Original Article

External validation of multiparametric magnetic resonance imaging-based decision rules for characterizing breast lesions and comparison to Kaiser score and breast imaging reporting and data system (BI-RADS) category

Yongyu An[1]^, Guoqun Mao[2], Sisi Zheng[1], Yangyang Bu[1], Zhen Fang[1], Jiangnan Lin[1], Changyu Zhou[1]^

[1]Department of Radiology, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, China; [2]Department of Radiology, Tongde Hospital of Zhejiang Province, Hangzhou, China

Contributions: (I) Conception and design: Y An, C Zhou; (II) Administrative support: G Mao, C Zhou; (III) Provision of study materials or patients: S Zheng, Y Bu; (IV) Collection and assembly of data: Z Fang, J Lin; (V) Data analysis and interpretation: Y An, C Zhou, G Mao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Changyu Zhou, MD. Department of Radiology, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), No. 54 Youdian Road, Hangzhou 310006, China. Email: tophorizon@zcmu.edu.cn.

Background: Breast imaging reporting and data system (BI-RADS) provides standard descriptors but not detailed decision rules for characterizing breast lesions. Diffusion-weighted imaging (DWI) and T2-weighted imaging (T2WI) are also not incorporated in the BI-RADS. Several multiparametric magnetic resonance imaging (mpMRI)-based decision rules have been developed to differentiate breast lesions, but lack external validation. This study aims to externally validate several mpMRI-based decision rules for characterizing breast lesions and compare them with Kaiser score and BI-RADS category.

Methods: There were 206 patients with 218 pathology-proven breast lesions (99 malignancies) included in this retrospective study from January 2018 to May 2018. Two radiologists blinded to pathology evaluated breast lesions according to the three mpMRI-based decision rules (Kim, Istomin, Zhong) and Kaiser score. BI-RADS category was extracted from radiology reports and also analysed. The diagnostic performances of the four decision rules and BI-RADS category were calculated and compared for different lesion types [mass and non-mass enhancement (NME)] and size (≤10 and >10 mm). The unnecessary biopsy rates for BI-RADS 4 lesions were calculated by the four decision rules.

Results: The three mpMRI-based decision rules showed area under the curve (AUC) of 0.81–0.87 for all lesions, 0.86–0.92 for mass lesions, 0.68–0.82 for NME, and 0.68–0.87 for lesion size ≤10 mm, 0.82–0.87 for lesion size >10 mm. Kaiser score showed the highest diagnostic performance for all subgroups except for lesion size ≤10 mm. No significant differences were found in AUC between Kaiser score and BI-RADS category. The mpMRI-based decision rules showed high sensitivity of 100% in all subgroups at the expense of low specificity (range, 2.9–41.2%). In contrast, Kaiser score demonstrated a significantly higher specificity of 73.5–92.9% than the three mpMRI-based decision rules at the cost of a decreased sensitivity (range, 60.0–93.6%) in different subgroups. The unnecessary biopsy rates for BI-RADS 4 lesions were 9.8% (Istomin), 12.2% (Zhong), 14.6% (Kim) and 70.7% (Kaiser score), respectively.

Conclusions: The mpMRI-based decision rules showed high sensitivity but low specificity for characterizing breast lesions, and their diagnostic efficiencies were inferior to Kaiser score and BI-RADS category.

Keywords: Breast disease; multiparametric magnetic resonance imaging (mpMRI); clinical decision support system

^ ORCID: Yongyu An, 0000-0003-0382-4454; Changyu Zhou, 0000-0002-5125-1737.

Submitted Dec 18, 2023. Accepted for publication Nov 29, 2024. Published online Dec 30, 2024.
doi: 10.21037/qims-23-1783
View this article at: https://dx.doi.org/10.21037/qims-23-1783

Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has a high sensitivity for breast cancer detection, but a relatively low specificity, contributing to unnecessary biopsy (1,2). The reason for this is the overlap in morphology and kinetics between benign and malignant lesions on DCE-MRI. To settle the dilemma, new imaging techniques have been investigated to reduce false-positive findings of DCE-MRI, for instance, diffusion-weighted imaging (DWI) (3), ultrafast imaging (4). Among these techniques, DWI combined with apparent diffusion coefficient (ADC) values has been shown to be the most clinically useful in identifying benign lesions classified as malignant on DCE-MRI and improving specificity (2,3,5). Additionally, T2-weighted imaging (T2WI) is also valuable for discrimination of breast lesions (6-8). Analysis of signal intensity on T2WI can reduce the false-positive findings and improve the specificity. Therefore, breast multiparametric magnetic resonance imaging (mpMRI) consisting of T2WI, DWI and DCE-MRI has been a standard protocol in clinical practice (9).

Breast imaging reporting and data system (BI-RADS) is widely used for reporting in the clinical scenario. It provides standard terminologies for imaging interpretation and management recommendation, facilitating communication between radiologists and clinicians. However, magnetic resonance imaging (MRI) BI-RADS lexicon does not provide a detailed diagnostic rule for categorization, leading to fair to moderate inter-reader agreement and variable diagnostic accuracies in final assessment (10,11). Furthermore, information obtained from T2WI and DWI is not incorporated into MRI BI-RADS lexicon. Therefore, as it has been done in prostate imaging reporting and data system (PI-RADS) (12), development of a straight decision algorithm in BI-RADS classification based on mpMRI is a trend to assist in the categorization of breast lesions and reduce variability between radiologists.

Several mpMRI-based diagnostic rules have been proposed by studies to differentiate breast malignancy from benign lesions (13-19). Qualitative and quantitative assessment, including BI-RADS descriptors, lesion size and ADC values, are integrated into these diagnostic algorithms. For instance, Kaiser score integrates five key imaging features [spiculated sign, time signal intensity curve (TIC), internal enhancement pattern, margin and edema] that are obtained from DCE-MRI and T2WI into an intuitive flowchart (20), and it demonstrates high accuracy for evaluation of breast lesions as well as low variability between radiologists (21-24). A simplified decision rule developed by Zhong *et al.* (18) integrates three imaging findings and it shows an area under the curve (AUC) of 0.844 for identifying malignant lesions with a cutoff of category ≥ BI-RADS 4. Additionally, diagnostic rules are designed in various clinical settings, for example, subcategorization of BI-RADS 4 lesions, classification of non-mass enhancement (NME). These decision rules have potential to assist radiologists in characterizing breast lesions in daily work and improve inter-reader agreement in BI-RADS classification. However, none of these mpMRI-based decision rules except Kaiser score have been externally validated, despite the importance of validation prior to clinical use. Most importantly, validation of these decision rules could serve as a reference for development of a decision algorithm for BI-RADS classification based on mpMRI in the future. Therefore, the purpose of the study is to validate several mpMRI-based decision rules proposed in the previous studies and compare them with Kaiser score and BI-RADS category. We present this article in accordance with the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1783/rc).

Methods

Study population

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Institutional Ethical Board of The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine) (No. 2021-KL-062-01). Informed consent from patients was waived due to retrospective nature of the study.

Patients who underwent breast MRI examinations were included in the study from January 2018 to May 2018. Indications for breast MRI included suspicious findings on mammography or ultrasound, abnormal clinical symptoms

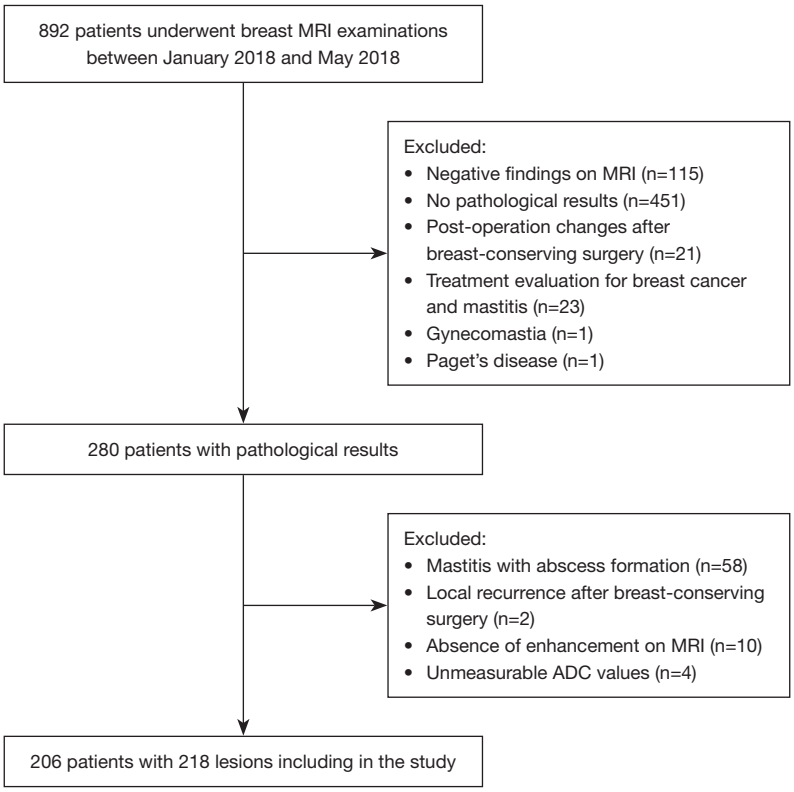

Figure 1 The patient selection flowchart of the study. MRI, magnetic resonance imaging; ADC, apparent diffusion coefficient.

as well as preoperative evaluation, screening and so on. The inclusion criteria in our study were as follows: (I) abnormal enhanced lesions were found on MRI. (II) MRI examinations were performed prior to surgery and treatment. (III) Lesions were proven by biopsy or surgical pathology. (IV) Imaging quality was good for analysis. We excluded 58 patients diagnosed mastitis with abscess formation due to typical imaging features on mpMRI and clinical presentation. These lesions were probably misdiagnosed as malignant on DCE-MRI. Additionally, 4 patients were excluded due to unmeasurable ADC values for breast lesions. Finally, 206 patients with 218 breast lesions (12 patients with bilateral breast lesions) were involved in our study. The patient selection flowchart of the study is shown in *Figure 1*. Part of the data has been published in another study (25) with different aim and design.

MRI protocol

Breast MRI examinations were performed on 1.5 Tesla (Avanto, Siemens, Erlangen, Germany, n=31) and 3.0 Tesla MRI scanners (Verio, Siemens, n=136; Discovery MR750, GE Healthcare, Chicago, IL, USA, n=39) with dedicated breast coils. All patients were in prone position. Imaging protocol consisted of T1WI with non-fat saturation (optional), T2WI with fat saturation, DWI and DCE sequences. The b-values used in DWI sequence were 0, 1,000 s/mm^2 on a 3.0 T GE scanner and 50, 400, 800 s/mm^2 on both 1.5 T and 3.0 T Siemens scanners. ADC maps were generated automatically by the workstation using monoexponential model and uploaded to Picture Archiving and Communication System (PACS). The DCE substraction images generated by the workstation without movement correction were also uploaded to the PACS. A dose (0.1 mmol/kg) of gadobenate dimeglumine (Beilu Pharmaceutical, Beijing, China) was injected intravenously at a rate of 2.0 mL/s, followed by a saline flush. The details of MRI scan parameters are listed in Table S1.

Selection of mpMRI-based decision rules

We searched previous studies in PubMed and referred to the study (17) that summarized decision rules. The inclusion criteria for a decision rule were in accordance

with the objectives of our study and also referred to PI-RADS (12) which provided a streamlined algorithmic approach to assigning categories. They were as follows: (I) imaging features in the decision rule were consistent with the 5th edition MRI BI-RADS descriptors. (II) It was intuitive and easy to use for radiologists in daily work, as was PI-RADS. Those that were displayed by complicated formulas were excluded. (III) It was designed to distinguish benign from malignant lesions regardless of BI-RADS category and could be applied to both mass and NME. Decision rules designed to subclassify BI-RADS 4 lesions were excluded. (IV) It was developed on mpMRI, of which DCE-MRI was the basis. Finally, three mpMRI-based decision rules met the inclusion criteria and were selected in our study, that was Kim *et al.* (19), Istomin *et al.* (17), Zhong *et al.* (18). The selection of mpMRI-based decision rules selection is summarized in Table S2. Previous studies have revealed that Kaiser score can be used as a valuable clinical decision tool for characterization of breast lesions (22-24). Hence, Kaiser score was also analysed in the study although the imaging features (spiculated sign, internal enhancement pattern and margin) involved were discordant with MRI BI-RADS descriptors for NME. Furthermore, BI-RADS category recorded in the radiology reports was also collected and analyzed in the study. In our department, breast MRI interpretation was a double-reading procedure performed by a junior and a senior breast radiologist. BI-RADS category was assigned based on their experience after assessing mpMRI. Lesions with BI-RADS 4–5 category were considered malignant. A brief summary of the decision rules included in our study is presented below, and details are provided in Appendix 1.

Kim *et al.* (19): it was developed based on imaging features on DCE images and signal intensity on DWI and T2WI (DWI-T2WI set). The total score for a lesion was the sum of DCE score and DWI-T2WI set score and it ranged from 3 to 10 points. A lesion with a total score >5 points was considered malignant, otherwise, it was considered benign. DCE score was equal to DCE BI-RADS category on basis of morphology, enhancement pattern and kinetics. The DWI-T2WI set score was given according to the signal intensity on T2WI and DWI with high b values, ranging from 1 (definite benign) to 5 points (highly suggestive of malignancy) (Table S3).

Istomin *et al.* (17): MRI features, including morphology, enhancement pattern, kinetics, signal intensity on T2WI and ADC values, were categorized into the minor, intermediate and major findings predictive of malignancy. A lesion with the minor features was set as BI-RADS 3 category, a lesion with intermediate and no more than one major feature was set as BI-RADS 4 category, otherwise, it was set as BI-RADS 5 category. Lesions with BI-RADS 4–5 category were considered malignant (Table S4).

Zhong *et al.* (18): the lesions were evaluated based on three main categories: morphology, TIC type, and ADC values. For each major category in which suspicious features were present, a score of 1 point was assigned. A lesion with a score of 0 points was equivalent to a BI-RADS 3 category, a score of 1 point was equivalent to a BI-RADS 4 category, and a score of 2–3 points was equivalent to a BI-RADS 5 category. In our study, lesions with a score of 0 points (corresponding to BI-RADS 3 category) were considered benign, whereas lesions with a score of 1–3 points (corresponding to BI-RADS 4–5 category) were considered malignant, in line with BI-RADS lexicon (Table S5).

Kaiser score (20): it was a decision tree that involved five imaging features (spiculated sign, TIC type, margin, enhancement pattern and edema) obtained from T2WI and DCE-MRI. Kaiser score ranged from 1 to 11 points, with greater score suggestive of a higher likelihood of malignancy. Lesions with a score >4 points were considered as malignant.

Imaging interpretation

Two breast radiologists (Y.A. and C.Z. with 4 and 19 years of experience in breast imaging, respectively) evaluated all breast images together and disagreement was settled by consensus. They were blinded to pathological results but aware of basic clinical information. Before imaging evaluation, a professor specialized in breast imaging made a detailed introduction of the three mpMRI-based decision rules and Kaiser score to the two radiologists according to the relevant studies (17-19,26). Then, 20 cases that were not included in the present study were provided to the two readers for practice. During imaging interpretation, the readers were required to document the following findings: lesion size, lesion type, shape, margin, distribution, internal enhancement pattern, TIC type, signal intensity on T2WI with fat saturation and DWI, edema, ADC values. The maximal diameter of a lesion was measured on the largest axis on DCE images with multiplanar reconstruction view. If lesions were found in bilateral breasts in one patient, each lesion in per breast was evaluated. For multiple lesions in the ipsilateral breast, the most suspicious lesion was assessed.

For ADC values measurement, several methods were reported in previous studies (13,27-29). In our study, we used

Table 1 Pathological results of the study cohort

Pathology	Values
Malignant	99
Invasive ductal carcinoma	75 (75.8)
Ductal carcinoma *in situ*	18 (18.2)
Mucinous carcinoma	3 (3.0)
Others	3 (3.0)
Benign	119
Fibroadenoma	63 (52.9)
Intraductal papilloma	18 (15.1)
Adenosis	24 (20.2)
Inflammatory diseases	7 (5.9)
Atypical ductal hyperplasia	2 (1.7)
Others	5 (4.2)

Numbers in the parentheses are percentages.

the method recommended by European Society of Breast Imaging International Breast DWI working group (5). Region of interest (ROI) was placed on the darkest part of ADC maps. Necro-cystic changes, hemorrhage, noisy and non-enhancing regions were avoided with reference to DCE images. The mean ADC values within ROI were recorded.

After imaging evaluation, a breast radiologist (Z.F.) who also attended the aforementioned training session on the decision rules organized these imaging features without knowledge of the pathology results and calculated the corresponding results for further analysis according to the decision rules.

Reference standard

Histopathology obtained from biopsy or surgical resection was used as the reference standard. In our study, high-risk lesions were diagnosed by surgical pathology and classified as benign.

Statistical analysis

Statistical analysis was performed using SPSS (version 26.0, IBM Corp, USA) and MedCalc (version 15.6.1, MedCalc Software bvba, Belgium). All calculations were performed on a per-lesion basis. Quantitative data with a normal distribution were expressed as the mean and standard deviation (SD). Otherwise, they were expressed as the median and interquartile range (IQR). Sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and their 95% confidence intervals (CIs) were calculated for all decision rules. Receiver operating characteristic (ROC) curve was performed to assess diagnostic efficiencies and AUC was compared by DeLong test. Diagnostic performance was also calculated stratifying by lesion type (mass and NME) and lesion size (≤10 and >10 mm). Cochran's *Q* test was performed for comparison of sensitivity, specificity between all decision rules and Bonferroni adjustment was used for multiple comparisons. Furthermore, unnecessary biopsy rates for BI-RADS 4 lesions were analysed using the three mpMRI-based decision rules and Kaiser score. A two-tailed P value <0.05 was considered statistically significant.

Results

Study population

A total of 206 patients with 218 lesions were included in the study. The mean age of patients was 48.1 years (range, 18–84 years, SD: 12.4 years). Of 218 breast lesions, 99 were proved malignant (malignancy rate: 45.4%), of which invasive ductal carcinoma (75.8%, 75/99) was the most common type, followed by ductal carcinoma *in situ* (18.2%, 18/99). Of 119 benign lesions, fibroadenoma (52.9%, 63/119) was the most common type, followed by adenosis (20.2%, 24/119). Pathological results are displayed in *Table 1*. The size of breast lesions ranged from 5 to 97 mm (median: 18 mm, IQR: 12–28 mm), with the malignancy being larger than the benign (median: 23 *vs.* 14 mm, P<0.001). There were 157 masses and 61 NME in the study. The malignancy rates in mass and NME were 45.9% (72/157) and 44.3% (27/61), respectively.

Diagnostic performance of BI-RADS category and the four decision rules

Of 218 breast lesions, 81 (37.2%), 73 (33.5%), and 64 (29.4%) lesions were classified as BI-RADS 3, BI-RADS 4, and BI-RADS 5 category, respectively. The corresponding malignancy rates were 4.9% (4/81), 43.8% (32/73), and 98.4% (63/64), respectively. Descriptive statistics for all decision rules stratified by individual scores are displayed in Table S6. The corresponding BI-RADS category based on Kaiser score and Zhong *et al.* is shown in Table S7. Notably, 9 benign lesions were unable to be scored by the DWI-T2WI set score in Kim

Table 2 Diagnostic performance of all decision rules for breast lesions stratified by lesion type and size

Lesions	Decision rules	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	AUC
All	Istomin *et al.* (17)	100 (96.3, 100)	21.0 (14.1, 29.4)	56.9 (50.3, 63.5)	51.3 (44.0, 58.5)	100 (86.3, 100)	0.81 (0.75, 0.86)
	Zhong *et al.* (18)	100 (96.3, 100)	31.9 (23.7, 41.1)	62.8 (56.4, 69.3)	55.0 (47.4, 62.4)	100 (90.8, 100)	0.87 (0.82, 0.92)
	Kim *et al.* (19)[†]	100 (96.3, 100)	30.9 (22.5, 40.4)	63.6 (56.7, 70.2)	56.6 (48.9, 64.0)	100 (89.7, 100)	0.84 (0.79, 0.89)
	Kaiser score (20)	91.9 (84.7, 96.5)	87.4 (80.1, 92.8)	89.4 (85.3, 93.6)	85.9 (77.7, 91.9)	92.9 (86.4, 96.9)	0.94 (0.89, 0.97)
	BI-RADS category	96.0 (90.0, 98.9)	64.7 (55.4, 73.2)	78.9 (73.4, 84.4)	69.3 (60.9, 76.9)	95.1 (87.8, 98.6)	0.91 (0.87, 0.95)
Mass	Istomin *et al.* (17)	100 (95.0, 100)	28.2 (19.0, 39.0)	61.1 (53.4, 68.9)	54.1 (45.3, 62.8)	100 (85.8, 100)	0.86 (0.79, 0.91)
	Zhong *et al.* (18)	100 (95.0, 100)	41.2 (30.6, 52.4)	68.2 (60.8, 75.5)	59.0 (49.8, 67.8)	100 (90.0, 100)	0.89 (0.83, 0.94)
	Kim *et al.* (19)[†]	100 (95.0, 100)	35.9 (25.3, 47.6)	66.7 (59.0, 74.3)	59.0 (49.8, 67.8)	100 (87.7, 100)	0.92 (0.87, 0.96)
	Kaiser score (20)	93.1 (84.5, 97.7)	92.9 (85.3, 97.4)	93.0 (89.0, 97.0)	91.8 (83.0, 96.9)	94.1 (86.7, 98.0)	0.96 (0.92, 0.99)
	BI-RADS category	94.4 (86.4, 98.5)	75.3 (64.8, 84.0)	84.1 (78.3, 89.9)	76.4 (66.2, 84.8)	94.1 (85.6, 98.4)	0.94 (0.89, 0.97)
NME	Istomin *et al.* (17)	100 (87.2, 100)	2.9 (0.07, 15.3)	45.9 (33.0, 58.8)	45.0 (32.1, 58.4)	100 (2.5, 100)	0.68 (0.55, 0.80)
	Zhong *et al.* (18)	100 (87.2, 100)	8.8 (1.9, 23.7)	49.2 (36.3, 62.1)	46.6 (33.3, 60.1)	100 (29.2, 100)	0.80 (0.68, 0.89)
	Kim *et al.* (19)[†]	100 (87.2, 100)	18.8 (7.2, 36.4)	55.9 (42.9, 69.0)	50.9 (36.8, 64.9)	100 (54.1, 100)	0.82 (0.70, 0.91)
	Kaiser score (20)	88.9 (70.8, 97.7)	73.5 (55.6, 87.1)	80.3 (70.1, 90.6)	72.7 (54.5, 86.7)	89.3 (71.8, 97.7)	0.84 (0.72, 0.92)
	BI-RADS category	100 (87.2, 100)	38.2 (22.2, 56.4)	65.6 (53.3, 77.8)	56.3 (41.2, 70.5)	100 (75.3, 100)	0.83 (0.71, 0.92)
Size ≤10 mm	Istomin *et al.* (17)	100 (47.8, 100)	20.5 (9.3, 36.5)	29.5 (16.8, 45.2)	13.9 (4.7, 29.5)	100 (63.1, 100)	0.68 (0.44, 0.93)
	Zhong *et al.* (18)	100 (47.8, 100)	30.8 (17.0, 47.6)	38.6 (24.4, 54.5)	15.6 (5.3, 32.8)	100 (73.5, 100)	0.87 (0.76, 0.99)
	Kim *et al.* (19)[‡]	100 (47.8, 100)	27.0 (13.8, 44.1)	35.7 (21.6, 52.0)	15.6 (5.3, 32.8)	100 (69.2, 100)	0.68 (0.48, 0.87)
	Kaiser score (20)	60.0 (14.7, 94.3)	92.3 (79.1, 98.4)	88.6 (75.4, 96.2)	50.0 (11.8, 88.2)	94.7 (82.3, 99.4)	0.86 (0.72, 1.0)
	BI-RADS category	80.0 (28.4, 99.5)	69.2 (52.4, 83.0)	70.5 (54.8, 83.2)	25.0 (7.3, 52.4)	96.4 (81.7, 99.9)	0.79 (0.57, 1.0)
Size >10 mm	Istomin *et al.* (17)	100 (96.2, 100)	21.3 (12.9, 31.8)	63.8 (56.2, 70.9)	59.9 (51.8, 67.6)	100 (80.5, 100)	0.82 (0.76, 0.88)
	Zhong *et al.* (18)	100 (96.2, 100)	32.5 (22.5, 43.9)	69.0 (61.5, 75.7)	63.5 (55.2, 71.3)	100 (86.8, 100)	0.87 (0.81, 0.92)
	Kim *et al.* (19)[‡]	100 (96.2, 100)	32.9 (22.3, 44.9)	70.7 (63.1, 77.4)	65.7 (57.3, 73.5)	100 (87.8, 100)	0.85 (0.79, 0.90)
	Kaiser score (20)	93.6 (86.6, 97.6)	85.0 (75.3, 92.0)	89.7 (84.1, 93.8)	88.0 (80.0, 93.6)	91.9 (83.2, 97.0)	0.94 (0.91, 0.98)
	BI-RADS category	96.8 (91.0, 99.3)	62.5 (51.0, 73.1)	81.0 (74.4, 86.6)	75.2 (66.5, 82.6)	94.3 (84.3, 98.8)	0.91 (0.87, 0.96)

Numbers in the parentheses are 95% confidence intervals. [†], 9 benign lesions, including 7 masses and 2 NME, cannot be scored by DWI-T2WI set score because of mismatch of their signal intensity on DWI and T2WI with DWI-T2WI set in the decision rule by Kim *et al.*, thus they are not involved in the analysis. [‡], 2 of these 9 lesions mentioned above have diameters less than or equal to 10 mm and are also not involved in this subgroup analysis, as well as for 7 lesions greater than 10 mm. PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; BI-RADS, breast imaging reporting and data system; NME, non-mass enhancement; DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging.

et al. because their signal intensity on DWI and T2WI could not match with the DWI-T2WI set. Among these 9 benign lesions, 77.8% (7/9) were fibroadenomas. The details of these 9 benign lesions are shown in Table S8, and the examples are given in Figures S1,S2. The correlation of the DWI-T2WI set score by Kim *et al.* with pathology results are summarized in Table S9.

The diagnostic performance of BI-RADS category and the four decision rules is shown in *Table 2* and *Figure 2*. Kaiser score achieved a higher AUC of 0.94 (95% CI: 0.89–0.97) than BI-RADS category (AUC: 0.91; 95% CI: 0.87–0.95; P=0.266) and the three mpMRI-based decision rules (AUC: 0.81–0.87, all P<0.05). Zhong *et al.* yielded a higher AUC of 0.87 (95% CI: 0.82–0.92) than Istomin *et al.* (AUC: 0.81; 95% CI: 0.75–0.86; P=0.042) and Kim *et al.* (AUC: 0.84; 95% CI: 0.79–0.89; P=0.427). The three mpMRI-based decision rules showed a sensitivity of 100% (95% CI: 96.3–100%), higher than Kaiser score (91.9%; 95%

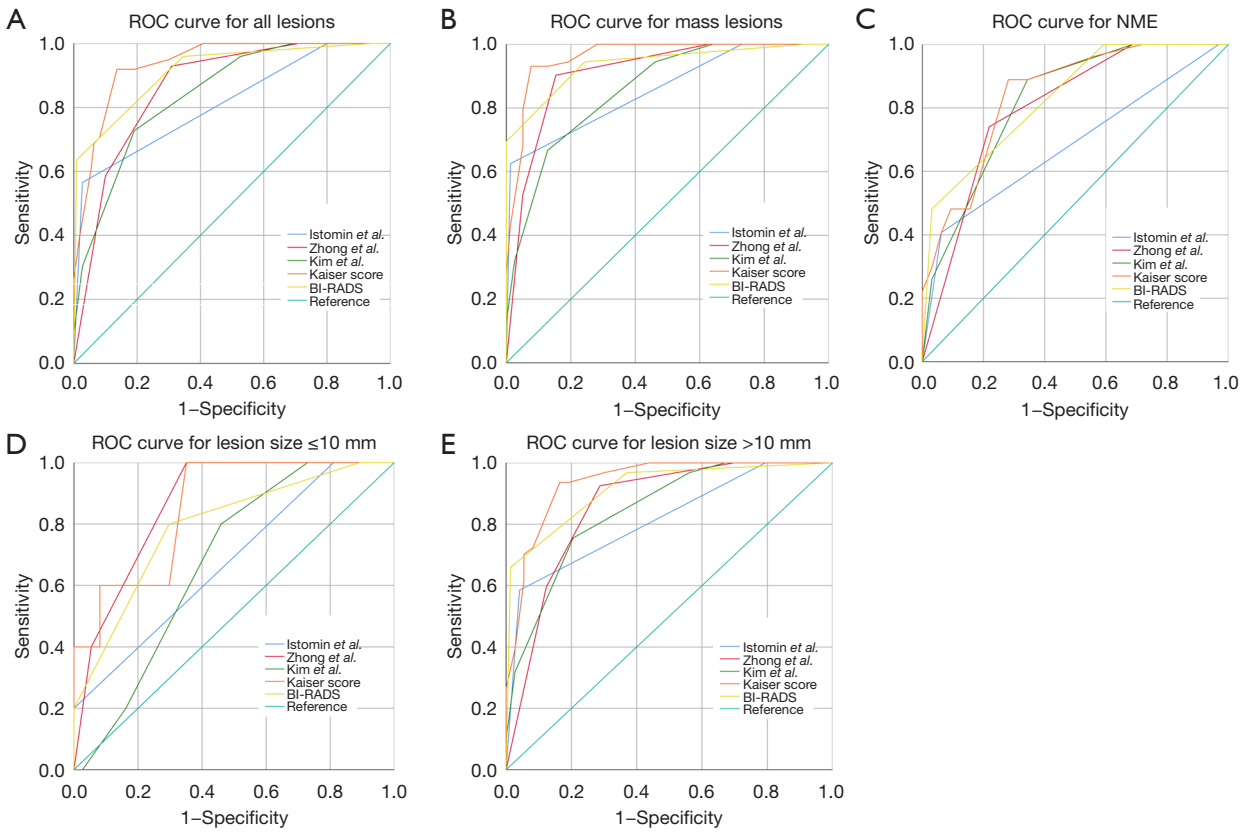

Figure 2 ROC curves of the decision rules for all breast lesions (A), mass lesions (B), NME (C), and lesion size ≤10 mm (D), lesion size >10 mm (E). ROC, receiver operating characteristic; BI-RADS, breast imaging reporting and data system; NME, non-mass enhancement.

CI: 84.7–96.5%; P=0.003) and BI-RADS category (96.0%; 95% CI: 90.0–98.9%; P=0.679). No statistical differences in specificity were found between Istomin *et al.* (21.0%; 95% CI: 14.1–29.4%), Kim *et al.* (30.9%; 95% CI: 22.5–40.4%) and Zhong *et al.* (31.9%; 95% CI: 23.7–41.1%) (all P>0.05), all of which were significantly lower than Kaiser score (87.4%; 95% CI: 80.1–92.8%) and BI-RADS category (64.7%; 95% CI: 55.4–73.2%) (all P<0.001). In comparison to BI-RADS category, Kaiser score showed a significantly higher specificity (87.4% *vs.* 64.7%, P=0.001), but as the cost of a lower sensitivity (91.9% *vs.* 96.0%, P=0.679). Among all decision rules, Kaiser score had the highest accuracy (89.4%; 95% CI: 85.3–93.6%), followed by BI-RADS category (78.9%; 95% CI: 73.4–84.4%), both higher than the three mpMRI-based decision rules (range, 56.9–63.6%). Comparison of diagnostic performance of the three mpMRI-based decision rules in the original studies and ours is presented in Table S10.

When stratified by lesion type, all decision rules showed higher diagnostic efficiencies for mass (AUC: 0.86–0.96) than for NME (AUC: 0.68–0.84) (*Table 2* and *Figure 2*). For mass, Kaiser score achieved a higher AUC of 0.96 (95% CI: 0.92–0.99) than BI-RADS category (AUC: 0.94, 95% CI: 0.89–0.97; P=0.164) and the three mpMRI-based decision rules (AUC: 0.86–0.92, all P<0.05). The sensitivity of the three mpMRI-based decision rules was 100% (95% CI: 95.0–100%), higher than BI-RADS category (94.4%; 95% CI: 86.4–98.5%; P=0.350) and Kaiser score (93.1%; 95% CI: 84.5–97.7%; P=0.084), however, the specificity of the three mpMRI-based decision rules (range, 28.2–41.2%) was significantly lower than BI-RADS category (75.3%; 95% CI: 64.8–84.0%) and Kaiser score (92.9%; 95% CI: 85.3–97.4%) (all P<0.001). For NME, Istomin *et al.* demonstrated lower performance (AUC: 0.68; 95% CI: 0.55–0.80) than BI-RADS category (AUC: 0.83; 95% CI: 0.71–0.92; P=0.057) and the other three decision rules (AUC: 0.80–0.84, all P<0.05). Kaiser score showed a higher AUC than BI-RADS, Zhong *et al.* and Kim *et al.* without significant difference. The three mpMRI-based decision rules and BI-RADS category showed a higher sensitivity (100% *vs.* 88.9%,

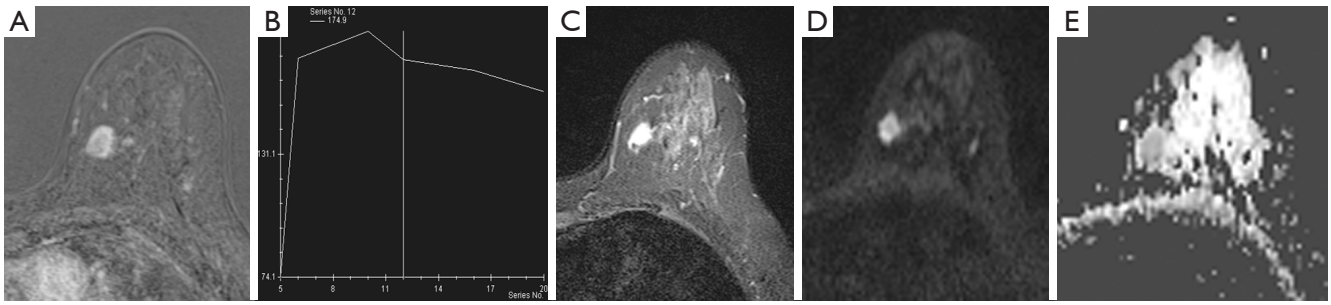

Figure 3 A 53-year-old woman presented with a mass in the left breast on MRI which was classified as BI-RADS 4 category. (A,B) The mass demonstrated oval shape, spiculated margin, rim enhancement and washout kinetics. (C,D) The lesion showed high signal intensity on both T2WI and DWI, with absence of edema. (E) ADC value was 0.911×10^{-3} mm^2/s. The lesion was considered malignant by Kaiser score and the three mpMRI-based decision rules. Pathology showed invasive ductal carcinoma. MRI, magnetic resonance imaging; BI-RADS, breast imaging reporting and data system; T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; mpMRI, multiparametric magnetic resonance imaging.

P=0.062) but a significantly lower specificity than Kaiser score (2.9–38.2% *vs.* 73.5%, P<0.05) for NME.

When stratified by lesion size, all decision rules except Zhong *et al.* yielded higher AUCs for lesion size >10 mm (range, 0.82–0.94) than for lesion size ≤10 mm (range, 0.68–0.86) (*Table 2* and *Figure 2*). The three mpMRI-based decision rules showed a sensitivity of 100% in both subgroups, higher than BI-RADS (80.0%; 95% CI: 28.4–99.5% for lesion size ≤10 mm; 96.8%; 95% CI: 91.0–99.3% for lesion size >10 mm, P>0.05) and Kaiser score (60.0%; 95% CI: 14.7–94.3% for lesion size ≤10 mm; 93.6%; 95% CI: 86.6–97.6% for lesion size >10 mm, P<0.05). However, the specificity of the three mpMRI-based decision rules was low, ranging from 20.5–30.8% for lesions size ≤10 mm and 21.3–32.9% for lesions size >10 mm, and no statistical differences in specificity were found between the three mpMRI-based decision rules in both subgroups (all P>0.05). In contrast, Kaiser score showed a high specificity of 92.3% (95% CI: 79.1–98.4%) for lesion size ≤10 mm and 85.0% (95% CI: 75.3–92.0%) for lesions size >10 mm, higher than BI-RADS (69.2%; 95% CI: 52.4–83.0% for lesion size ≤10 mm, P=0.218; 62.5%; 95% CI: 51.0–73.1% for lesion size >10 mm, P=0.011) and the three mpMRI-based decision rules (all P<0.001).

Unnecessary biopsy rate for BI-RADS 4 lesions

Of 73 lesions classified as BI-RADS 4 category, 41 lesions had false-positive findings. There were 4, 5, 6 and 29 lesions that were correctly downgraded to benign by Istomin *et al.*, Zhong *et al.*, Kim *et al.* and Kaiser score, respectively. The unnecessary biopsy rates for BI-RADS 4 lesions using Istomin *et al.*, Zhong *et al.*, Kim *et al.* and Kaiser score were 9.8%, 12.2%, 14.6% and 70.7%, respectively. The examples are given in *Figures 3*,*4*.

Discussion

Our study indicated that the three mpMRI-based decision rules were inferior to BI-RADS and Kaiser score for all breast lesions, and demonstrated a higher sensitivity of 100% but a lower specificity in different subgroups stratified by lesion size and type in comparison to BI-RADS and Kaiser score. Among these decision rules, Kaiser score exhibited the highest specificity than the others decision rules and it was more valuable for BI-RADS 4 lesions to reduce unnecessary biopsy due to its high specificity.

MRI BI-RADS lexicons define suspicious imaging features for breast lesions, such as spiculated margin, rim enhancement and washout kinetics. However, a detailed diagnostic algorithm is not provided. The three mpMRI-based decision rules were based on the principle that all suspicious features were equally weighted, and lesions were assigned as malignant if they were present. This approach accounted for a sensitivity of 100% for breast cancer diagnosis, which was higher than BI-RADS and Kaiser score. Consequently, the specificity was relatively low in different subgroups stratified by lesion type and size, ranging from 2.9–35.9% in our study, while it was 11.5–91.7% in the original studies (17-19). Discrepancies in specificity of the three mpMRI-based decision rules

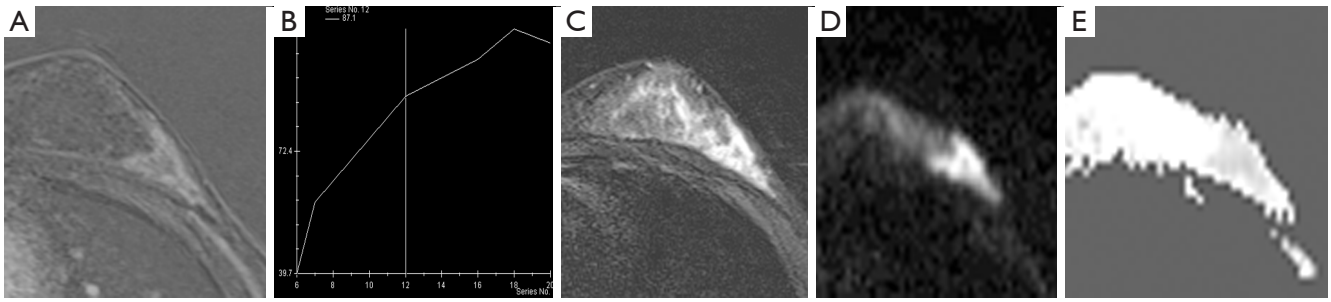

Figure 4 A 18-year-old woman presented with NME in the left breast on MRI which was classified as BI-RADS 4 category. (A,B) The lesion displayed segmental distribution, heterogeneous internal enhancement, circumscribed margin and persistent kinetics. (C,D) The lesion showed high signal intensity on both T2WI and DWI. (E) ADC value was 1.12×10^{-3} mm^2/s. The lesion was classified as benign by Kaiser score and as malignant by the three mpMRI-based decision rules. Pathology revealed adenosis. NME, non-mass enhancement; MRI, magnetic resonance imaging; BI-RADS, breast imaging reporting and data system; T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; mpMRI, multiparametric magnetic resonance imaging.

between the original studies and ours may be related to malignancy rate and variety of breast lesions in the study cohorts. For instance, the malignancy rate in the study by Istomin *et al.* (17) was 76.2% (531/697), while it was 45.4% (99/218) in our study. The high prevalence of malignant lesions is associated with low specificity (30). In the study by Zhong *et al.* (18), 88 benign lesions classified as BI-RADS 1–2 category were involved, including symmetric background parenchymal enhancement, cyst and non-enhanced duct dilation. In contrast, only enhanced lesions classified as BI-RADS 3–5 category were included in our study. However, the main reasons for the low specificity of the three mpMRI-based decision rules are probably due to the following factors. Firstly, the risk weights for malignancy vary among suspicious features (15,17,31). For example, spiculated margin is a major predictor of malignancy for breast mass lesions (15,17), while heterogeneous enhancement is intermediate and plateau is minor (15). Therefore, equal weighting of all suspicious features regardless of their risk ratios for malignancy in the mpMRI-decision rules may result in low specificity. It is noteworthy that round shape is assigned as an intermediate predictor for mass in the decision rule by Istomin *et al.* (17), leading to its low specificity. Secondly, DWI-T2WI set score in the decision rule by Kim *et al.* (19) that combined signal intensity on DWI and T2WI showed a sensitivity of 81.8%, specificity of 87.5% in their study (19). In our study, the DWI-T2WI set score demonstrated a sensitivity of 96.0%, specificity of 33.6%. The majority of false-positive lesions in the DWI-T2WI set were fibroadenomas (42.5%, 31/73) in our study, which represented the most common benign breast disease in clinical practice. Fibroadenomas often present hyperintensity on both T2WI and DWI (32,33), corresponding to a score of 3 points (possibly malignant) in the DWI-T2WI set score. In such cases, fibroadenomas may be misdiagnosed as malignant if they are assigned BI-RADS 3–4 category on DCE images, resulting in reduced specificity of the decision rule by Kim *et al.* The high specificity of 91.7% of the decision rule by Kim *et al.* in the original study may be associated with the low prevalence of fibroadenomas, which was 12.5% in benign lesions. It is, however, unreasonable to reflect the true diffusion level of breast lesions by the DWI-T2WI set score because of the influence of signal intensity of T2WI on DWI, referring to T2WI shine-through and blackout (5). Signal intensity on ADC maps or ADC values instead of DWI is a more favorable indicator of diffusion level of breast lesions. Additionally, nine benign lesions cannot be scored by the DWI-T2WI set score in our study, indicating its limited generalizability. Lastly, in the decision rule by Zhong *et al.* (18), breast lesions with ADC values below a certain cutoff ($\leq1.05\times10^{-3}$ mm^2/s for mass, $\leq1.35\times10^{-3}$ mm^2/s for NME) were classified as malignant, potentially leading to a decrease in specificity. Given the high sensitivity of DCE-MRI, the value of DWI is used as an adjunct to DCE-MRI to improve its specificity, with the aim of reducing biopsies of benign lesions (3,5). Moreover, benign findings on DCE-MRI are reliable to exclude malignancy due to its high NPV of nearly 100% (2,34,35). Consequently, it is preferable to set a high cutoff for ADC values to downgrade suspicious lesions on DCE-MRI rather than a low cutoff to upgrade benign lesions. For instance, a high cutoff of

ADC values of 1.53×10^{-3} mm^2/s is useful to reduce MRI-prompted biopsies (36). Similarly, adding T2WI to DCE-MRI can reduce the number of false-positive findings and increase the specificity. This is because breast lesions with hyperintensity on T2WI are commonly benign (6-8). However, upgrading breast lesions with low-intermediate signal intensity on T2WI to malignant also leads to low specificity of the decision rule by Istomin *et al.* (17).

In contrast, the major predictors for malignancy are selected by Chi-squared automatic interaction detection methods in the Kaiser score, which is displayed by a decision tree flowchart (15). In our study, Kaiser score outperformed the three mpMRI-based decision rules and it showed an AUC of 0.94, sensitivity of 91.9% and specificity of 87.4% for all breast lesions. The performance of Kaiser score in our study is within the reported ranges of 0.796–0.941 for AUC, 80.6–98.9% for sensitivity, 45.1–83.9% for specificity in different clinical scenarios (21-24,37). The application of Kaiser score could reduce unnecessary biopsy rates by 70.7% for BI-RADS 4 lesions due to its high specificity, though at the cost of decreased sensitivity. Our study once again demonstrated the value of Kaiser score used as a clinical decision tool for characterizing breast lesions, specifically in reducing unnecessary biopsies of breast benign lesions (22-24,37). Additionally, BI-RADS category based on readers' experience also demonstrated higher performance than the three mpMRI-based decision rules in our study. The results can be owing to the double-reading procedure performed by two breast radiologists and accumulation of experience in breast imaging interpretation with wide usage of breast mpMRI.

When stratified by lesion type, all decision rules achieved lower diagnostic performance for NME than for mass. In particular, the specificity of BI-RADS category and the mpMRI-based decision rules for NME was notably low, with a range of 2.9–38.2%. This finding was in concordance with previous studies (10,38,39), indicating that it was challenging to accurately diagnose NME on breast MRI using BI-RADS descriptors and DWI. On DCE-MRI, breast NME was a major cause of false-positive findings in imaging interpretation (38) and affected the diagnostic performance of radiologists (10). The addition of DWI to DCE-MRI also showed the limited value for NME (40). In comparison, Kaiser score yielded a significantly higher specificity of 73.5% for NME, but at the expense of a lower sensitivity of 88.9%. However, it has to mention that the imaging features including spiculated sign, margin and internal enhancement pattern in Kaiser score are discordant with MRI BI-RADS descriptors for NME, making it sometimes difficult to evaluate.

Small breast lesions are often detected on MRI, especially in high-risk screening. Accurate assessment of these small lesions is of great importance both to increase breast cancer detection and to avoid unnecessary biopsies. In our study, all decision rules except Zhong *et al.* showed higher AUCs for lesion size >10 mm than for lesion size ≤10 mm. The mpMRI-based decision rules still yielded a low specificity despite a high sensitivity of 100% in both subgroups. In contrast, BI-RADS and Kaiser score showed a lower sensitivity for lesion size ≤10 mm than for lesion size >10 mm. The results may indicate that accurate interpretation of small breast lesions may be challenging, in line with previous studies (41-43). The typical imaging features of breast lesions on MRI depend on lesion size, and small breast cancers often have benign appearance on MRI (44,45).

The assignment of a BI-RADS category in breast MRI interpretation requires the assessment of various imaging features and multiple sequences, which is complicated, especially for inexperienced radiologists. This complexity leads to variability in diagnostic accuracies and inter-reader agreement in BI-RADS classification (10,11). In such a clinical setting, a decision algorithm for BI-RADS categorization is warranted. A decision algorithm can standardize imaging interpretation and provide radiologists with an objective, straightforward diagnostic approach that compensates for experience. With the help of a decision rule, inexperienced radiologists can perform well in imaging evaluation, and they even outperform experienced radiologists (46). The finding that Kaiser score outperforms BI-RADS category in our study suggests the benefit of a decision rule. So far, several decision algorithms in other fields have been proposed and validated, including PI-RADS (12), the MRI clear cell likelihood score for renal mass (47). The findings in our study provide insights for development of a decision algorithm for BI-RADS classification in the future. The specificity of decision rules need to be improved to reduce necessary biopsies, while maintaining high sensitivity. The key imaging features involved in a decision rule should be selected from various breast imaging finding, as is done with Kaiser score. DWI together with ADC values and T2WI are valuable for characterizing breast lesions. How they can be integrated into DCE-MRI in a final categorization, as in PI-RADS, needs further investigation. Particularly, ADC values are influenced by many factors, such as measurement methods, MRI protocols and b values. In our study, two paired b

values (0–1,000 and 50–800 s/mm^2) were used to calculate ADC values by mono-exponential model. There were also some differences between ours and the mpMRI-based decision rules in the b values used to calculate ADC values, which may have impact on the results in the study. The cutoff of ADC values to identify breast malignancy was discordant in the decision rules by Istomin *et al.* and Zhong *et al.* This emphasizes the necessity of standardizing DWI and ADC values to facilitate their clinical application. A single cutoff of ADC values may be narrow and simplistic, considering the variation between readers and MRI protocols. An ADC category using multiple cutoffs may be more practical than a single cutoff of ADC values (48,49).

There are some limitations in our study. Firstly, the study was designed retrospectively and sample size was relatively small. Prospective studies with large samples are warranted to validate and further refine these decision rules, focusing on improving the specificity while maintaining high sensitivity. Secondly, we did not analyse the decision rules based on DCE-MRI. Breast mpMRI that consists of DCE-MRI, T2WI and DWI has become the standard scan protocol in clinical practice due to its superior performance over DCE-MRI (9,50). The decision rules based on mpMRI are aligned with clinical practice. Further studies can perform a comparative analysis between decision rules based on mpMRI and DCE-MRI. Thirdly, MRI scanners used in the study were not uniform. However, in our study, for DCE imaging, the time per dynamic acquisition was 60 seconds for all MRI scanners, in line with the recommended temporal resolution (51), and there were small differences in spatial resolution between the MRI scanners, which potentially had little impact on imaging evaluation. Additionally, the two paired b values used in the study conform to clinical practice (5,50), as there is no consensus on the optimal b values in breast DWI. From a clinical standpoint, the inhomogeneity of MRI scanners can be regarded as a strength of our study, as it reflects clinical practice. The inhomogeneity of MRI scanners can test the generalization ability of a decision rule, which is of critical importance for clinical use. Lastly, images were assessed by two breast radiologists by consensus, in accordance with our daily work. Inter-reader agreement and diagnostic efficiencies of readers with varying experience for the decision rules can be analyzed in further studies.

Conclusions

In conclusion, the mpMRI-based decision rules showed high sensitivity but low specificity for characterizing breast lesions and their performance was inferior to that of Kaiser score and BI-RADS category for all breast lesions. Kaiser score had a significantly higher specificity than the mpMRI-based decision rules and BI-RADS category but at the expense of decreased sensitivity. For BI-RADS 4 lesions, Kaiser score could reduce the number of unnecessary biopsy by 70.7%, while the mpMRI-based decision rules were of limited value.

Acknowledgments

Funding: The study was funded by Medical Health Science and Technology Project of Zhejiang Provincial Health Commission (Nos. 2021KY611 and 2021KY224), Chinese Medicine Science and Technology Project of Zhejiang Province (No. 2021ZB089), and Zhejiang Provincial Natural Science Foundation of China (No. LTGY24H180007).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-1783/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1783/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Institutional Ethical Board of The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine) (No. 2021-KL-062-01). Informed consent from patients was waived due to retrospective nature of the study.

References

1. Zhang L, Tang M, Min Z, Lu J, Lei X, Zhang X. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. Acta Radiol 2016;57:651-60.
2. Dijkstra H, Dorrius MD, Wielema M, Pijnappel RM, Oudkerk M, Sijens PE. Quantitative DWI implemented after DCE-MRI yields increased specificity for BI-RADS 3 and 4 breast lesions. J Magn Reson Imaging 2016;44:1642-9.
3. Rahbar H, Zhang Z, Chenevert TL, Romanoff J, Kitsch AE, Hanna LG, Harvey SM, Moy L, DeMartini WB, Dogan B, Yang WT, Wang LC, Joe BN, Oh KY, Neal CH, McDonald ES, Schnall MD, Lehman CD, Comstock CE, Partridge SC. Utility of Diffusion-weighted Imaging to Decrease Unnecessary Biopsies Prompted by Breast MRI: A Trial of the ECOG-ACRIN Cancer Research Group (A6702). Clin Cancer Res 2019;25:1756-65.
4. Cao Y, Wang X, Shi J, Zeng X, Du L, Li Q, Nickel D, Zhou X, Zhang J. Multiple parameters from ultrafast dynamic contrast-enhanced magnetic resonance imaging to discriminate between benign and malignant breast lesions: Comparison with apparent diffusion coefficient. Diagn Interv Imaging 2023;104:275-83.
5. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, Martincich L, Partridge SC, Patterson A, Pinker K, Thibault F, Camps-Herrero J, Le Bihan D; EUSOBI international Breast Diffusion-Weighted Imaging working group. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group. Eur Radiol 2020;30:1436-50.
6. Arponen O, Masarwah A, Sutela A, Taina M, Könönen M, Sironen R, Hakumäki J, Vanninen R, Sudah M. Incidentally detected enhancing lesions found in breast MRI: analysis of apparent diffusion coefficient and T2 signal intensity significantly improves specificity. Eur Radiol 2016;26:4361-70.
7. Ballesio L, Savelli S, Angeletti M, Porfiri LM, D'Ambrosio I, Maggi C, Castro ED, Bennati P, Fanelli GP, Vestri AR, Manganaro L. Breast MRI: Are T2 IR sequences useful in the evaluation of breast lesions? Eur J Radiol 2009;71:96-101.
8. Westra C, Dialani V, Mehta TS, Eisenberg RL. Using T2-weighted sequences to more accurately characterize breast masses seen on MRI. AJR Am J Roentgenol 2014;202:W183-90.
9. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. Radiology 2019;292:520-36.
10. Baltzer PAT, Kaiser WA, Dietzel M. Lesion type and reader experience affect the diagnostic accuracy of breast MRI: a multiple reader ROC study. Eur J Radiol 2015;84:86-91.
11. Grimm LJ, Anderson AL, Baker JA, Johnson KS, Walsh R, Yoon SC, Ghate SV. Interobserver Variability Between Breast Imagers Using the Fifth Edition of the BI-RADS MRI Lexicon. AJR Am J Roentgenol 2015;204:1120-4.
12. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, Tempany CM, Choyke PL, Cornud F, Margolis DJ, Thoeny HC, Verma S, Barentsz J, Weinreb JC. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. Eur Urol 2019;76:340-51.
13. Pinker K, Bickel H, Helbich TH, Gruber S, Dubsky P, Pluschnig U, Rudas M, Bago-Horvath Z, Weber M, Trattnig S, Bogner W. Combined contrast-enhanced magnetic resonance and diffusion-weighted imaging reading adapted to the "Breast Imaging Reporting and Data System" for multiparametric 3-T imaging of breast lesions. Eur Radiol 2013;23:1791-802.
14. Baltzer A, Dietzel M, Kaiser CG, Baltzer PA. Combined reading of Contrast Enhanced and Diffusion Weighted Magnetic Resonance Imaging by using a simple sum score. Eur Radiol 2016;26:884-91.
15. Fujiwara K, Yamada T, Kanemaki Y, Okamoto S, Kojima Y, Tsugawa K, Nakajima Y. Grading System to Categorize Breast MRI in BI-RADS 5th Edition: A Multivariate Study of Breast Mass Descriptors in Terms of Probability of Malignancy. AJR Am J Roentgenol 2018;210:W118-27.
16. Zhang M, Horvat JV, Bernard-Davila B, Marino MA, Leithner D, Ochoa-Albiztegui RE, Helbich TH, Morris EA, Thakur S, Pinker K. Multiparametric MRI model with dynamic contrast-enhanced and diffusion-weighted imaging enables breast cancer diagnosis with high accuracy. J Magn Reson Imaging 2019;49:864-74.
17. Istomin A, Masarwah A, Okuma H, Sutela A, Vanninen R, Sudah M. A multiparametric classification system for lesions detected by breast magnetic resonance imaging. Eur J Radiol 2020;132:109322.

18. Zhong Y, Li M, Zhu J, Zhang B, Liu M, Wang Z, Wang J, Zheng Y, Cheng L, Li X. A simplified scoring protocol to improve diagnostic accuracy with the breast imaging reporting and data system in breast magnetic resonance imaging. Quant Imaging Med Surg 2022;12:3860-72.
19. Kim KW, Kuzmiak CM, Kim YJ, Seo JY, Jung HK, Lee MS. Diagnostic Usefulness of Combination of Diffusion-weighted Imaging and T2WI, Including Apparent Diffusion Coefficient in Breast Lesions: Assessment of Histologic Grade. Acad Radiol 2018;25:643-52.
20. Baltzer PA, Dietzel M, Kaiser WA. A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography. Eur Radiol 2013;23:2051-60.
21. An Y, Mao G, Ao W, Mao F, Zhang H, Cheng Y, Yang G. Can DWI provide additional value to Kaiser score in evaluation of breast lesions. Eur Radiol 2022;32:5964-73.
22. Woitek R, Spick C, Schernthaner M, Rudas M, Kapetas P, Bernathova M, Furtner J, Pinker K, Helbich TH, Baltzer PAT. A simple classification system (the Tree flowchart) for breast MRI can reduce the number of unnecessary biopsies in MRI-only lesions. Eur Radiol 2017;27:3799-809.
23. Wengert GJ, Pipan F, Almohanna J, Bickel H, Polanec S, Kapetas P, Clauser P, Pinker K, Helbich TH, Baltzer PAT. Impact of the Kaiser score on clinical decision-making in BI-RADS 4 mammographic calcifications examined with breast MRI. Eur Radiol 2020;30:1451-9.
24. Milos RI, Pipan F, Kalovidouri A, Clauser P, Kapetas P, Bernathova M, Helbich TH, Baltzer PAT. The Kaiser score reliably excludes malignancy in benign contrast-enhancing lesions classified as BI-RADS 4 on breast MRI high-risk screening exams. Eur Radiol 2020;30:6052-61.
25. Zhang R, Xu M, Zhou C, Ding X, Lu H, Ge M, Du L, Bu Y. The value of noncontrast MRI in evaluating breast imaging reporting and data system category 0 lesions on digital mammograms. Quant Imaging Med Surg 2022;12:4069-80.
26. Dietzel M, Krug B, Clauser P, Burke C, Hellmich M, Maintz D, Uder M, Bickel H, Helbich T, Baltzer PAT. A Multicentric Comparison of Apparent Diffusion Coefficient Mapping and the Kaiser Score in the Assessment of Breast Lesions. Invest Radiol 2021;56:274-82.
27. Bougias H, Ghiatas A, Priovolos D, Veliou K, Christou A. Whole-lesion histogram analysis metrics of the apparent diffusion coefficient as a marker of breast lesions characterization at 1.5 T. Radiography (Lond) 2017;23:e41-6.
28. Kul S, Metin Y, Kul M, Metin N, Eyuboglu I, Ozdemir O. Assessment of breast mass morphology with diffusion-weighted MRI: Beyond apparent diffusion coefficient. J Magn Reson Imaging 2018;48:1668-77.
29. Wielema M, Dorrius MD, Pijnappel RM, De Bock GH, Baltzer PAT, Oudkerk M, Sijens PE. Diagnostic performance of breast tumor tissue selection in diffusion-weighted imaging: A systematic review and meta-analysis. PLoS One 2020;15:e0232856.
30. Murad MH, Lin L, Chu H, Hasan B, Alsibai RA, Abbas AS, Mustafa RA, Wang Z. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. CMAJ 2023;195:E925-31.
31. Gutierrez RL, DeMartini WB, Eby PR, Kurland BF, Peacock S, Lehman CD. BI-RADS lesion characteristics predict likelihood of malignancy in breast MRI for masses but not for nonmasslike enhancement. AJR Am J Roentgenol 2009;193:994-1000.
32. Basara Akin I, Balci P. Fibroadenomas: a multidisciplinary review of the variants. Clin Imaging 2021;71:83-100.
33. Kul S, Eyuboglu I, Cansu A, Alhan E. Diagnostic efficacy of the diffusion-weighted imaging in the characterization of different types of breast lesions. J Magn Reson Imaging 2014;40:1158-64.
34. Bennani-Baiti B, Bennani-Baiti N, Baltzer PA. Diagnostic Performance of Breast Magnetic Resonance Imaging in Non-Calcified Equivocal Breast Findings: Results from a Systematic Review and Meta-Analysis. PLoS One 2016;11:e0160346.
35. Spick C, Szolar DHM, Preidler KW, Tillich M, Reittner P, Baltzer PA. Breast MRI used as a problem-solving tool reliably excludes malignancy. Eur J Radiol 2015;84:61-4.
36. Youn I, Biswas D, Hippe DS, Winter AM, Kazerouni AS, Javid SH, Lee JM, Rahbar H, Partridge SC. Diagnostic Performance of Point-of-Care Apparent Diffusion Coefficient Measures to Reduce Biopsy in Breast Lesions at MRI: Clinical Validation. Radiology 2024;310:e232313.
37. Jajodia A, Sindhwani G, Pasricha S, Prosch H, Puri S, Dewan A, Batra U, Doval DC, Mehta A, Chaturvedi AK. Application of the Kaiser score to increase diagnostic accuracy in equivocal lesions on diagnostic mammograms referred for MR mammography. Eur J Radiol 2021;134:109413.
38. Baltzer PA, Benndorf M, Dietzel M, Gajda M, Runnebaum IB, Kaiser WA. False-positive findings at contrast-enhanced breast MRI: a BI-RADS descriptor study. AJR Am J Roentgenol 2010;194:1658-63.

39. Avendano D, Marino MA, Leithner D, Thakur S, Bernard-Davila B, Martinez DF, Helbich TH, Morris EA, Jochelson MS, Baltzer PAT, Clauser P, Kapetas P, Pinker K. Limited role of DWI with apparent diffusion coefficient mapping in breast lesions presenting as non-mass enhancement on dynamic contrast-enhanced MRI. Breast Cancer Res 2019;21:136.
40. Marino MA, Avendano D, Sevilimedu V, Thakur S, Martinez D, Lo Gullo R, Horvat JV, Helbich TH, Baltzer PAT, Pinker K. Limited value of multiparametric MRI with dynamic contrast-enhanced and diffusion-weighted imaging in non-mass enhancing breast tumors. Eur J Radiol 2022;156:110523.
41. Schlossbauer T, Leinsinger G, Wismuller A, Lange O, Scherr M, Meyer-Baese A, Reiser M. Classification of small contrast enhancing breast lesions in dynamic magnetic resonance imaging using a combination of morphological criteria and dynamic analysis based on unsupervised vector-quantization. Invest Radiol 2008;43:56-64.
42. Gibbs P, Onishi N, Sadinski M, Gallagher KM, Hughes M, Martinez DF, Morris EA, Sutton EJ. Characterization of Sub-1 cm Breast Lesions Using Radiomics Analysis. J Magn Reson Imaging 2019;50:1468-77.
43. Clauser P, Dietzel M, Weber M, Kaiser CG, Baltzer PA. Motion artifacts, lesion type, and parenchymal enhancement in breast MRI: what does really influence diagnostic accuracy? Acta Radiol 2019;60:19-27.
44. Dietzel M, Baltzer PA, Vag T, Gröschel T, Richter C, Burmeister H, Kaiser WA. Magnetic resonance mammography in small vs. advanced breast lesions - systematic comparison reveals significant impact of lesion size on diagnostic accuracy in 936 histologically verified breast lesions. Rofo 2011;183:126-35.
45. Meissnitzer M, Dershaw DD, Feigin K, Bernard-Davila B, Barra F, Morris EA. MRI appearance of invasive subcentimetre breast carcinoma: benign characteristics are common. Br J Radiol 2017;90:20170102.
46. Pötsch N, Korajac A, Stelzer P, Kapetas P, Milos RI, Dietzel M, Helbich TH, Clauser P, Baltzer PAT. Breast MRI: does a clinical decision algorithm outweigh reader experience? Eur Radiol 2022;32:6557-64.
47. Schieda N, Davenport MS, Silverman SG, Bagga B, Barkmeier D, Blank Z, Curci NE, Doshi AM, Downey RT, Edney E, Granader E, Gujrathi I, Hibbert RM, Hindman N, Walsh C, Ramsay T, Shinagare AB, Pedrosa I. Multicenter Evaluation of Multiparametric MRI Clear Cell Likelihood Scores in Solid Indeterminate Small Renal Masses. Radiology 2022;303:590-9.
48. Bickel H, Clauser P, Pinker K, Helbich T, Biondic I, Brkljacic B, Dietzel M, Ivanac G, Krug B, Moschetta M, Neuhaus V, Preidler K, Baltzer P. Introduction of a breast apparent diffusion coefficient category system (ADC-B) derived from a large multicenter MRI database. Eur Radiol 2023;33:5400-10.
49. Zuiani C, Mansutti I, Caronia G, Linda A, Londero V, Girometti R. Added value of the EUSOBI diffusion levels in breast MRI. Eur Radiol 2024;34:3352-63.
50. Lo Gullo R, Sevilimedu V, Baltzer P, Le Bihan D, Camps-Herrero J, Clauser P, Gilbert FJ, Iima M, Mann RM, Partridge SC, Patterson A, Sigmund EE, Thakur S, Thibault FE, Martincich L, Pinker K; EUSOBI International Breast Diffusion-Weighted Imaging working group. A survey by the European Society of Breast Imaging on the implementation of breast diffusion-weighted imaging in clinical practice. Eur Radiol 2022;32:6588-97.
51. Kuhl C. The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. Radiology 2007;244:356-78.

Cite this article as: An Y, Mao G, Zheng S, Bu Y, Fang Z, Lin J, Zhou C. External validation of multiparametric magnetic resonance imaging-based decision rules for characterizing breast lesions and comparison to Kaiser score and breast imaging reporting and data system (BI-RADS) category. Quant Imaging Med Surg 2025;15(1):648-661. doi: 10.21037/qims-23-1783

