from fair to moderate (Kappa=0.396-0.604,  $P<0.001$ ). **Conclusion:** The inter-observer consistency in the BI-RADS classification of mammographic density is moderate, and the reliability of classification of breast density assessed by original report is limited.

[ **Key words** ] Breast density; Mammography; Breast Imaging Reporting and Data System; Reproducibility of results

乳腺癌是女性最常见的恶性肿瘤之一, 乳腺X线摄影是早期发现乳腺癌并降低其死亡率的主要筛查手段<sup>[1]</sup>。致密型乳腺是乳腺癌的独立危险因素之一, 致密型乳腺妇女罹患乳腺癌的概率比非致密型增加4~6倍<sup>[2]</sup>。同时致密型乳腺也是乳腺X线摄影筛查早期乳腺癌灵敏度及特异度偏低的主要因素之一<sup>[3]</sup>。2009年美国康涅狄格州通过的致密乳腺告知法案要求致密型乳腺女性需了解致密乳腺存在的风险因素及辅助筛查手段对致密乳腺的局限, 并选择是否进行补充筛查, 目前美国已有31个州通过了该项立法<sup>[4]</sup>。

2013年美国放射学会提出第5版乳腺影像报告和数据系统(Breast Imaging Reporting and Data System, BI-RADS)乳腺密度分类标准, 建立了4个分类, 分别为脂肪类、散在纤维腺体类、不均匀致密类和极度致密类。该分类标准将不再明确不同乳腺构成分级中高密度组织的构成比例, 乳腺密度直接决定在无钙化病变中致密乳腺纤维组织的遮挡性强弱<sup>[5-6]</sup>。尽管第5版BI-RADS对于预测遮蔽风险可能更为合理, 以更有意义的方式对乳腺密度进行分类。然而去除第4版中的量化百分比值系统, 其主观性更强, 不同医师对同一乳腺实质可能有不同的分类, 且第5版并未像第4版一样有良好的已证实的公认的观察间一致性的分析结果。

本研究针对不同医师采用第5版BI-RADS对同一组患者乳腺密度分类的一致性进行回顾分

析, 探讨其可重复性和原始报告对乳腺密度分类评估的可靠性, 以期更准确地判读致密型乳腺, 指导进一步的补充筛查工作。

## 1 资料和方法

### 1.1 一般资料

2018年1—5月在南方医科大学南方医院行乳腺X线摄影的女性患者共774例。患者年龄23~81岁, 中位年龄47岁。阅片前随机收集每位筛查女性的一侧头尾位乳腺图像, 共收集774幅图像。纳入标准: ① 无临床症状的健康查体女性; ② 乳腺X线筛查图像未发现明确阳性征象; ③ 标准摄影体位图像。排除标准: ① 乳房假体植入术后/注射式隆胸术后; ② 乳腺肿物切除术后/放疗术后; ③ 有临床症状/影像阳性征象者; ④ 附加体位图像。

### 1.2 影像学检查方法

采用美国Hologic公司的Selenia数字化乳腺X线机, 拍摄双侧乳腺头尾位和内外侧斜位。标准化压迫乳房, 采用自动曝光模式。

### 1.3 图像分析

女性双侧乳腺不同体位的乳腺密度类型基本一致, 为了减少内外侧斜位胸大肌对乳腺密度评估的影响, 本研究选取了每例患者的单侧头尾位片进行阅片试验。由3名不同年资的乳腺影像诊断医师R1、R2、R3(年资分别为2、5、10年,

代表低、中、高年资)在不知患者年龄及临床资料的情况下参照BI-RADS分类标准于乳腺诊断专用工作站高分辨5M显示器上独立分析图像,评估乳腺密度类型。乳腺X线密度评估目前的标准是放射科医师依据BI-RADS分类标准进行判定,该结果依赖于放射科医师的主观判断。为了减少这种主观差异,本研究以3名医师中占多数的分类结果为金标准,图像原始的报告分类以下记为原始报告。本研究以金标准分类依据,分析金标准与3名不同年资医师、原始报告之间以及医师

间的分类一致性。

美国放射学会BI-RADS乳腺密度分类:a脂肪类,几乎全部是脂肪组织(图1A);b散在纤维腺体类,乳腺内散在纤维腺体密度区域(图1B);c不均匀致密类,乳腺组织密度不均,可能使小的肿块被遮挡(图1C~D);d极度致密类,乳腺组织极其致密,使X线检查灵敏度降低(图1E)。将a和b类乳腺统一定义为非致密型乳腺,c和d类乳腺统一定义为致密型乳腺。

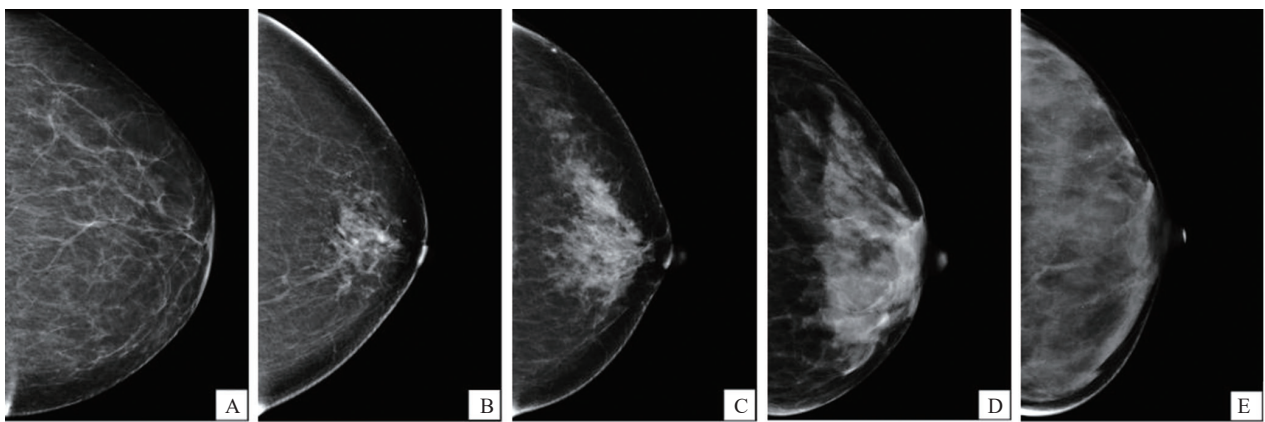


图1 第5版BI-RADS乳腺密度分类示意图

A: 脂肪类; B: 散在纤维腺体类; C~D: 不均匀致密类; E: 极度致密类。

#### 1.4 统计学处理

采用SPSS 20.0软件进行统计分析。采用 $\chi^2$ 检验分析不同年龄组乳腺癌筛查女性乳腺X线密度差异。采用Kappa系数来评估观察者间的分类一致性;  $Kappa < 0.40$ 为一致性较差;  $0.40 \leq Kappa < 0.75$ 为一致性中等;  $Kappa \geq 0.75$ 为一致性较好,  $P < 0.05$ 为差异有统计学意义。

## 2 结果

### 2.1 不同年龄组乳腺X线密度分布情况

以金标准为分类依据, 774幅图像中83.85% (649/774) 女性乳腺呈致密型, 16.15% (125/774) 女性乳腺呈非致密型。不同年龄乳腺癌筛查女性乳腺X线密度的分布情况见表1, 乳腺X线密度比例差异在60岁发生改变,  $< 60$ 岁与 $\geq 60$ 岁女性间乳腺X线密度差异有统计学意义 [非致密型: 11.85% (82/692) vs 52.44%

(43/82), 致密型: 88.15% (610/692) vs 47.56% (39/82);  $P < 0.001$  ]。

表1 不同年龄乳腺癌筛查女性乳腺X线密度的分布情况

年龄/岁	[ n (%) ]	
	非致密型	致密型
$\leq 30$ (n=5)	0 (0.00)	5 (100.00)
30~39 (n=87)	5 (5.75)	82 (94.25)
40~49 (n=370)	27 (7.30)	343 (92.70)
50~59 (n=230)	50 (21.74)	180 (78.26)
60~69 (n=64)	32 (50.00)	32 (50.00)
$> 69$ (n=18)	11 (61.11)	7 (38.89)
合计	125 (16.15)	649 (83.85)

### 2.2 BI-RADS乳腺X线密度分类结果

各观察者与原始报告对乳腺X线密度的BI-RADS分类结果见表2。以金标准为分类依据, R1与原始报告分别有45、96例将致密型图像评估为非致密型, R3有20例将非致密型图像评估为致密型。

表2 各观察者与原始报告对乳腺X线密度的BI-RADS分类结果

(n)

项目	BI-RADS分类			
	a类	b类	c类	d类
观察者				
R1	32	138	559	45
R2	12	113	479	170
R3	10	95	511	158
原始报告	37	184	423	130
金标准	13 (1.7%)	112 (14.5%)	526 (67.9%)	123 (15.9%)

a: 脂肪类; b: 散在纤维腺体类; c: 不均匀致密类; d: 极度致密类。

### 2.3 各观察者与原始报告乳腺密度分类的评估效能

与金标准对照, 各观察者及原始报告的分​​类准确率见表3, 四分类方法R1、R2、R3的分类准确率均>80.00%, 原始报告的分​​类准确率偏低, 为67.70%; 二分类方法, 4组分类的准确率均>80.00%。

表3 各观察者及原始报告对乳腺X线密度的分类准确率

项目	分类准确率/%	
	四分类	二分类
观察者		
R1	81.14 (628/774)	93.93 (727/774)
R2	87.86 (680/774)	94.57 (732/774)
R3	90.96 (704/774)	96.90 (750/774)
原始报告	67.70 (524/774)	85.79 (664/774)

各观察者分类结果与金标准的一致性见表4, 四分类方法R1与金标准的一致性中等 (Kappa=0.602), R2、R3与金标准的一致性

较好 (Kappa=0.766、0.817), 原始报告的一致性中等 (Kappa=0.430); 二分类方法R1、R2、R3与金标准的一致性较好 (Kappa=0.804、0.800、0.878), 原始报告的一致性中等 (Kappa=0.599, P均<0.001)。

表4 各观察者及原始报告分类结果与金标准的一致性

组别	Kappa值	
	四分类	二分类
R1 vs 金标准	0.602	0.804
R2 vs 金标准	0.766	0.800
R3 vs 金标准	0.817	0.878
原始报告 vs 金标准	0.430	0.599

各观察者及原始报告与金标准的不一致结果见表5, 4组分类的不一致主要集中在b/c与c/d分类差异上, 其中原始报告有107例在归类为b/c上与金标准不一致, 有3例在归类为a/c上与金标准不一致。

表5 各观察者及原始报告与金标准的不一致结果

[n (%)]

项目	R1	R2	R3	原始报告
绝对一致率	628 (81.14)	680 (87.86)	704 (90.96)	524 (67.10)
a/b不一致率	19 (2.45)	1 (0.13)	3 (0.39)	25 (3.23)
b/c不一致率	47 (6.07)	42 (5.43)	24 (3.10)	107 (13.82)
c/d不一致率	80 (10.34)	51 (6.59)	43 (5.56)	115 (14.86)

a: 脂肪类; b: 散在纤维腺体类; c: 不均匀致密类; d: 极度致密类。

## 2.4 观察者间乳腺X线密度分类的一致性分析

如表6, 四分类方法R1、R2、R3的观察者间一致性中等(总体Kappa=0.671), 两两一致性从较差到中等(Kappa为0.396~0.604)。二分类方法R1、R2、R3的观察者间一致性中等(总体Kappa=0.660), 两两一致性中等(Kappa为0.629~0.690,  $P$ 均 $<0.001$ )。

表6 观察者间乳腺X线密度分类的一致性

项目	Kappa值	
	四分类	二分类
R1 vs R2	0.396	0.629
R1 vs R3	0.429	0.690
R2 vs R3	0.604	0.663
总体Kappa	0.671	0.660

## 3 讨 论

早在1976年, Wolfe等<sup>[7]</sup>首先提出用乳腺X线摄影来判断乳腺密度, 并对乳腺实质进行分型。2003年美国放射学会制定了BI-RADS标准, 依据乳腺纤维腺体组织与脂肪组织所占比例将乳腺分为脂肪型( $<25\%$ )、少量纤维腺体型( $25\% \sim 50\%$ )、多量纤维腺体型( $50\% \sim 75\%$ )和极度致密型乳腺( $>75\%$ )<sup>[8]</sup>。2013年第5版BI-RADS则不再明确不同乳腺构成分级中高密度组织的构成比例范围, 而更强调致密乳腺纤维组织的遮挡性强弱对显示小结节及微钙化的影响, 将其分为脂肪类、散在纤维腺体类、不均匀致密类和极度致密类。其中非致密型乳腺包括脂肪类和散在纤维腺体类, 致密型乳腺包括不均匀致密类和极度致密类。在乳腺癌风险因素中, 基于乳腺X线摄影发现的致密型乳腺是一个有效的危险因素, 致密型乳腺的女性患乳腺癌的可能性要比脂肪型乳腺高4~6倍<sup>[9-10]</sup>。此外, 致密型乳腺也是乳腺X线摄影筛查早期乳腺癌灵敏度及特异度偏低的主要因素之一。研究发现无论年龄, 绝经状态和激素使用如何, 放射科医师检出乳腺癌的灵敏度随着乳腺密度的增加而降低, 低乳腺密度乳腺癌检出率为80%~98%, 而高乳腺密度的检出

率下降到30.0%~64.4%<sup>[11-12]</sup>。近年随着计算机技术的发展, 计算机定量测量软件逐渐获得应用, 其以像素为基本单位在3D的基础上计算乳腺密度<sup>[13]</sup>。但软件属于商业软件, 且并不能反映第5版BI-RADS分类的修改意义, 临床未普及使用。目前国内外主要应用第5版BI-RADS评估标准, 因此对BI-RADS乳腺密度分类的解读规范和不同观察者间较好、较稳定的可重复性对乳腺影像诊断有非常很重要的意义。

本研究结果显示,  $<60$ 岁与 $\geq 60$ 岁筛查女性间乳腺X线密度差异有统计学意义, 随着年龄增长, 乳腺X线密度呈减低趋势。来自中国的大样本数据<sup>[14]</sup>(28 388名无症状妇女)结果表明, 参照第4版BI-RADS标准, 12.4%归为脂肪型( $<25\%$ ), 38.4%归为少量纤维腺体型( $25\% \sim 50\%$ ), 40.6%归为多量纤维腺体型( $50\% \sim 75\%$ )和8.6%为极度致密型( $>75\%$ ), 乳腺X线密度的比例也是随着年龄的增长而减少。本研究结果显示, 参照第5版BI-RADS标准, 1.7%归为脂肪型, 14.5%归为散在纤维腺体型, 67.9%归为不均匀致密型和15.9%为极度致密型。与该文献相比, 本研究致密型乳腺(不均匀致密型+极度致密型)的比例较高, 提示部分中等量纤维组织密度区域( $<50\%$ )的乳腺可能遮盖小的肿块而在第5版归为致密型, 使得致密型乳腺的比例较高。

Irshad等<sup>[15]</sup>研究表明, 相比于参照第4版BI-RADS标准, 参照第5版标准给出了更多比例的致密型乳腺, 其观察者间一致性中等(总体Kappa=0.57), 且比第4版较低。本研究结果显示, 3名医师的组间一致性中等(总体Kappa=0.671), 与上述研究结果一致。Ekpo等<sup>[16]</sup>研究表明, 各观察者与多数共识之间的一致性从中等到较好(Kappa为0.650~0.790), 观察者间两两一致性从较差到中等(Kappa为0.380~0.680)。本研究结果显示, 观察者与金标准的一致性同样是从中等到较好(Kappa为0.602~0.817), 且观察者间一致性同样是从较差到中等(Kappa为0.396~0.604)。由此可见, 不同观察者对第5版BI-RADS标准掌握的熟练程度

不同,对乳腺X线密度遮蔽程度的感知不同,而第5版BI-RADS标准正是基于观察者根据乳腺纤维腺体的遮挡性强弱判断的。

本研究中,观察者R1、R2、R3分类乳腺X线密度的准确率与一致性良好,低年资医师与金标准的一致性中等,中、高年资医师与金标准的一致性较好,而原始报告的准确率偏低,一致性中等,我们推测可能原因是本研究的设计使观察者能够将所有注意力集中在乳腺密度类型上,使乳腺密度成为阅片的重点,而实际临床工作中并非如此,实际工作中关注的多为是否有阳性征象且征象是否为恶性可能。但基于上述结果我们认为本研究中的一致性趋势可能会或可能不会在常规临床实践中完全转化。以上结果提示,临床可通过推广与训练增进观察者的经验;亦或是利用计算机技术制定人工智能标准,开发基于第5版BI-RADS的自动乳腺密度分类工具,以达到乳腺密度分类的可重复的目的和更高的准确率。

本研究的不足:由于乳腺密度分布的差异,本研究非致密型患者相对较少;本研究仅评价组间一致性,缺乏组内一致性的检验,后续研究中会加以补充。

综上所述,乳腺X线密度是影像筛查中的重点,第5版BI-RADS标准是临床最广泛使用的规范,为致密型乳腺和风险评估、补充筛查提供重要信息。本研究结果表明,乳腺密度分类的一致性尚有潜在的可提升的空间,临床可通过专门训练增进诊断医师的经验或开发基于第5版BI-RADS的自动乳腺密度分类的人工智能软件,以达到乳腺密度分类的可重复的目的和更高的准确率。

#### [参 考 文 献]

- [1] 张建兴. 多种影像学方法在中国乳腺癌筛查中的应用 [J]. 实用医学杂志, 2017, 33(9): 1365-1368.
- [2] BOYD N F, MARTIN L J, YAFFE M J, et al. Mammographic density and breast cancer risk: current understanding and future prospects [J]. *Breast Cancer Res*, 2011, 13(6): 223.
- [3] 胡从依, 柳 杰, 刘佩芳. 数字化乳腺X线摄影评估乳腺密度、诊断乳腺癌的研究进展 [J]. *中国医学影像技术*, 2015, 31(10): 1601-1604.
- [4] MOHAMED A A, LUO Y, PENG H, et al. Understanding clinical mammographic breast density assessment: a deep learning perspective [J]. *J Digit Imaging*, 2018, 31(4): 387-392.
- [5] SPAK D A, PLAXCO J S, SANTIAGO L, et al. BI-RADS® fifth edition: a summary of changes [J]. *Diagn Interv Imaging*, 2017, 98(3): 179-190.
- [6] RAO A A, FENEIS J, LALONDE C, et al. A pictorial review of changes in the BI-RADS fifth edition [J]. *Radiographics*, 2016, 36(3): 623-639.
- [7] WOLFE J N. Risk for breast cancer development determined by mammographic parenchymal pattern [J]. *Cancer*, 1976, 37(5): 2486-2492.
- [8] D'ORSI C, SICKLES E, MENDELSON E, et al. ACR BI-RADS atlas, breast imaging reporting and data system [M]. Reston: American College of Radiology, 2013.
- [9] BOYD N F, MARTIN L J, YAFFE M J, et al. Mammographic density and breast cancer risk: current understanding and future prospects [J]. *Breast Cancer Res*, 2011, 13(6): 223.
- [10] 杜铁桥, 王永利, 张 超, 等. 数字化乳腺X线摄影观察年龄和乳腺密度与乳腺良恶性病变的关系 [J]. *放射学实践*, 2009, 24(7): 739-743.
- [11] MANDELSON M T, OESTREICHER N, PORTER P L, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers [J]. *J Natl Cancer Inst*, 2000, 92(13): 1081-1087.
- [12] CARNEY P A, MIGLIORETTI D L, YANKASKAS B C, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography [J]. *Ann Intern Med*, 2003, 138(3): 168-175.
- [13] 秦乃姗, 郭 丽, 党 易, 等. 不同影像方法评价乳腺密度的一致性研究 [J]. *中华放射学杂志*, 2011, 45(3): 284-287.
- [14] DAI H, YAN Y, WANG P, et al. Distribution of mammographic density and its influential factors among Chinese women [J]. *Int J Epidemiol*, 2014, 43(4): 1240-1251.
- [15] IRSHAD A, LEDDY R, ACKERMAN S, et al. Effects of changes in BI-RADS density assessment guidelines (fourth versus fifth edition) on breast density assessment: intra- and interreader agreements and density distribution [J]. *AJR Am J Roentgenol*, 2016, 207(6): 1366-1371.
- [16] EKPO E U, MELLO-THOMS C, RICKARD M, et al. Breast density (BD) assessment with digital breast tomosynthesis (DBT): agreement between Quantra™ and 5th edition BI-RADS® [J]. *Breast*, 2016, 30: 185-190.

(收稿日期: 2020-02-26 修回日期: 2020-02-29)