Diffusion-weighted Imaging Allows for Downgrading MR BI-RADS 4 Lesions in Contrast-enhanced MRI of the Breast to Avoid Unnecessary Biopsy

Paola Clauser[1], Barbara Krug[2], Hubert Bickel[1], Matthias Dietzel[3], Katja Pinker[4], Victor-Frederic Neuhaus[2], Maria Adele Marino[5], Marco Moschetta[6], Nicoletta Troiano[6], Thomas H. Helbich[1], and Pascal A.T. Baltzer[1]

ABSTRACT

Purpose: Diffusion-weighted imaging with the calculation of an apparent diffusion coefficient (ADC) has been proposed as a quantitative biomarker on contrast-enhanced MRI (CE-MRI) of the breast. There is a need to approve a generalizable ADC cutoff. The purpose of this study was to evaluate whether a predefined ADC cutoff allows downgrading of BI-RADS 4 lesions on CE-MRI, avoiding unnecessary biopsies.

Experimental Design: This was a retrospective, multicentric, cross-sectional study. Data from five centers were pooled on the individual lesion level. Eligible patients had a BI-RADS 4 rating on CE-MRI. For each center, two breast radiologists evaluated the images. Data on lesion morphology (mass, non-mass), size, and ADC were collected. Histology was the standard of reference. A previously suggested ADC cutoff ($\geq 1.5 \times 10^{-3}$ mm^2/second) was applied. A negative likelihood ratio of 0.1 or lower was considered as a rule-out criterion for breast cancer. Diagnostic performance indices were calculated by ROC analysis.

Results: There were 657 female patients (mean age, 42; SD, 14.1) with 696 BI-RADS 4 lesions included. Disease prevalence was 59.5% (414/696). The area under the ROC curve was 0.784. Applying the investigated ADC cutoff, sensitivity was 96.6% (400/414). The potential reduction of unnecessary biopsies was 32.6% (92/282).

Conclusions: An ADC cutoff of $\geq 1.5 \times 10^{-3}$ mm^2/second allows downgrading of lesions classified as BI-RADS 4 on breast CE-MRI. One-third of unnecessary biopsies could thus be avoided.

Introduction

Because of its very high sensitivity, contrast-enhanced MRI (CE-MRI) has gained a pivotal role in breast cancer diagnosis. Breast MRI allows the detection of breast cancers not visible on mammography in women with dense breast tissue (1, 2) and can aid in the management of equivocal findings on mammography and ultrasound (3–5). The currently recommended standard breast MRI protocols include T2-weighted and CE T1-weighted sequences (6–8). The use of intravenous contrast agent is required, as the absence of enhancement practically rules out breast cancer and enhancement characteristics are useful to distinguish benign from malignant breast lesions. However, lesion characterization with CE-MRI remains challenging due to the overlapping characteristics of benign and malignant lesions (9, 10). This leads to a significant number of biopsy recommendations, a substantial number of which are benign, highlighting the potential to avoid harm and costs by improved pre-interventional lesion assessment (7, 9–11).

Diffusion-weighted imaging (DWI) is an imaging technique that enables quantitative evaluation of water diffusion hindrance in the extracellular space *in vivo*. The most commonly used metric in clinical practice is the apparent diffusion coefficient (ADC; refs. 12–14). Generally, high ADC values are rarely found in malignant lesions (15, 16) and are considered to be typical of benign lesions. This is why several investigators have measured ADC cut-off values that could be used to safely downgrade suspicious enhancing lesions initially classified as BI-RADS 4 on breast MRI, and avoid unnecessary interventions (17–20). The cutoffs suggested in these studies vary considerably. Furthermore, they were derived from the analyzed cases and they were not validated in an external dataset (17, 20). Only one previous study fulfilled the prospective trial conditions, providing a rigorously determined ADC cutoff (18). Nevertheless, an international expert consensus stated that it still remains unclear whether a predefined ADC cutoff with which to downgrade BI-RADS 4 lesions seen on CE-MRI is feasible across different centers, applying similar, but not identical, DWI technology (14).

Therefore, the aim of this study was to independently validate the effectiveness of the previously suggested ADC cutoff (18) with which to downgrade BI-RADS 4 lesions seen on CE-MRI in a multicentric study.

[1]Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria. [2]Department of Diagnostical and Interventional Radiology, University Hospital Cologne, Cologne, Germany. [3]Department of Radiology, Friedrich-Alexander-University Hospital Erlangen-Nürnberg, Erlangen, Germany. [4]Department of Radiology, Breast Imaging Service, Memorial Sloan Kettering Cancer Center, New York, New York. [5]Department of Biomedical Sciences and Morphologic and Functional Imaging, Policlinico Universitario G. Martino, University of Messina, Messina, Italy. [6]DETO Breast Care Unit, University of Bari Medical School, Bari, Italy.

Materials and Methods

Data collection

This was a retrospective, multicentric, mega-analysis. Data from five centers were pooled on the individual lesion level. Data were collected at five sites in three European countries and each single-center study was approved by the local Institutional Review Board (IRB). Because of the retrospective nature of the data analysis, the IRB waived the need for a signed informed consent. Data collection and aggregation was

Translational Relevance

Diffusion-weighted imaging (DWI) with the calculation of the apparent diffusion coefficient (ADC) has been proposed as a quantitative biomarker with which to downgrade lesions classified as BI-RADS 4 on contrast-enhanced MRI (CE-MRI) of the breast to potentially avoid unnecessary biopsies. To facilitate the clinical implementation of ADC, there is a need to approve a generalizable ADC cutoff. We tested a previously recommended cut-off value ($\geq 1.5 \times 10^{-3}$ mm^2/second) on a multicentric dataset of 696 lesions initially classified as suspicious (BI-RADS 4) at CE-MRI of the breast. We found that this ADC cutoff would have allowed downgrading of lesions initially classified as BI-RADS 4, with a potential reduction of unnecessary biopsies by 32.6% and a sensitivity of 96.6%. These results suggest that a single ADC cutoff could be used to downgrade lesions initially classified as BI-RADS 4 on CE-MRI, regardless of MRI vendor, DWI sequence, or reading center.

Figure 1.
Flowchart showing the included and excluded cases. The standard of reference was histopathology. BI-RADS: Breast Imaging Reporting and Data System.

performed in a fully anonymized way and in line with international legislation. Part of these data was already published as single-center studies in different scientific contexts (see Supplementary Materials and Methods S1). The study was performed in accordance with the Declaration of Helsinki statement for medical research involving human subjects.

Patients undergoing CE-MRI of the breast according to current recommendations (11, 21) were eligible for this study. Indications to perform CE-MRI of the breast included work-up of suspicious clinical abnormalities, equivocal findings on mammography and/or ultrasound, high-risk screening, and staging for a known breast cancer (further details are given in Supplementary Materials and Methods S1). Inclusion criteria were CE-MRI protocol that included DWI sequences, presence of an enhancing lesion in the examination, lesion classified as BI-RADS 4 (6) on CE-MRI, and availability of a standard of reference defined as histopathologic verification obtained with image-guided biopsy or surgery. Note that DWI results were not taken into consideration at the initial BI-RADS category assignment.

Exclusion criteria were nondiagnostic examinations (examination interrupted by the patient, artifacts); examinations performed during or after neoadjuvant chemotherapy; and lesions classified as BI-RADS 2, 3, or 5 on CE-MRI. The data selection process is shown in **Fig. 1**.

CE-MRI acquisition

Breast MRI was performed on either 1.5 or 3 T scanners, with dedicated breast coils and with patients lying in a prone position. All protocols followed international guidelines and recommendations (6, 21) and included a T2-weighted sequence and a T1-weighted series acquired before and after the injection of a gadolinium-based contrast agent. The technical parameters of the DWI sequences are shown in Supplementary Materials and Methods S2. In all centers, ADC maps used for the evaluation were generated by inline mono-exponential fitting of the highest and lowest b-value data by the scanner software.

Image interpretation and ADC measurements

The readings were performed by on-site readers for each center, with variable readers across sites. All readers were experienced breast radiologists and/or breast imaging fellows. The readers were blinded to the clinical history and histological results. Measurements were performed using dedicated workstations. Readers evaluated lesion
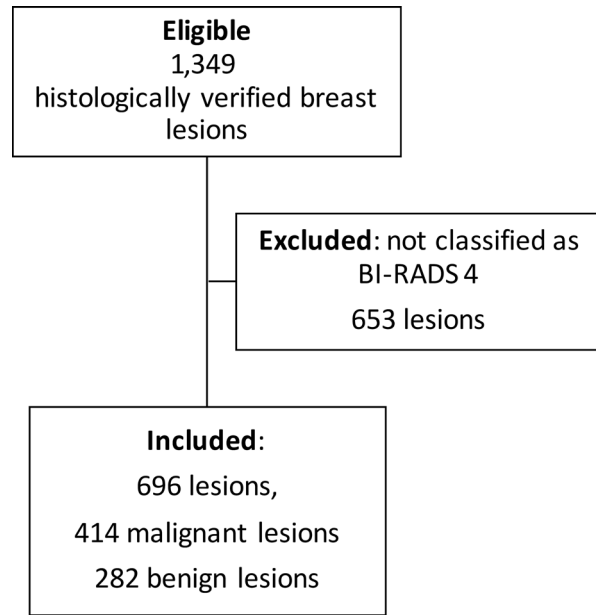
type, lesion size, and assigned a BI-RADS category (6) to each lesion. The same readers measured the mean ADC values either by drawing a small region of interest (ROI) in the area of lowest diffusion, which was, thus, the area within the lesion with the lowest signal intensity on ADC and brightest on DWI (data subsets 1–6, see Supplementary Materials and Methods S1), or by covering with an ROI the whole lesion while excluding areas of necrosis, hemorrhage, or cystic areas within the lesion (data subset 7). These ROI types have been classified in ref. 22 and are considered diagnostically equivalent.

Statistical analysis

The prevalence of breast cancer was calculated as the number of cancers in the total number of included lesions. The area under the ROC curves (AUC) were calculated for the entire database considering histopathology (i.e., benign vs. malignant) as the standard of reference, and separately for mass and non-mass lesions, as well as for lesions with a maximum diameter equal to or below 10 mm and above 10 mm. Differences between independent AUCs were calculated using the Hanley and McNeil methodology. We aimed at independently validating the cutoff $\geq 1.5 \times 10^{-3}$ mm^2/second, as reported by Rahbar and colleagues (18) in their prospective trial. To provide context, diagnostic performance indices both for lower [1.4×10^{-3} mm^2/second (20)] and higher [1.6×10^{-3} mm^2/second, 1.8×10^{-3} mm^2/second (15, 17)] ADC cut-off values are provided as Supplementary Materials and Methods S3. Considering interreader variability (23), cutoffs where rounded to one digit.

We defined a clinically acceptable ADC cutoff to rule out breast cancer if a negative likelihood ratio of 0.1 was achieved (24). As only lesions with a biopsy recommendation (BI-RADS 4) were examined, specificity equals the rate of reduced false-positives or potentially avoidable biopsies at this ADC cutoff. The cutoff was applied to calculate sensitivity, specificity, negative and positive predictive value (NPV and PPV), and negative and positive likelihood ratios (LR−, LR+). ADC values $\geq 1.5 \times 10^{-3}$ mm^2/second were considered

indicative of benignity (i.e., malignancy could be excluded); ADC values $< 1.5 \times 10^{-3}$ mm^2/second were considered potentially malignant (i.e., malignancy could not be excluded). Results are presented with 95% confidence intervals (95% CI). Independent proportions were compared using the $n-1 \chi^2$ test.

The Fagan nomogram was used to calculate and visualize posttest probabilities depending on pretest probabilities. This analysis was performed to clarify, whether the investigated ADC cutoff would be applicable to BI-RADS 4 lesions with low, intermediate, or high risk of breast cancer alike. The pretest probability generally reflects breast cancer prevalence. The lower the pretest probability is, the lower is the LR− needed to achieve a posttest probability $\leqq$ 2% that would formally allow to downgrade a BI-RADS 4 lesion to BI-RADS 3. The individual pretest probability can be estimated by using clinical or imaging findings to calculate cancer risk. On the basis of the calculated LR−, a maximum disease prevalence (pretest probability) was estimated, for which a negative test (ADC cutoff $\geq 1.5 \times 10^{-3}$ mm^2/second) would determine a posttest probability $\leqq$ 2% (BI-RADS 3 cutoff), and consequently, eliminate the need for biopsy. An open online source was used for calculations (http://araw.mede.uic.edu/cgi-bin/testcalc.pl).

To analyze potential technical confounders on the diagnostic results of the applied ADC cutoff, univariate fixed effects inverse variance meta-regression was performed. The target variables were sensitivity and specificity at the applied ADC cutoff $\geq 1.5 \times 10^{-3}$ mm^2/second; magnetic field strength, DWI b-values, MRI vendor, type of ROI, and acquisition before or after gadolinium-based contrast material injection were used as independent predictors.

The calculations were performed using IBM SPSS Statistics v. 20.0.0 (IBM) and MedCalc v. 12.5.0.0 (MedClac Software) and OpenMeta analyst (http://www.cebm.brown.edu).

The P values from statistical tests were interpreted as indicating low ($P < 0.05$), moderate ($P < 0.01$), or strong ($P < 0.001$) evidence.

Results

Lesion characteristics

From the initial database, 657 female patients formed the study dataset [mean age (years), 42; range, 18–91; SD, 14.1], with 696 lesions (Fig. 1). By histology, 282 lesions (40.5%) were classified as

benign and 414 (59.5%) as malignant, giving a disease prevalence of 59.5% (Table 1).

Lesions presented as masses in 512 of 696 cases (73.6%) and non-masses in 184 of 696 cases (26.4%). Disease prevalence was 68.8% (352/512) for mass lesions and 33.7% (63/184) for non-mass lesions ($P < 0.001$).

Lesion diameter ranged from 3 to 100 mm (mean, 20.6 mm; SD, 16.4). Mean lesion diameter was 20.0 mm for masses (SD, 15.6) and 22.6 mm for non-masses (SD, 18.3). The database included 220 lesions with a maximum diameter equal to or below 10 mm (31.6%) and 476 lesions with a diameter above 10 mm (68.4%). Disease prevalence was 39.5% (87/220) for lesions $\leqq$ 10 mm and 68.7% (327/476) for lesions > 10 mm ($P < 0.001$). The technical DWI success rate ranged between 82.4% and 92.5% and was provided by five of seven data subsets (see Supplementary Materials and Methods S1).

Diagnostic performance

The AUC of ADC values to distinguish benign from malignant lesions was 0.784 (Fig. 2). At the ADC cutoff of $\geq 1.5 \times 10^{-3}$ mm^2/second, sensitivity was 96.6%, with a LR− of 0.1 (Table 2). A total of 106 lesions presented ADC values that exceeded this threshold, 92 benign and 14 malignant. To provide context, alternative ADC thresholds and corresponding diagnostic parameters are provided in Supplementary Materials and Methods S3.

Of the 282 benign lesions, 92 would have been accurately classified as nonsuspicious, and a biopsy would have been correctly avoided, with a reduction in false-positives of 32.6%.

In the remaining 190 cases, the ADC cutoff did not alter the conventional BI-RADS 4 category assignment. These 190 false-positives included 79 fibrosis, fibrocystic changes, or fibroadenomatoid hyperplasia; 28 adenosis or complex sclerosing adenosis; 16 inflammatory changes; 15 papillomas; 15 atypical ductal hyperplasia or lobular neoplasia; 10 fibroadenomas; and 27 other lesions (i.e., pseuodoangiomatous stromal hyperplasia, epithelial proliferation, flat epithelial atypia, apocrine metaplasia).

Fourteen of 414 malignant lesions (3.4%) would have been incorrectly classified as nonsuspicious. These 14 false-negatives were six ductal carcinomas in situ (DCIS), four invasive mucinous carcinomas, two invasive cancers (nonspecial type), and two

Table 1. Detailed histology of the lesions included in the study.

Histology	Total N° (%)	Mass N° (%)	Non-mass N° (%)
Malignant	414/696 (59.5)	352/512 (68.7)	62/184 (33.7)
Invasive carcinoma NST	277	258 (93.1)	19 (6.9)
Ductal carcinoma in situ	58	29 (50.0)	29 (50.0)
Invasive lobular carcinoma	62	52 (83.8)	10 (16.1)
Mucinous carcinoma	10	6 (60.0)	4 (40.0)
Other malignant lesions	7	7 (100)	0 (0.0)
Benign	282/696 (40.5)	160/512 (31.3)	122/184 (66.3)
Fibrosis/FCC/fibroadenomatoid hyperplasia	136	73 (53.7)	63 (46.3)
Radial scar/CSA/adenosis	34	18 (52.9)	16 (47.1)
Fibroadenoma	26	22 (84.6)	4 (15.4)
ADH/LN	26	11 (42.3)	15 (57.7)
Abscess/inflammation	16	9 (56.3)	7 (43.7)
Papilloma/papillomatosis	16	9 (56.3)	7 (43.7)
Fat necrosis	3	1 (33.3)	2 (66.7)
Other benign lesions	25	17 (68.0)	8 (32.0)

Abbreviations: ADH, atypical ductal hyperplasia; CSA, complex sclerosing adenosis; FCC, fibrocystic changes; LN, lobular neoplasia; NST, nonspecial type.
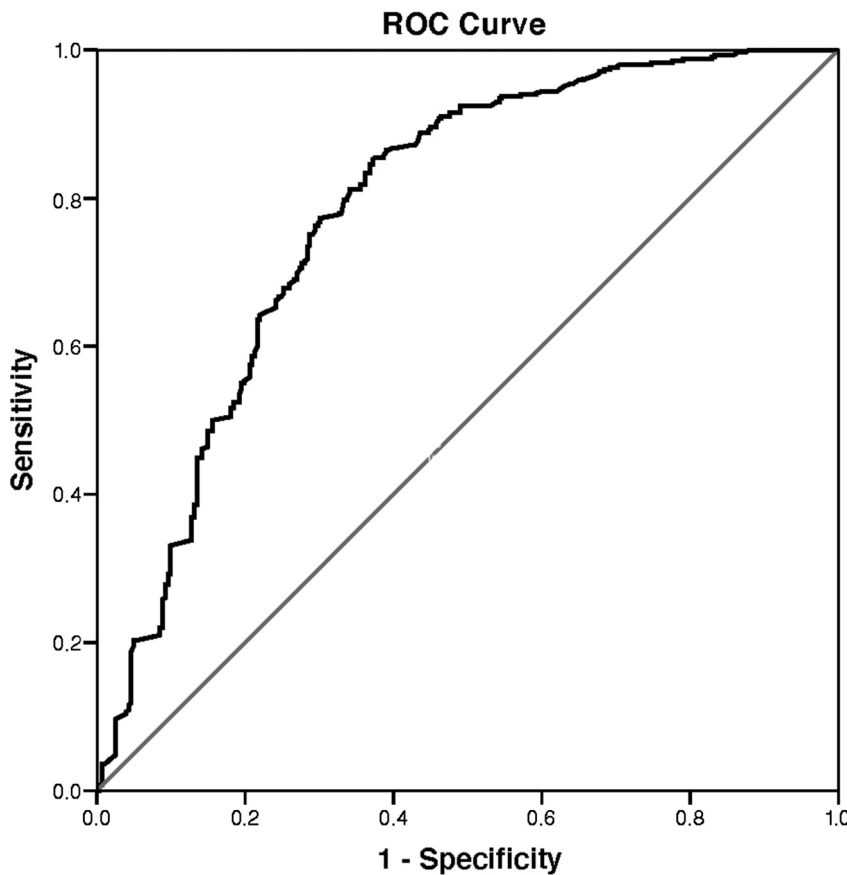
ROC Curve

Figure 2.
AUC for the ADC values to distinguish benign from malignant breast lesions initially classified as suspicious (BI-RADS 4) at CE-MRI.

invasive lobular carcinomas. Details are given in Supplementary Materials and Methods S4.

Influence of lesion characteristics on ADC cut-off performance

The AUC was significantly higher for mass lesions than for non-mass lesions ($P = 0.001$) and for lesions above 10 mm in size ($P = 0.036$). Sensitivity and specificity were nominally lower in non-mass lesions compared to mass lesions, but only the difference in sensitivity was significant on the low evidence level ($P = 0.028$) while specificity was not ($P = 0.222$; **Table 2**). Sensitivity and specificity at the applied cutoff did not differ depending on lesion size ($P > 0.05$; **Table 2**).

The AUC for lesions $\leqq 10$ mm was lower than for lesions > 10 mm, also considering mass lesions and non-mass lesions separately. The AUC difference for lesions $\leqq 10$ versus >10 mm was more evident for masses (0.780 vs. 0.863, $P = 0.065$) than for non-mass lesions (0.689 vs. 0.678, $P = 0.900$).

Details on number of false-positives and false-negatives divided per lesion morphology and size are given in **Table 3**.

Influence of technical confounders on ADC cut-off performance

Univariate meta-regression (Supplementary Materials and Methods S5) did not identify a significant effect of technical confounders on neither sensitivity nor specificity.

Pretest and posttest probabilities

The negative likelihood ratio calculated on the full study dataset was 0.10. This value would allow to exclude breast cancer with a resulting posttest probability of $\leqq 2\%$ if the underlying prevalence of malig-

nancy (pretest probability) does not exceed 17% (**Fig. 3**). The LR− ranged between 0.47 and <0.01 within the seven separate study subsets, with five of seven (71.4%) subsets achieving an LR− of 0.1 or less (**Fig. 3**).

In the subgroup analysis, a higher LR− was found in non-mass lesions (0.28) compared with mass lesions (0.07) and LR− was higher in lesions >10 mm (0.11) compared with lesions $\leqq 10$ mm (0.07).

Discussion

We validated the diagnostic performance of an ADC cutoff, suggested previously by a prospective multicentric trial (18), by applying it to an independent large, heterogeneous, multicentric dataset. Our results confirmed a cutoff of $\geq 1.5 \times 10^{-3}$ mm^2/second as suitable to downgrade lesions initially classified as BI-RADS 4. This ADC cutoff would have allowed a potential reduction of unnecessary biopsies by 32.6%.

Several studies have pointed out the possibility of using the ADC to rule out breast cancer, thereby potentially avoiding unnecessary breast biopsies (15, 17, 19, 20). In the only prospective trial, Rahbar and colleagues (18) defined a threshold to decrease false-positives in CE-MRI of the breast. They selected the ADC cutoff that reached 100% sensitivity on the ROC analysis ($\geq 1.53 \times 10^{-3}$ mm^2/second). Using this cutoff, they reported a potential reduction of unnecessary biopsies by 35.9%. Our study confirms this ADC cutoff in a considerably large, multicentric dataset across independent centers, MRI vendors, and readers. In addition, we obtained a potential reduction in false-positives of 32.6%, which is in line with Rahbar and colleagues (18).

Table 2. AUC, sensitivity, specificity, NPV, PPV, negative likelihood ratio (LR−), and positive likelihood ratio (LR+) for all lesions were calculated using a cutoff ADC of $\geq 1.5 \times 10^{-3}$ mm^2/second.

	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	NPV (95% CI)	PPV (95% CI)	LR− (95% CI)	LR+ (95% CI)
Total	0.784 (0.752–0.814)	400/414 96.6% (94.3–98.1)	92/282 32.6% (28.2–37.4)	92/106 86.8% (83.1–89.8)	400/590 67.8% (63.0–72.2)	0.1 (0.06–0.18)	1.43 (1.32–1.56)
Mass lesions	0.816 (0.779–0.848)	343/352 97.4% (95.3–98.7)	57/160 35.6% (31.0–40.5)	57/66 86.4% (82.6–89.4)	343/446 76.9% (72.5–80.8)	0.07 (0.04–0.14)	1.51 (1.35–1.70)
Non-mass lesions	0.671 (0.598–0.738)	57/62 91.9% (88.8–94.3)	35/122 28.7% (24.4–33.4)	35/40 87.5% (83.8–90.5)	57/144 39.6% (34.9–44.5)	0.28 (0.12–0.68)	1.29 (1.13–1.47)
Size ≦10 mm	0.748 (0.685–0.804)	85/87 97.7% (95.6–98.8)	41/133 30.8% (26.5–35.6)	41/43 95.3% (92.7–97.1)	85/177 48.0% (43.1–53.0)	0.07 (0.02–0.30)	1.41
Size >10 mm	0.828 (0.791–0.861)	315/327 96.3% (93.9–97.8)	51/149 34.2% (29.7–39.0)	51/63 81.0% (76.8–84.6)	315/413 76.3% (71.8–80.2)	0.11 (0.06–0.20)	1.46 (1.30–1.65)

Note: In addition, mass and non-mass lesions and lesions with a maximum diameter equal to and below or above 10 mm were considered separately.
Abbreviation: 95% CI, 95% confidence interval.

Retrospective studies proposed various ADC cutoffs to exclude malignancy, ranging between 1.4 and 1.8×10^{-3} mm^2/second. Woodhams and colleagues aimed at maximum sensitivity using a cutoff of 1.6×10^{-3} mm^2/second (15), but they did not investigate the potential of reducing false-positive findings. Baltzer and colleagues (20) suggested a cutoff of 1.4×10^{-3} mm^2/second to rule out malignancy by means of a scoring system, reporting an improvement in specificity of 10.4%. The study did not draw conclusions on the number of avoidable biopsies. This was done by Partridge and colleagues (17), who selected a cutoff at 100% sensitivity (1.8×10^{-3} mm^2/second) to potentially reduce 33% of false-positive biopsies in lesions initially classified as BI-RADS 4 and 5. As no independent validation across centers, vendors, and readers was performed, there currently is no generally accepted cutoff for ADC measurements with which to exclude breast cancer (14).

ADC values depend on several technical factors: the administration of gadolinium-based contrast, the choice of b-values and the ROI method applied to measure the ADC can affect ADC metrics (22, 25). Though no significant influence of these factors on the overall diagnostic performance of DWI has been confirmed (22, 25), it remains a matter of debate whether technical confounders challenge the application of an ADC cutoff across centers (14). Our results do not indicate a significant influence of technical factors on diagnostic sensitivity and specificity, and thus support the broad applicability of the investigated ADC cutoff. This does, however, not challenge the importance of a systematic approach on DWI standardization as suggested in ref. 14.

Avoiding unnecessary biopsies comes at the price of risking false-negative cases. In our study, the majority of false-negative lesions were DCIS (six of 58 intraductal carcinomas in the dataset were incorrectly classified) or mucinous type invasive carcinomas (four of 10 mucinous carcinomas in the dataset). DCIS (26, 27) and mucinous carcinomas (28, 29) are characterized by higher ADC values, therefore, the cutoff of $\geq 1.5 \times 10^{-3}$ mm^2/second is prone to miss these lesions. A second, higher, ADC cutoff, though, could be difficult to apply in clinical practice. Applying a higher ADC cutoff would lead to a minor increase of sensitivity, while sacrificing most of the benefits in terms of reduction of avoidable biopsies (Supplementary Materials and Methods S3; refs. 27 and 29). Therefore, it may be rather appropriate to suggest a 12-month follow-up in lesions exceeding the ADC cutoff validated in this article, as possible false-negative lesions usually show a less aggressive biological behavior. This management will safely avert negative outcomes, as disease progression in such time frames is unlikely.

We wondered whether there are specific lesion characteristics that are associated with a better or worse diagnostic performance of the ADC to diagnose cancer. Upon subgroup analysis, we found a reduced AUC for lesions with a maximum diameter equal to or smaller than 10 mm. A similar finding was also reported by Rahbar and colleagues (18). In our dataset, this reduction was related to a reduction in specificity, rather than in sensitivity, thereby not challenging the clinical applicability of the tested ADC cutoff. DWI sequences have a lower spatial resolution compared with contrast and T2-weighted sequences; thus, small lesions might be more difficult to detect and evaluate (14). It could be inferred that the ADC cutoff can be applied to all lesions, regardless of the maximum diameter, as long as the lesion is clearly visible on DWI and the ADC map. This is backed up by a report from Partridge and colleagues (30), who found no effect of size on ADC performance in a set of 166 mass and non-mass lesions. Conflicting results were obtained by Wan and colleagues (31), in a study including only mass lesions. The authors found a significantly lower sensitivity

Table 3. Number of false-negative and false-positive results obtained after applying the ADC cutoff $\geq 1.5 \times 10^{-3}$ mm^2/second, divided per lesion morphology and size.

| | Mass | | Non-mass | | |
	Size $\leq$ 10 mm	Size > 10 mm	Size $\leq$ 10 mm	Size > 10 mm	Total
False-negative	2/70	7/282	0/17	5/45	14/414
	2.9%	2.5%	0.0%	11.1%	3.4%
False-positive	62/92	41/68	30/41	57/81	190/282
	67.4%	60.3%	73.2%	70.4%	67.4%

Note: Denominators are the total number of malignant lesions per group for false-negatives and the total number of benign lesions for false-positives.

and specificity for lesions below 10 mm. Their study included only 22 lesions below 10 mm; thus, the different results might be related to a different case selection.

We found a lower AUC for non-mass lesions compared with mass lesions. In line with our own findings, the Rahbar and colleagues (18) reported a lower biopsy rate reduction in non-mass compared with mass lesions by keeping the same ADC cutoff. Their rate of avoidable unnecessary biopsies in benign lesions differed by 24.5% between mass and non-mass lesions but did not prove statistical significance ($P = 0.136$). Our results show a similar, but lower difference in specificity of 6.9% between mass and non-mass lesions,

also not proving statistical significance. Still, non-mass lesions are a known diagnostic dilemma, and constitute a relevant portion of false-positive findings in breast MRI (32, 33). The evaluation of ADC for non-mass lesions is more challenging due to the absence of an obviously defined space-occupying nodule. In addition, the majority of DCIS present as non-mass lesions (28). Concurrently, benign lesions that present as non-mass enhancement, such as fibrosis and scar tissue, might present with lower ADC values due to a low water content (14). Similar results were obtained in other studies (30, 34), suggesting that the measurement of ADC should be interpreted with more caution for non-mass lesions.
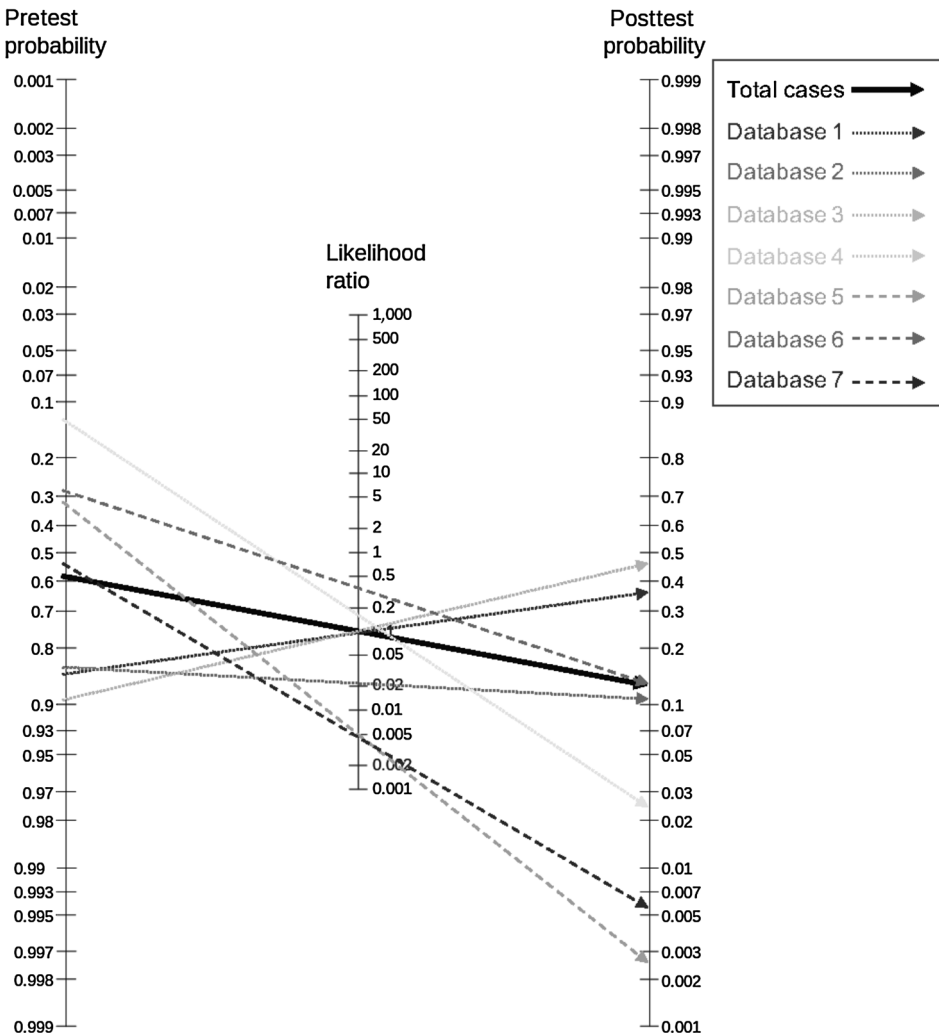

Figure 3.
Fagan nomogram showing the pretest and posttest probabilities of malignancy overall and separately in the included datasets from the different centers.

Downgrade of BI-RADS 4 lesions

Our analysis was done using a diverse dataset, with ADC values measured from different readers and acquired with different systems, sequences, and b-values. We obtained an accuracy comparable with that of smaller studies with more homogenous datasets (16, 18, 20, 35). Because of the chosen standard of reference, the dataset had a relatively high cancer prevalence, mathematically decreasing the prevalence-dependent NPV. Following Bayes' theorem, our data indicate that ADC measurements could be safely used to downgrade lesions for which the probability of malignancy is below 20%. ADC measurements could therefore be a safe, and thus, a valuable tool to eliminate the need for biopsies in BI-RADS 4 lesions (36).

In clinical practice, malignancy rates in BI-RADS 4 lesions vary considerably (37, 38), and decision making must consider individual patient factors such as the indication for MRI and imaging findings (39). Consequently, we used the calculated likelihood ratios to determine up to which pretest probability a negative DWI test result (e.g., an ADC value $\geq 1.5 \times 10^{-3}$ mm^2/second) would yield a posttest probability $\leq 2\%$, thereby allowing a formal downgrade of BI-RADS 4 lesions fulfilling these criteria. The feasibility of distinguishing BI-RADS a, b, and c scores has been demonstrated earlier (36, 40) and may be formalized by objective scoring systems to compensate for reader experience (41).

Limitations and conclusions

Our study has some limitations. The included datasets represent a heterogenous prevalence of cancer, as CE-MRI of the breast was performed for different indications depending on the routine of each institution. This, however, allowed us to evaluate simultaneously the performance of different readers and different systems and sequences, and therefore, reflects a very realistic, real-world scenario. The study design was retrospective. Only five of the data subsets provided information on DWI technical success: the rate of nondiagnostic DWI examinations ranged between 7.5% and 17.6% (Supplementary Materials and Methods S1) and is somewhat lower as compared with Rahbar and colleagues (18). As only biopsied lesions were included, the overall disease prevalence was high, close to 60%. Expectedly, cancer prevalence and thus BI-RADS 4 PPV's are higher in nonscreening settings. Still, a clinical selection bias toward more complicated cases is possible. This bias can, however, not be referred to DWI as none of the centers used DWI data for clinical decision making at the time of data acquisition. Specificity and NPV are underestimated in populations with a high disease prevalence (39), and this was also true in our analysis. Our results suggest that, if the probability of malignancy in a given BI-RADS 4 is not too high, the additional quantitative information of ADC could significantly improve diagnostic accuracy and reduce false-positive findings.

In conclusion, a cutoff of $\geq 1.5 \times 10^{-3}$ mm^2/second, as proposed by a prospective multicentric study (18), showed a high sensitivity and could have reduced the rate of unnecessary breast biopsies by 32.6%. The accuracy of ADC was only moderately influenced by lesion size, and it was reduced in non-mass lesions and for DCIS. A single ADC cutoff could be used to downgrade lesions initially classified as BI-RADS 4 on CE-MRI, a result not significantly influenced by MRI vendor, DWI acquisition parameters, or ROI method applied.

Authors' Disclosures

Authors' Contributions

P. Clauser: Data curation, formal analysis, investigation, methodology, writing–original draft. B. Krug: Data curation, formal analysis, investigation, writing–review and editing. H. Bickel: Conceptualization, data curation, investigation, formal analysis, methodology, writing–review and editing. M. Dietzel: Data curation, supervision, writing–review and editing. K. Pinker: Data curation, supervision, validation, investigation, writing–review and editing. V.-F. Neuhaus: Data curation, investigation, writing–review and editing. M.A. Marino: Data curation, formal analysis, investigation, writing–review and editing. M. Moschetta: Data curation, formal analysis, investigation, writing–review and editing. N. Troiano: Data curation, formal analysis, investigation, writing–review and editing. T.H. Helbich: Supervision, validation, writing–review and editing. P.A.T. Baltzer: Conceptualization, investigation, data curation, project administration, software, formal analysis, supervision, methodology, writing–original draft, writing–review and editing.

References

1. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI screening for women with extremely dense breast tissue. N Engl J Med 2019;381:2091–102.
2. Comstock CE, Gatsonis C, Newstead GM, Snyder BS, Gareen IF, Bergin JT, et al. Comparison of abbreviated breast MRI vs digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening. JAMA 2020;323:746–56.
3. Amitai Y, Scaranelo A, Menes TS, Fleming R, Kulkarni S, Ghai S, et al. Can breast MRI accurately exclude malignancy in mammographic architectural distortion? Eur Radiol 2020;30:2751–60.
4. Niell BL, Bhatt K, Dang P, Humphrey K. Utility of breast MRI for further evaluation of equivocal findings on digital breast tomosynthesis. AJR Am J Roentgenol 2018;211:1171–8.
5. 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.
6. D'Orsi Carl J, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. 5th ed. Reston, VA: American College of Radiology; 2013.

7. Mann RM, Balleyguier C, Baltzer PA, Bick U, Colin C, Cornford E, et al. Breast MRI: EUSOBI recommendations for women's information. Eur Radiol 2015;25:3669–78.

8. Sardanelli F. Overview of the role of pre-operative breast MRI in the absence of evidence on patient outcomes. Breast 2010;19:3–6.

9. Baltzer PAT, 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.

10. 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.

11. Mann RM, Cho N, Moy L. Breast MRI: state of the art. Radiology 2019;292: 520–36.

12. Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion mr imaging can tell us about tissue structure. Radiology 2013;268:318–22.

13. Bogner W, Gruber S, Pinker K, Grabner G, Stadlbauer A, Weber M, et al. Diffusion-weighted MR for differentiation of breast lesions at 3.0 T: how does selection of diffusion protocols affect diagnosis? Radiology 2009;253: 341–51.

14. Baltzer P, Mann RM, Iima M, Sigmund EE, Clauser P, Gilbert FJ, et al. 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.

15. Woodhams R, Matsunaga K, Iwabuchi K, Kan S, Hata H, Kuranami M, et al. Diffusion-weighted imaging of malignant breast tumors: the usefulness of apparent diffusion coefficient (ADC) value and ADC map for the detection of malignant breast tumors and evaluation of cancer extension. J Comput Assist Tomogr 2005;29:644–9.

16. Partridge SC, Rahbar H, Murthy R, Chai X, Kurland BF, DeMartini WB, et al. Improved diagnostic accuracy of breast MRI through combined apparent diffusion coefficients and dynamic contrast-enhanced kinetics. Magn Reson Med 2011;65:1759–67.

17. Partridge SC, DeMartini WB, Kurland BF, Eby PR, White SW, Lehman CD. Quantitative diffusion-weighted imaging as an adjunct to conventional breast MRI for improved positive predictive value. AJR Am J Roentgenol 2009;193: 1716–22.

18. Rahbar H, Zhang Z, Chenevert TL, Romanoff J, Kitsch AE, Hanna LG, et al. 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.

19. Spick C, Pinker-Domenig K, Rudas M, Helbich TH, Baltzer PA. MRI-only lesions: application of diffusion-weighted imaging obviates unnecessary MR-guided breast biopsies. Eur Radiol 2014;24:1204–10.

20. 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.

21. Sardanelli F, Boetes C, Borisch B, Decker T, Federico M, Gilbert FJ, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. Eur J Cancer 2010;46:1296–316.

22. Wielema M, Dorrius MD, Pijnappel RM, De Bock GH, Baltzer PAT, Oudkerk M, et al. Diagnostic performance of breast tumor tissue selection in diffusion weighted imaging: a systematic review and meta-analysis. PLoS One 2020;15: e0232856.

23. Clauser P, Marcon M, Maieron M, Zuiani C, Bazzocchi M, Baltzer PAT. Is there a systematic bias of apparent diffusion coefficient (ADC) measurements of the breast if measured on different workstations? An inter- and intra-reader agreement study. Eur Radiol 2016;26:2291–6.

24. McGee S. Simplifying likelihood ratios. J Gen Intern Med 2002;17:646–9.

25. Dorrius MD, Dijkstra H, Oudkerk M, Sijens PE. Effect of b value and pre-admission of contrast on diagnostic accuracy of 1.5-T breast DWI: a systematic review and meta-analysis. Eur Radiol 2014;24:2835–47.

26. Iima M, Le Bihan D, Okumura R, Okada T, Fujimoto K, Kanao S, et al. Apparent diffusion coefficient as an MR imaging biomarker of low-risk ductal carcinoma in situ: a pilot study. Radiology 2011;260:364–72.

27. Bickel H, Pinker-Domenig K, Bogner W, Spick C, Bagó-Horváth Z, Weber M, et al. Quantitative apparent diffusion coefficient as a noninvasive imaging biomarker for the differentiation of invasive breast cancer and ductal carcinoma in situ. Invest Radiol 2015;50:95–100.

28. Baltzer PAT, Bickel H, Spick C, Wengert G, Woitek R, Kapetas P, et al. Potential of noncontrast magnetic resonance imaging with diffusion-weighted imaging in characterization of breast lesions: intraindividual comparison with dynamic contrast-enhanced magnetic resonance imaging. Invest Radiol 2018;53:229–35.

29. Woodhams R, Kakita S, Hata H, Iwabuchi K, Umeoka S, Mountford CE, et al. Diffusion-weighted imaging of mucinous carcinoma of the breast: evaluation of apparent diffusion coefficient and signal intensity in correlation with histologic findings. AJR Am J Roentgenol 2009;193:260–6.

30. Partridge SC, Mullins CD, Kurland BF, Allain MD, DeMartini WB, Eby PR, et al. Apparent diffusion coefficient values for discriminating benign and malignant breast lesions: effects of lesion type and size. AJR Am J Roentgenol 2010;194: 1664–73.

31. Wan CWS, Lee CY, Lui CY, Fong CY, Lau KCH. Apparent diffusion coefficient in differentiation between malignant and benign breast masses: does size matter? Clin Radiol 2016;71:170–7.

32. 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.

33. 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.

34. Avendano D, Marino MA, Leithner D, Thakur S, Bernard-Davila B, Martinez DF, et al. 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.

35. Pinker K, Bickel H, Helbich TH, Gruber S, Dubsky P, Pluschnig U, et al. 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.

36. Strigel RM, Burnside ES, Elezaby M, Fowler AM, Kelcz F, Salkowski LR, et al. Utility of BI-RADS assessment category 4 subdivisions for screening breast MRI. AJR Am J Roentgenol 2017;208:1392–9.

37. Lee CI, Ichikawa L, Rochelle MC, Kerlikowske K, Miglioretti DL, Sprague BL, et al. Breast MRI BI-RADS assessments and abnormal interpretation rates by clinical indication in US community practices. Acad Radiol 2014;21:1370–6.

38. Spick C, Szolar DHM, Preidler KW, Reittner P, Rauch K, Brader P, et al. 3 Tesla breast MR imaging as a problem-solving tool: diagnostic performance and incidental lesions. PLoS One 2018;13:e0190287.

39. Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol 2009;62: 5–12.

40. Honda M, Kataoka M, Kawaguchi K, Iima M, Miyake KK, Kishimoto AO, et al. Subcategory classifications of Breast Imaging and Data System (BI-RADS) category 4 lesions on MRI. Jpn J Radiol 2021;39:56–65.

41. Marino MA, Clauser P, Woitek R, Wengert GJ, Kapetas P, Bernathova M, et al. A simple scoring system for breast MRI interpretation: does it compensate for reader experience? Eur Radiol 2016;26:2529–37.