Original Article

Quantitative evaluation of Kaiser score in diagnosing breast dynamic contrast-enhanced magnetic resonance imaging for patients with high-grade background parenchymal enhancement

Hui Wang[1,2], Ling Gao[1], Xu Chen[3], Shou-Ju Wang[2]

[1]Department of Radiology, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China; [2]Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China; [3]Department of Thyroid and Breast Surgery, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China

Contributions: (I) Conception and design: H Wang, SJ Wang; (II) Administrative support: SJ Wang, L Gao; (III) Provision of study materials or patients: SJ Wang, X Chen; (IV) Collection and assembly of data: H Wang, SJ Wang; (V) Data analysis and interpretation: H Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Shou-Ju Wang, MD. Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing 210000, China. Email: Shouju.wang@gmail.com.

Background: High-grade background parenchymal enhancement (BPE), including moderate and marked, poses a considerable challenge for the diagnosis of breast disease due to its tendency to increase the rate of false positives and false negatives. The purpose of our study was to explore whether the Kaiser score can be used for more accurate assessment of benign and malignant lesions in high-grade BPE compared with the Breast Imaging Reporting and Data System (BI-RADS).

Methods: A retrospective review was conducted on consecutive breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) scans from 2 medical centers. Included were patients who underwent DCE-MRI demonstrating high-grade BPE and who had a pathology-confirmed diagnosis. Excluded were patients who had received neoadjuvant chemotherapy or who had undergone biopsy prior to MRI examination. Two physicians with more than 7 years of experience specializing in breast imaging diagnosis jointly reviewed breast magnetic resonance (MR) images. The Kaiser score was used to determine the sensitivity, specificity, and positive predictive value (PPV), and negative predictive value (NPV) of the BI-RADS from different BPE groups and different enhancement types. The performance of the Kaiser score and BI-RADS were compared according to diagnostic accuracy.

Results: A total of 126 cases of high-grade BPE from 2 medical centers were included in this study. The Kaiser score had a higher specificity and PPV than did the BI-RADS (87.5% *vs.* 46.3%) as well as a higher PPV (94.3% *vs.* 79.8%). The value of diagnostic accuracy and 95% confidence interval (CI) for the Kaiser score (accuracy 0.928; 95% CI: 0.883–0.973) was larger than that for BI-RADS (accuracy 0.810; 95% CI: 0.741–0.879). Moreover, the Kaiser score had a significantly higher value of diagnostic accuracy for both mass and non-mass enhancement, especially mass lesions (Kaiser score: accuracy 0.947, 95% CI: 0.902–0.992; BI-RADS: accuracy 0.821, 95% CI: 0.782–0.860), with a P value of 0.006.

Conclusions: The Kaiser score is a useful diagnostic tool for the evaluation of high-grade BPE lesions, with a higher specificity, PPV, and diagnostic accuracy as compared to the BI-RADS.

Keywords: Kaiser score; Breast Imaging Reporting and Data System (BI-RADS); dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI); specificity; positive predictive value (PPV)

Submitted Jan 30, 2023. Accepted for publication Jul 28, 2023. Published online Aug 17, 2023.
doi: 10.21037/qims-23-113
View this article at: https://dx.doi.org/10.21037/qims-23-113

Introduction

Background parenchymal enhancement (BPE) is the normal physiological state of the breast and is seen as normal enhancement of the breast fibrous tissue in breast magnetic resonance imaging (MRI) after the injection of a contrast agent. BPE can be identified in imaging examinations (1-5). The 2013 edition of the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) Atlas increased the classification of BPE to minimal, mild, moderate, and marked. Although BPE is a normal state of the breast, the presence of BPE decreases lesion contrast, and early lesions are covered with light enhancement. This leads to missed diagnosis, blurred lesion edges, errors in the interpretation of the morphology of the lesions, and overestimation of the scope of the lesions (6,7). BI-RADS, a common diagnostic tool used every day, has a number of diagnostic descriptors that make it easier for physicians to discuss disease diagnosis. BI-RADS is currently recognized as the global standard for diagnosing breast diseases. However, the interpretation of images using BI-RADS may be subject to a certain degree of subjectivity, particularly for less experienced physicians, which could result in diagnostic errors (8-10). Because the presence of BPE increases the rate of false-negative and false-positive diagnoses, some studies have indicated that the diagnostic accuracy of MRI for high-grade BPE (moderate and marked) is lower than that for low-grade BPE (minimal and mild) (1,11-13); therefore, the use of BI-RADS in the diagnosis of high-grade BPE cases results in a high misdiagnosis rate.

In 2013, Baltzer and colleagues proposed the preliminary Kaiser score model, which is an intuitive flow chart that combines 5 separate diagnostic criteria: root sign, time-intensity curve (TIC), margins, internal enhancement patterns, and peritumoral edema. "Root sign" refers to any spicule-like margin irregularity, even in an otherwise smooth bordered lesion, while "peritumoral edema" refers to the presence of a markedly prolonged T2 of soft tissue not being caused by duct ectasia (14). The Kaiser score is based on 11 classifications and has a cutoff value of 4, with scores greater than 4 indicating malignancy (15). Some studies have shown that the Kaiser score can improve consistency among image readers (15-17). One meta-analysis (3) proposed that women who exhibit minimal or mild background enhancement do not have an elevated risk of breast cancer. Similar studies (1,5,11) have also suggested that women with high-grade BPE are more likely to develop breast cancer compared with women with minimal or mild BPE. With the increasing incidence of breast cancer, accurate preoperative imaging diagnosis is becoming increasingly vital. This study was thus designed to evaluate the diagnostic performance of the Kaiser score in patients with high-grade BPE. We aimed to analyze whether the Kaiser score is a valuable diagnostic tool for high-grade BPE lesions and to address the deficiencies in the diagnostic accuracy of BI-RADS. We present this article in accordance with the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-23-113/rc).

Methods

Study participants

This study was approved by the institutional review boards of 2 medical centers (The First Affiliated Hospital of Nanjing Medical University and Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine) and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Written informed consent was waived due to the retrospective nature of the study.

A retrospective review was conducted of consecutive breast dynamic contrast-enhanced MRI (DCE-MRI) scans from 2 medical centers between January 2021 and December 2021. The inclusion criteria were in accordance with the 2013 edition of the American College of Radiology BI-RADS Atlas. The first postcontrast subtraction image and patients with moderate and marked BPE were selected. Additional inclusion criteria were complete imaging data, diagnosis confirmed by biopsy or surgical pathology (pathology is the gold standard of diagnosis), clear image quality, complete scanning sequence, and no image artifacts. The exclusion criteria were patients who had neoadjuvant chemotherapy or biopsy before MRI examination. As a result, 126 cases were included in the study: 100 cases from institution 1 and 26 cases from institution 2 (*Figure 1*).

MRI protocol and interpretation

Institution 1

The First Affiliated Hospital of Nanjing Medical University. All patients underwent MRI in the prone position on 1.5-Tesla scanner (MAGNETON Aera XJ, Siemens Healthineers, Erlangen, Germany) units with the following imaging protocol: (I) axial turbo inversion recovery magnitude [repetition time (TR)/echo time (TE), 5,000/61 ms; field of view (FOV), 340 mm × 340 mm;

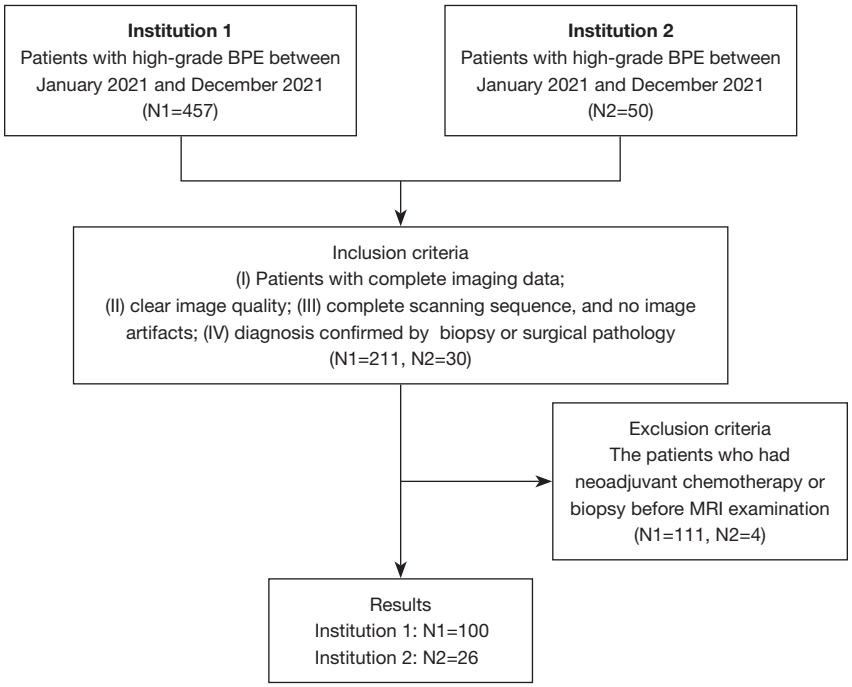

Figure 1 Flow diagram for the inclusion and exclusion of patients. BPE, background parenchymal enhancement; MRI, magnetic resonance imaging.

matrix, 576×403; flip angle, 80°; and slice thickness, 4 mm]; (II) axial diffusion-weighted imaging (b value, 50/800 s/mm^2; TR/TE, 5,400/86 ms; FOV, 360 mm × 180 mm; matrix, 192×82; flip angle, 180°; and slice thickness, 4 mm); (III) DCE sequence (TR/TE, 4.23/1.57 ms; FOV, 340 mm × 340 mm; matrix, 448×296; slice thickness, 1 mm; flip angle, 10°; pixel resolution, 1.1×0.8×0.9 mm^3; and temporal resolution, 1 min); and (IV) sagittal fat-suppressed T2-weighted imaging (TR/TE, 3,000/72 ms; FOV, 340 mm × 340 mm; matrix, 269×384; flip angle, 80°; and slice thickness, 4 mm). Gadolinium-diethylenetriaminepentaacetic acid (DTPA) (Magnevist, Bayer Healthcare, Berlin, Germany) was injected with an automated injector at a dose of 0.1 mmol/kg and a rate of 3 mL/s, which was followed by a 20-mL injection of saline solution.

Institution 2

Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine. All patients underwent MRI in the prone position on 3.0-Tesla scanner (MAGNETOM Prisma, Siemens Healthineers) units, with the following imaging protocol: (I) axial Dixon sequence (TR/TE, 3,500/93 ms; FOV, 384 mm × 384 mm; matrix, 320×320; flip angle, 80°; and slice thickness, 4 mm); (II) axial diffusion-weighted imaging (b value, 0/800 s/mm^2; TR/TE, 4,900/56 ms; FOV, 100 mm × 170 mm; matrix, 220×110; flip angle, 80°; and slice thickness, 4 mm); and (III) DCE sequence (TR/TE, 3.72/1.47 ms; FOV, 768 mm × 768 mm; matrix, 448×448; slice thickness, 1 mm; flip angle, 12°; pixel resolution, 0.8×0.8×1.0 mm^3; and temporal resolution, 1 min). Gadolinium-DTPA (Omniscan, GE HealthCare, Chicago, IL, USA) was injected with an automated injector at the dose of 0.5 mmol/mL and the rate of 2 mL/s, which was followed by a 16-mL injection of saline solution.

Image interpretation and management

The breast MRI images were reviewed in a double-blinded manner by 2 physicians with more than 7 years of experience specializing in breast imaging diagnosis. The Kaiser score, also known as a tree flow chart (18,19), consists of 5 diagnostic criteria derived from the BI-RADS and includes 11 outcome classifications. The score ranges from 1 to 11. The higher the Kaiser score is, the higher probability of a malignant tumor, and when the score exceeds 4, biopsy is recommended (15,17). If there is a

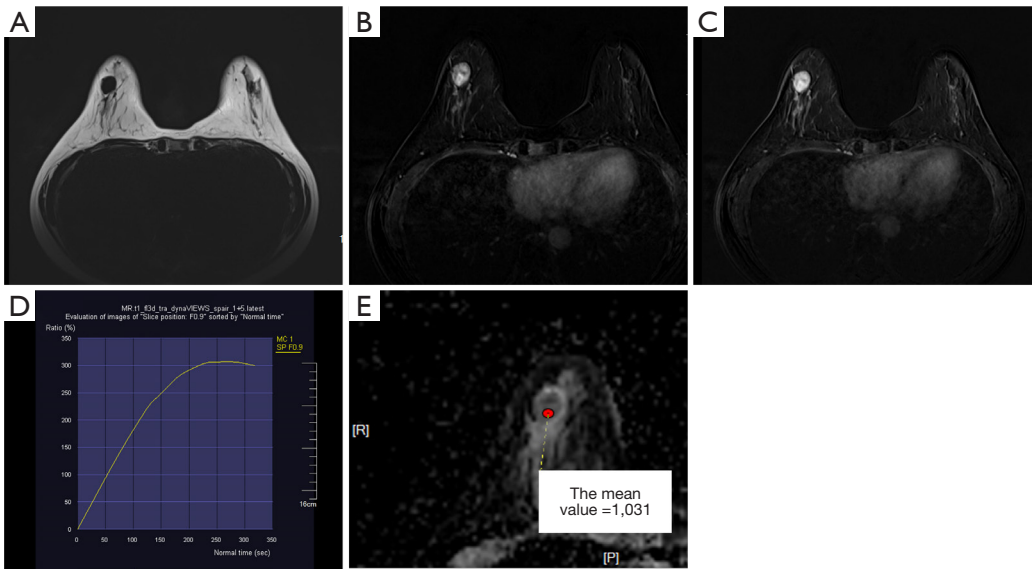

Figure 2 The Kaiser score diagnostic process. A 39-year-old woman with a mass found during physical examination. DCE-MRI showed a moderate BPE. The lesion was isointense on T1WI, with a smooth edge and no root sign observed on either early and late subtraction. The TIC was persistent type, and the mean ADC value obtained was 1.03×10^{-3} mm^2/s. The Kaiser score was 1 point, falling within the 1–4-point category and corresponding to BI-RADS 2/3. The final pathological result was adenosis. (A) T1WI; (B) the early subtraction of DCE-MRI; (C) the late subtraction of DCE-MRI; (D) TIC; (E) ADC map. DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; BPE, background parenchymal enhancement; T1WI, T1-weighted imaging; TIC, time-intensity curve; ADC, apparent diffusion coefficient; BI-RADS, Breast Imaging Reporting and Data System.

suspicious calcification on the corresponding mammogram, 2 points are added to the score. If the apparent diffusion coefficient value is greater than 1.4×10^3 mm^2/s, the total score is reduced by 4 points (*Figure 2*). Both physicians determined the Kaiser scores for all lesions, and for cases with differing initial opinions, the final result was obtained after a chief physician with 15 years of experience in breast diagnosis was consulted.

Data analysis

SPSS 23.0 (IBM Corp., Armonk, NY, USA) was used for the statistical analysis. The *t*-test and Fisher exact test were used to compare the benign and malignant cases of each group in the population data. Pathological results were used as diagnostic criteria to analyze the diagnosability between the Kaiser score and BI-RADS in different BPE groups and different lesion enhancement categories. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of each group were also calculated. The McNemar test was used for comparisons between groups, and $P<0.05$ was considered statistically significant.

Results

Patient cohort

A total of 126 cases were enrolled in this study, including 125 women and 1 man, comprising 40 benign cases (31.7%) and 86 malignant cases (68.3%). The mean ± standard deviation age among all patients was 44.2±10.0 years; the average age of benign cases was 39.0±8.7 years, and that of the malignant cases was 46.6±9.9 years. The difference in mean age between the benign and malignant cases was statistically significant ($P<0.001$). There were 124 total cases (98.4%) who were premenopausal, 40 (32.3%) of whom were in the benign group and 84 (67.4%) of whom were in the malignant group. Two postmenopausal patients had malignancies. Patients were grouped according to whether they had a family history of breast cancer: 7 cases (5.6%) had a family history of breast cancer, among whom 1 (14.3%) had a benign lesion and 6 (85.7%) had malignant lesions. There were 119 cases (94.4%) without a family history of breast cancer, among whom 39 (32.8%) had benign lesions and 80 (67.2%) had malignant lesions. No patients had any personal history of breast cancer (*Table 1*). Of the

Table 1 Demographic data (n=126)

General data	Total (n=126)	Begin (n=40)	Malignant (n=86)	P value
Age (years)	44.2±10.0	39.0±8.7	46.6±9.9	<0.001*
Menopausal status				0.335
Premenopausal	124 (98.4)	40 (32.3)	84 (67.4)	
Postmenopausal	2 (1.6)	0 (0)	2 (100.0)	
Family history of breast cancer				0.232
Yes	7 (5.6)	1 (14.3)	6 (85.7)	
No	119 (94.4)	39 (32.8)	80 (67.2)	
Personal history of breast cancer				–
Yes	0	0	0	
No	126 (100.0)	40 (31.7)	86 (68.3)	

Data are presented as the mean ± standard deviation or number (percentage). *, significantly different.

Table 2 Final histological characteristics of the biopsied lesions

Histological characteristics	Pathology	No. of participants
Benign (n=40)	Intraductal papilloma	11 (27.5)
	Phyllodes tumor	2 (5.0)
	Fibroadenoma	11 (27.5)
	Adenosis	9 (22.5)
	Sclerosing adenosis	3 (7.5)
	Idiopathic granulomatous mastitis	4 (10.0)
Malignant (n=86)	Invasive ductal carcinoma	42 (48.8)
	Invasive ductal carcinoma with ductal carcinoma in situ	31 (36.0)
	Ductal carcinoma in situ	10 (11.6)
	Papillary carcinoma	3 (3.5)

Data are presented as number (percentage).

40 benign lesions, 11 intraductal were papilloma (27.5%), 2 were phyllodes tumor (5%), 11 were fibroadenoma (27.5%), 9 were adenosis (22.5%), 3 were sclerosing adenosis (7.5%), and 4 were idiopathic granulomatous mastitis (10.0%). Of the 86 malignant lesions, 42 were invasive ductal carcinoma (48.8%), 31 were invasive ductal carcinoma with ductal carcinoma *in situ* (36.0%), 10 were ductal carcinomas *in situ* (11.6%), and 3 were papillary carcinoma (3.5%) (*Table 2*).

BPE groups

We divided all participants into 3 groups: group 1 was the high-grade BPE group, consisting of moderate and marked BPE; group 2 consisted of only moderate BPE; and group 3 consisted of only marked BPE. The diagnostic performances of the Kaiser score and BI-RADS for these groups were compared to evaluate the scores' test evaluation indices and diagnostic accuracy, with the 95% confidence intervals (CIs) being calculated. There were 126 cases in group 1, and the sensitivity, specificity, PPV, and NPV values for the Kaiser score were 95.3%, 87.5%, 94.3%, and 89.7%, respectively; meanwhile, the values for BI-RADS were 96.5%, 46.3%, 79.8%, and 90.4%, respectively. The diagnostic accuracy for Kaiser score was 0.928 (95% CI: 0.883–0.973), and that for BI-RADS was 0.810 (95% CI: 0.741–0.879). There were 93 cases in group 2, and the sensitivity, specificity, PPV, and NPV values for the Kaiser score were 95.2%, 93.6%, 96.7%, and 90.6%, respectively; meanwhile, the values for BI-RADS were 96.8%, 54.8%, 81.1%, and 89.5%, respectively. The diagnostic accuracy for the Kaiser score was 0.946 (95% CI: 0.901–0.991), and that for BI-RADS was 0.828 (95% CI: 0.752–0.904). There were 33 cases in group 3, and the sensitivity, specificity, PPV, and NPV values for the Kaiser score were 95.8%, 66.7%, 88.5%, and 85.7%, respectively; meanwhile, the values for BI-RADS were 95.8%, 22.2%, 76.7%, and 66.7%, respectively. The diagnostic accuracy for the Kaiser score was 0.879 (95% CI: 0.767–0.991), and that for BI-RADS was 0.758 (95% CI: 0.611–0.905). The P values of the diagnostic accuracy for group 1 and 2 were all <0.05, indicating statistical significance (*Table 3*).

Enhancement types

In accordance with the classification from the 2013 edition of the American College of Radiology BI-RADS, lesion enhancement can be divided into 3 types: focus, mass, and non-mass enhancement. For the 126 included cases, 6 (4.8%) had focus lesions, 95 (75.4%) had mass lesions, and 25 (19.8%) had non-mass enhancement lesions. Owing to the small number of cases with focus lesions, statistical analysis was not performed for this group. The sensitivity, specificity, PPV, and NPV for the Kaiser score for mass lesions were 95.7%, 92.0%, 97.1%, and 88.4%, respectively, and the diagnostic accuracy was 0.947 (95% CI: 0.902–0.992). The sensitivity, specificity, PPV, and NPV for BI-RADS were 96.5%, 47.5%, 79.8%, and 86.4%,

Table 3 Diagnosis by characteristics for the Kaiser score and BI-RADS based on DCE-MRI

Classification	Group	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	DAc[#]	95% CI	P value
BPE degree	All lesions (n=126)							0.001*
	Kaiser score	95.3	87.5	94.3	89.7	0.928	0.883–0.973	
	BI-RADS	96.5	46.3	79.8	90.4	0.810	0.741–0.879	
	Moderate BPE lesions (n=93)							0.007*
	Kaiser score	95.2	93.6	96.7	90.6	0.946	0.901–0.991	
	BI-RADS	96.8	54.8	81.1	89.5	0.828	0.752–0.904	
	Marked BPE lesions (n=33)							0.125
	Kaiser score	95.8	66.7	88.5	85.7	0.879	0.767–0.991	
	BI-RADS	95.8	22.2	76.7	66.7	0.758	0.611–0.905	
Enhancement type	Mass lesions (n=95)							0.006*
	Kaiser score	95.7	92.0	97.1	88.4	0.947	0.902–0.992	
	BI-RADS	96.5	47.5	79.8	86.4	0.821	0.782–0.860	
	Non-mass enhancement lesions (n=25)							0.125
	Kaiser score	93.8	88.9	93.8	88.9	0.920	0.814–1.026	
	BI-RADS	100.0	55.6	80.0	100.0	0.840	0.697–0.983	

*, significantly different; [#], determined as (true positive + true negative)/all. BI-RADS, Breast Imaging Reporting and Data System; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; PPV, positive predictive value; NPV, negative predictive value; CI confidence interval; DAc, diagnostic accuracy; BPE, background parenchymal enhancement.

respectively, and the diagnostic accuracy was 0.821 (95% CI: 0.782–0.860). In the non-mass enhancement lesions, the sensitivity, specificity, PPV, and NPV for the Kaiser score were 93.8%, 88.9%, 93.8%, and 88.9%, respectively, and the diagnostic accuracy was 0.920 (95% CI: 0.814–1.026). The sensitivity, specificity, PPV, and NPV for BI-RADS were 100.0%, 55.6%, 80.0%, and 100.0%, respectively, and the diagnostic accuracy was 0.840 (95% CI: 0.697–0.983). The P values of the diagnostic accuracy was 0.006 for mass lesions, indicating a statistically significant difference. However, the P value for non-mass enhancement lesions was 0.125 and not statistically significant (*Table 3*).

False positives and negatives for DCE-MRI

Among the 126 cases, 40 (31.7%) were benign and 86 (68.3%) were malignant. Among the benign lesions, 19 (47.5%) cases were true negatives and 21 (52.5%) cases were false positives using BI-RADS. Using the Kaiser score, 35 (87.5%) cases were true negatives and 5 (12.5%) cases were false positives. Among the 86 malignant cases, there were 83 (96.5%) true-positive cases and 3 (3.5%) false-negative cases using BI-RADS. Using the Kaiser score, 82 (95.3%) cases were true positive and 4 (4.7%) cases were false negative. Thus, with the Kaiser score, there were 5 false-positive cases and 4 false-negative cases. Among the 4 false-negative cases, there were 3 cases with moderate BPE and 1 with marked BPE, along with 3 with mass lesions and 1 with non-mass enhancement lesions. There were 2 cases of invasive ductal carcinomas with ductal carcinoma *in situ*, 1 case of ductal carcinoma *in situ*, and 1 case of papillary carcinoma. Among the 5 false-positive cases, there were 2 moderate BPE cases and 3 marked BPE cases, along with 2 focus cases, 2 mass cases, and 1 non-mass-enhanced case. The pathological results indicated 4 cases of intraductal papilloma and 1 case of idiopathic granulomatous mastitis (*Table 4*).

The diagnostic concordance between BI-RADS and the Kaiser score

Kappa statistic was employed to evaluate the diagnostic concordance to all lesions between BI-RADS and the Kaiser score. The kappa value was 0.515 while the P value

Table 4 Characteristics of false negatives and false positives using the Kaiser score

Variables	Case	BPE degree	Enhancement type	Pathology
False negative (n=4)	1	Moderate	Mass	Invasive ductal carcinoma with ductal carcinoma *in situ*
	2	Moderate	Mass	Invasive ductal carcinoma with ductal carcinoma *in situ*
	3	Moderate	Non-mass enhancement	Papillary carcinoma
	4	Marked	Mass	Invasive ductal carcinoma
False positive (n=5)	1	Moderate	Focus	Intraductal papilloma
	2	Moderate	Mass	Intraductal papilloma
	3	Marked	Non-mass enhancement	Idiopathic granulomatous mastitis
	4	Marked	Focus	Intraductal papilloma
	5	Marked	Mass	Intraductal papilloma

BPE, background parenchymal enhancement.

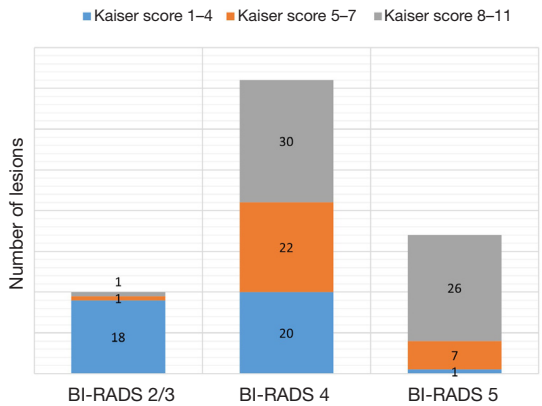

Figure 3 The total statistics of lesions grouped by the results of BI-RADS and Kaiser score. BI-RADS, Breast Imaging Reporting and Data System.

was <0.05, indicating a statistically significant difference. The moderate degree of concordance suggested by the kappa value indicated a certain degree of difference in the diagnosis between the BI-RADS and Kaiser score. The total results of all lesions diagnosed using the BI-RADS and Kaiser score were summarized for further study, as shown in *Figure 3*. The number of lesions was counted according to the BI-RADS group (2/3, 4, and 5) and Kaiser score group (1–4, 5–7, and 8–11). The results clearly show that the diagnosis between Kaiser score and BI-RADS groups 2/3 and 5 was consistent. However, there was no significant association between the diagnosis Kaiser score and BI-RADS group 4. This finding is consistent with the kappa statistic mentioned above. This additionally suggests that some of the lesions were incorrectly assessed in this group with BI-RADS, and thus the Kaiser score demonstrated a degree of superiority in accurate diagnosis.

Discussion

Compared with other breast examinations, DCE-MRI has high sensitivity (20); however, this imaging method is associated with a high rate of false-positive cases and low specificity. Several factors affect BPE, including both iatrogenic and physiological factors, with age, day of menstrual cycle, menopausal status, and lactational status also being associated with BPE (7). In the 2013 BI-RADS, BPE was officially included in the BI-RADS lexicon and is used to describe the extent of potential masking of suspicious lesions on breast MRI. Umatsu *et al.* (12) proposed BPE to be directly related to the impact of the diagnostic results, with higher BPE being associated with a higher degree of lesion cover-up, which is a limitation for accurate interpretation of breast MRI. Kaiser score is a clinical decision-making tool derived from machine learning that can guide image readers to evaluate the pathological conditions of lesions through an intuitive and progressively graded decision tree (15,17). The consistency of Kaiser scores has been proven in several studies (15,17,19,21,22). The results of BPE assessment using Kaiser scores are consistent even if they are derived from different units, MR machines, or field strengths. Our results showed that Kaiser scores have better diagnostic ability than do BI-RADS, with a specificity of 87.5%, which was 41.2% higher than that of BI-RADS at 46.3%; moreover, the Kaiser

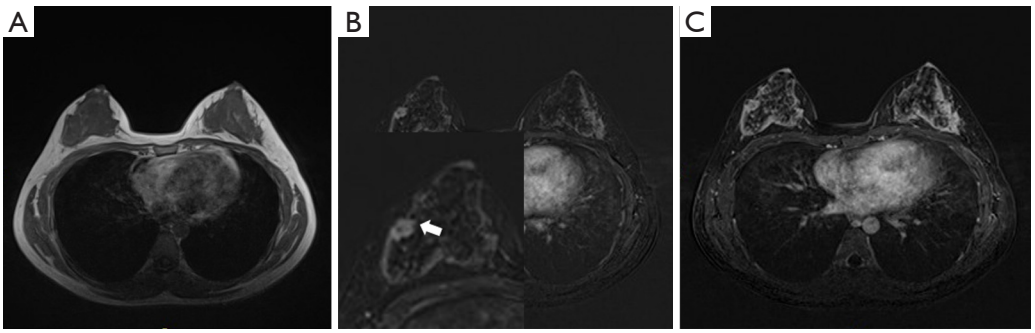

Figure 4 A false-positive case. The patient was a 25-year-old woman with a mass found on ultrasonography 1 week earlier. DCE-MRI showed marked BPE. The lesion was isointense on T1WI, the root sign was found around the lesion, the TIC curve was flat, and surrounding edema was present. The Kaiser score was 10, falling within the 8–11 category, the result was intraductal papilloma, and the BI-RADS grade was 4a. Both assessment methods yielded false-positive results. The images showed that the lesion was small, at approximately 10 mm × 9 mm in size; root sign could be seen around the lesion, with marked BPE and blurred lesion edges; and the TIC curve did not indicate a high possibility of a benign lesion. (A) T1WI; (B) the early subtraction of DCE-MRI, with the arrow indicating the root sign of the lesion; (C) the late subtraction of DCE-MRI. DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; BPE, background parenchymal enhancement; BI-RADS, Breast Imaging Reporting and Data System; T1WI, T1-weighted imaging; TIC, time-intensity curve.

score had a higher PPV of 94.3% compared to the 79.8% of BI-RADS. The Kaiser score can address the deficiencies of BI-RADS for diagnosis, with its low specificity and low PPV for high-grade BPE lesions (11,23), making the Kaiser score a helpful clinical decision tool. In the 3 groups in this study, we found advantages with the Kaiser score compared to BI-RADS for the total lesion group, moderate BPE group, and marked BPE group. The specificity and PPV values for the Kaiser score in each group were significantly higher than those of BI-RADS, as was the diagnostic accuracy. The Kaiser score had the best performance with the moderate BPE group, with the highest diagnostic accuracy, while the performance with the marked BPE group was lower than that of the other 2 groups. Our results showed that BPE was directly related to diagnostic accuracy, and high-grade BPE increased the extent of lesion occlusion and affected the display of the lesion edges and the morphology. Our conclusions are consistent with those of Umatsu *et al.* (12).

In the breast DCE-MRI examinations, mass lesions account for the majority of the lesion enhancement types, and lesion edge features are the most predictive single feature. TIC is also a key qualitative assessment, and the cancer risk ratio of lesions with washout curves is significantly (5:1) higher than that of lesions with persistent or plateau curves (24). Our data showed that the specificity and PPV of the Kaiser score were significantly higher than those of BI-RADS for mass lesions. For non-mass enhancement lesions, the misdiagnosis rate was higher with BI-RADS than with the Kaiser score because the BI-RADS descriptor results in a greater degree of overlap of benign and malignant lesions in morphology and hemodynamics. Our results showed that the specificity of the Kaiser score for the diagnosis of non-mass enhancement lesions was 33.3% higher than that of BI-RADS, and the PPV value of the Kaiser score was also higher than that of BI-RADS. These findings are in line with those reported by Marino *et al.* (16), who found that the Kaiser score provides better diagnostic performance for mass lesions compared with BI-RADS, and that the Kaiser score also has good diagnostic value for non-mass enhancement lesions.

BPE is a clinically important biomarker for breast cancer and an important predictor (2,4,5); however, the presence of high-grade BPE can lead to an increase in false-negative or false-positive results (25,26). Use of the Kaiser score in this study yielded 5 false-positive cases and 4 false-negative cases. We speculate that these misdiagnoses occurred for the following reasons: (I) with high-grade BPE, the biopsy recommendation rate of breast MRI reading is higher compared with lower grades of BPE (27). Owing to the particularity of the data in this study, the inclusion criteria were moderate and marked BPE, and among the 9 false-positive and false-negative cases, 6 (85.7%) of the lesions were smaller than 10 mm in diameter, making it almost impossible to separate the lesions from the breast parenchyma (28). (II) Among the 5 false-positive cases, 4 cases were intraductal papillomas (*Figure 4*) and all were

tumor-like papilloma types (29). Some scholars (30) argue that DCE-MRI cannot distinguish between tumor-like papillomas and malignant tumors. Furthermore, some tumor-like papillomas may have imaging characteristics similar to those of invasive malignant tumors; therefore, biopsy is essential. Pathologically, intraductal papilloma is often accompanied by peripheral fibrosis, which can lead to distortion of the papillary arrangement, resulting in blurred tumor edges, forming a "pseudo-infiltrating" structure. (III) The presence of heterogeneous tumors (31) can lead to a false-negative diagnosis, and this is especially true for ductal carcinoma in situ, which may show no enhancement on DCE-MRI. Among our 4 false-negative cases, there were 2 invasive ductal carcinomas with ductal carcinoma in situ and 1 ductal carcinoma in situ. In a previous study of 1,707 cases, 12 cases were false negatives (27), with invasive ductal carcinoma accounting for 0.3% of the lesions and ductal carcinoma *in situ* accounting for 0.4%; moreover, the proportion of false negatives was significantly lower than that in our study. However, other studies (32,33) have reported false-negative rates that are consistent with our results. We believe this discrepancy can be attributed to the variability in inclusion and exclusion criteria across the studies. We further found that 1 of the 5 false-positive cases diagnosed with the Kaiser score was assessed correctly by BI-RADS while 3 of the 4 false-negative cases were assessed correctly with BI-RADS. The reason of these misdiagnosis using the Kaiser score was examined in an analysis of the imaging results. In 3 of the 4 misdiagnosed cases, the diameter of the lesions was measured as all less than 10 mm while the morphology was also difficult to distinguish from that of high-grade BPE. We consider this to be the main reason for the misdiagnosis. Nonetheless, the Kaiser score still achieved superior diagnostic accuracy in the general lesions of high-grade BPE.

There are several limitations in this study. First, as a retrospective design was employed and the cohort of cases were limited, further, prospective research is needed with a larger number of cases. Second, considering the many factors that affect BPE, some cases in this study were not examined during the optimal period of breast MRI examination (the second week after menstruation); therefore, the results might have been affected.

Conclusions

The Kaiser score is a useful and advantageous clinical decision tool. The Kaiser score addresses the low specificity of BI-RADS and improves the diagnostic performance in high-grade BPE. The Kaiser score also has higher diagnostic performance for both mass lesions and non-mass enhancement lesions compared with BI-RADS. Owing to the large extent of lesion cover-up by high-grade BPE, misdiagnosis is more likely for small lesions and intraductal papilloma.

Acknowledgments

Funding: This study was supported by the National Natural Science Foundation of China (Nos. 82022034 and 81871420 to S Wang) and the Jiangsu Province Natural Science Foundation of China (No. BK20200032 to S Wang).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-113/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-113/coif). SJW was supported by grants from the National Natural Science Foundation of China (Nos. 82022034 and 81871420) and the Jiangsu Province Natural Science Foundation of China (No. BK20200032). The other 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 was approved by the institutional review boards of 2 medical centers (The First Affiliated Hospital of Nanjing Medical University and Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine). Written informed consent was waived due to the retrospective nature of the study.

References

1. Rella R, Bufi E, Belli P, Contegiacomo A, Giuliani M, Rosignuolo M, Rinaldi P, Manfredi R. Background parenchymal enhancement in breast magnetic resonance imaging: A review of current evidences and future trends. Diagn Interv Imaging 2018;99:815-26.
2. Sung JS, Corben AD, Brooks JD, Edelweiss M, Keating DM, Lin C, Morris EA, Patel P, Robson M, Woods M, Bernstein JL, Pike MC. Histopathologic characteristics of background parenchymal enhancement (BPE) on breast MRI. Breast Cancer Res Treat 2018;172:487-96.
3. Hu N, Zhao J, Li Y, Fu Q, Zhao L, Chen H, Qin W, Yang G. Breast cancer and background parenchymal enhancement at breast magnetic resonance imaging: a meta-analysis. BMC Med Imaging 2021;21:32.
4. Arasu VA, Miglioretti DL, Sprague BL, Alsheik NH, Buist DSM, Henderson LM, Herschorn SD, Lee JM, Onega T, Rauscher GH, Wernli KJ, Lehman CD, Kerlikowske K. Population-Based Assessment of the Association Between Magnetic Resonance Imaging Background Parenchymal Enhancement and Future Primary Breast Cancer Risk. J Clin Oncol 2019;37:954-63.
5. King V, Brooks JD, Bernstein JL, Reiner AS, Pike MC, Morris EA. Background parenchymal enhancement at breast MR imaging and breast cancer risk. Radiology 2011;260:50-60.
6. Chen L. Improving the understanding of background parenchymal enhancement of the breast. Chinese Journal of Radiology 2019;(9):721-2.
7. Baltzer PA, Dietzel M, Vag T, Burmeister H, Gajda M, Camara O, Pfleiderer SO, Kaiser WA. Clinical MR mammography: impact of hormonal status on background enhancement and diagnostic accuracy. Rofo 2011;183:441-7.
8. 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.
9. Pleasant V. Benign Breast Disease. Clin Obstet Gynecol 2022;65:448-60.
10. Zhang J, Cai L, Pan X, Chen L, Chen M, Yan D, Liu J, Luo L. Comparison and risk factors analysis of multiple breast cancer screening methods in the evaluation of breast non-mass-like lesions. BMC Med Imaging 2022;22:202.
11. Hambly NM, Liberman L, Dershaw DD, Brennan S, Morris EA. Background parenchymal enhancement on baseline screening breast MRI: impact on biopsy rate and short-interval follow-up. AJR Am J Roentgenol 2011;196:218-24.
12. Uematsu T, Kasami M, Watanabe J. Does the degree of background enhancement in breast MRI affect the detection and staging of breast cancer? Eur Radiol 2011;21:2261-7.
13. Telegrafo M, Rella L, Stabile Ianora AA, Angelelli G, Moschetta M. Effect of background parenchymal enhancement on breast cancer detection with magnetic resonance imaging. Diagn Interv Imaging 2016;97:315-20.
14. 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.
15. 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.
16. Marino MA, Clauser P, Woitek R, Wengert GJ, Kapetas P, Bernathova M, Pinker-Domenig K, Helbich TH, Preidler K, Baltzer PA. A simple scoring system for breast MRI interpretation: does it compensate for reader experience? Eur Radiol 2016;26:2529-37.
17. 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.
18. Dietzel M, Baltzer PAT. How to use the Kaiser score as a clinical decision rule for diagnosis in multiparametric breast MRI: a pictorial essay. Insights Imaging 2018;9:325-35.
19. Grippo C, Jagmohan P, Helbich TH, Kapetas P, Clauser P, Baltzer PAT. Correct determination of the enhancement curve is critical to ensure accurate diagnosis using the Kaiser score as a clinical decision rule for breast MRI. Eur J Radiol 2021;138:109630.
20. 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.
21. 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.
22. Meng L, Zhao X, Lu L, Xing Q, Wang K, Guo Y, Shang H, Chen Y, Huang M, Sun Y, Zhang X. A Comparative Assessment of MR BI-RADS 4 Breast Lesions With Kaiser Score and Apparent Diffusion Coefficient Value. Front Oncol 2021;11:779642.
23. Ray KM, Kerlikowske K, Lobach IV, Hofmann MB, Greenwood HI, Arasu VA, Hylton NM, Joe BN. Effect of Background Parenchymal Enhancement on Breast MR Imaging Interpretive Performance in Community-based Practices. Radiology 2018;286:822-9.
24. Schnall MD, Blume J, Bluemke DA, DeAngelis GA, DeBruhl N, Harms S, Heywang-Köbrunner SH, Hylton N, Kuhl CK, Pisano ED, Causer P, Schnitt SJ, Thickman D, Stelling CB, Weatherall PT, Lehman C, Gatsonis CA. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. Radiology 2006;238:42-53.
25. Niell BL, Abdalah M, Stringfield O, Raghunand N, Ataya D, Gillies R, Balagurunathan Y. Quantitative Measures of Background Parenchymal Enhancement Predict Breast Cancer Risk. AJR Am J Roentgenol 2021;217:64-75.
26. Jung Y, Jeong S, Kim JY, Kang DK, Kim TH. Correlations of female hormone levels with background parenchymal enhancement and apparent diffusion coefficient values in premenopausal breast cancer patients: Effects on cancer visibility. Eur J Radiol 2020;124:108818.
27. Song D, Kang BJ, Kim SH, Lee J, Park GE. The Frequency and Causes of Not-Detected Breast Malignancy in Dynamic Contrast-Enhanced MRI. Diagnostics (Basel) 2022;12:2575.
28. Vreemann S, Dalmis MU, Bult P, Karssemeijer N, Broeders MJM, Gubern-Mérida A, Mann RM. Amount of fibroglandular tissue FGT and background parenchymal enhancement BPE in relation to breast cancer risk and false positives in a breast MRI screening program : A retrospective cohort study. Eur Radiol 2019;29:4678-90.
29. Daniel BL, Gardner RW, Birdwell RL, Nowels KW, Johnson D. Magnetic resonance imaging of intraductal papilloma of the breast. Magn Reson Imaging 2003;21:887-92.
30. Wang W, Ding J, Yang W, Li Y, Zhou L, Zhang S, Zhu H, Mao J, Tang J, Gu Y, Peng W. MRI characteristics of intraductal papilloma. Acta Radiol 2015;56:276-83.
31. Zhu M, Pi Y, Jiang Z, Wu Y, Bu H, Bao J, Chen Y, Zhao L, Peng Y. Application of deep learning to identify ductal carcinoma in situ and microinvasion of the breast using ultrasound imaging. Quant Imaging Med Surg 2022;12:4633-46.
32. Obdeijn IM, Loo CE, Rijnsburger AJ, Wasser MN, Bergers E, Kok T, Klijn JG, Boetes C. Assessment of false-negative cases of breast MR imaging in women with a familial or genetic predisposition. Breast Cancer Res Treat 2010;119:399-407.
33. Shimauchi A, Jansen SA, Abe H, Jaskowiak N, Schmidt RA, Newstead GM. Breast cancers not detected at MRI: review of false-negative lesions. AJR Am J Roentgenol 2010;194:1674-9.

Cite this article as: Wang H, Gao L, Chen X, Wang SJ. Quantitative evaluation of Kaiser score in diagnosing breast dynamic contrast-enhanced magnetic resonance imaging for patients with high-grade background parenchymal enhancement. Quant Imaging Med Surg 2023;13(10):6384-6394. doi: 10.21037/qims-23-113

