Breast Lesions: Diagnosis by Using Proton MR Spectroscopy at 1.5 and 3.0 T—Systematic Review and Meta-Analysis[1]

Pascal A. T. Baltzer, MD
Matthias Dietzel, MD

Purpose: To perform a systematic review and meta-analysis to estimate the diagnostic performance of breast proton magnetic resonance (MR) spectroscopy in differentiating benign from malignant lesions and to identify variables that influence the accuracy of MR spectroscopy.

Materials and Methods: A comprehensive search of the PubMed database was performed on articles listed until January 6, 2012. The Medical Subject Headings and text words for the terms "breast," "spectroscopy," and "magnetic resonance" were used. Investigations including more than 10 patients at 1.5 T or 3.0 T applying one-dimensional single-voxel MR spectroscopy or spatially resolved MR spectroscopy for differentiation between benign and malignant breast lesions were eligible. A reference standard had to be established either by means of histopathologic examination or imaging follow-up of 12 or more months. Statistical analysis included pooling of diagnostic accuracy, control for data inhomogeneity, and identification of publication bias.

Results: Nineteen studies were used for general data pooling. The studies included a total of 1183 patients and 1198 lesions (773 malignant, 452 benign). Pooled sensitivity and specificity were 73% (556 of 761; 95% confidence interval [CI]: 64%, 82%) and 88% (386 of 439; 95% CI: 85%, 91%), respectively. The pooled diagnostic odds ratio (DOR) was 34.30 (95% CI: 16.71, 70.43). For breast cancers versus benign lesions, the area under the symmetric summary receiver operating characteristic curve of MR spectroscopy was 0.88, and the Q* index was 0.81. There was evidence of between-studies heterogeneity regarding sensitivity and DOR ($P < .0001$). No significant influences of higher field strength, postcontrast acquisition, or qualitative versus quantitative MR spectroscopy measurements were identified. Egger testing confirmed significant publication bias in studies including small numbers of patients ($P < .0001$).

Conclusion: Breast MR spectroscopy shows variable sensitivity and high specificity in the diagnosis of breast lesions, independent from the technical MR spectroscopy approach. Because of significant publication bias, pooled diagnostic measures might be overestimated.

[1] From the Department of Radiology, Medical University Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria (P.A.T.B.); and Department of Neuroradiology, University Hospital Erlangen, Erlangen, Germany (M.D.). Received August 28, 2012; revision requested October 4; revision received November 2; accepted November 12; final version accepted December 5. **Address correspondence to** P.A.T.B. (e-mail: *pascal.baltzer@meduniwien.ac.at*).

Dynamic contrast material–enhanced magnetic resonance (MR) imaging is the most sensitive method for detection of breast cancer (1,2). The high detection rate of this method for breast cancer is based on T1-weighted studies that allow the measurement of the extracellular distribution of paramagnetic contrast agents. Although cancers show a characteristically early and strong enhancement with rapid washout, substantial overlap of enhancement characteristics between benign and malignant breast lesions has been described (3,4). Consequently, for lesion classification in clinical practice, a combination of morphologic criteria and dynamic enhancement pattern analysis is applied (5). Morphology assessment is a subjective task and thus prone to experience-related variation and interobserver bias. Nonmass and small lesions especially frequently cause false-positive findings owing to a limited diagnostic performance of established criteria used in dynamic contrast-enhanced MR imaging (6–8). An adjunct method providing high specificity would thus be of diagnostic value.

Advances in Knowledge

- In a meta-analysis of 19 studies of the diagnostic performance of proton MR spectroscopy for the differentiation of malignant from benign breast lesions, the pooled overall sensitivity and specificity were 73% (556 of 761; 95% confidence interval [CI]: 64%, 82%) and 88% (386 of 439; 95% CI: 85%, 91%), respectively.
- According to our results, the diagnostic performance of MR spectroscopy was not significantly influenced by field strength, spectroscopy sequence, method of spectra analysis, study design, or quality criteria.
- We identified significant publication bias: studies including low numbers of patients reported systematically higher sensitivities for MR spectroscopy ($P < .0001$).

Proton MR spectroscopy (specifically, hydrogen 1 MR spectroscopy) is a noninvasive examination technique for the assessment of biochemical tissue properties. The presence of a compound resonance at around 3.23 ppm is attributed to choline metabolites such as choline, phosphocholine, and glycerophosphocholine and is simply referred to as total choline (tCho). Increased levels of tCho have been detected in malignant cancers and are ascribed to an increased cellular membrane turnover (9–11). In vivo qualitative and quantitative tCho measurements have been used as a diagnostic test in the work-up of neoplastic breast lesions (12–32).

However, the clinical value of MR spectroscopy of the breast still remains unclear and is controversial. This is because of many factors, and at this point two should be discussed. First, the number of studies investigating MR spectroscopy of the breast in a clinical setting is rather low. This substantially limits the statistical power of the data published to date. Second, study designs in the present literature are heterogeneous, in terms of both technical criteria and the characteristics of the patients studied. Variations in patient characteristics and spectroscopic methodology have been described as confounders of spectroscopic results (11,33). Accordingly, integrating such data into clinical practice is challenging. To solve this task, there is a need for systematic control of both patient characteristics and technical specifications.

Accordingly, we performed a systematic review and meta-analysis to investigate the diagnostic performance of tCho measurements for the differentiation of breast lesions, focusing on the state-of-the-art field strengths for clinical breast imaging of 1.5 T and 3.0 T.

Implication for Patient Care

- Owing to its high specificity, MR spectroscopy may be helpful for the diagnosis of breast lesions; however, owing to its lower and variable reported sensitivity, further systematic research is necessary to verify the diagnostic value of clinically applied MR spectroscopy.

Materials and Methods

No financial support was received for this research.

Search Strategy

A computerized search was performed by using the free-access PubMed database *(www.ncbi.nlm.nih.gov/pubmed/)*, including articles listed until January 6, 2012. The following Medical Subject Headings and text search terms were used: "breast," "magnetic resonance imaging," "magnetic resonance spectroscopy," "mri," "mr," "spectrum analysis," and "spectroscopy and 1h."

Eligibility Criteria for Study Selection

Eligibility criteria for study selection were as follows: peer-reviewed studies on human subjects applying one-dimensional single-voxel spectroscopy or spatially resolved multivoxel spectroscopic imaging for differentiation of benign from malignant breast lesions. Furthermore, the applied field strength had to be 1.5 or 3.0 T to represent current technical standards

Published online before print
10.1148/radiol.13121856 **Content code:** BR

Radiology 2013; 267:735–746

Abbreviations:
AUC = area under the ROC curve
CI = confidence interval
DOR = diagnostic odds ratio
FWHM = full width at half maximum
QUADAS = Quality Assessment of Diagnostic Accuracy Studies
ROC = receiver operating characteristic
tCho = total choline

Author contributions:
Guarantor of integrity of entire study, P.A.T.B.; study concepts/study design or data acquisition or data analysis/interpretation, P.A.T.B., M.D.; manuscript drafting or manuscript revision for important intellectual content, P.A.T.B., M.D.; manuscript final version approval, P.A.T.B., M.D.; literature research, P.A.T.B., M.D.; clinical studies, P.A.T.B.; statistical analysis, P.A.T.B., M.D.; and manuscript editing, P.A.T.B., M.D.

Conflicts of interest are listed at the end of this article.

in clinical breast imaging. A reference standard had to be established, either by means of histopathologic sampling or by means of imaging follow-up of at least 12 months. Not eligible were studies with fewer than 10 patients, studies investigating only malignant lesions, and studies comparing malignant lesions with benign breast parenchyma. No further restrictions were used. Titles and abstracts of search results were reviewed by two independent observers (P.A.T.B., with 10 years of experience in breast MR imaging and 6 years of experience in breast MR spectroscopy, and M.D., with 8 years of experience in breast MR imaging). A study was included if diagnostic data could be summarized in a 2 × 2 contingency table to assess true-positive, true-negative, false-positive, and false-negative findings. If an overlap between studies was identified, the more recent report was chosen to avoid data redundancy.

Data Collection and Quality Assessment

Data collection included the following parameters: publication year, study design (retrospective vs prospective), number of patients, age, number of benign and malignant lesions, lesion size, applied field strength, voxel size and spectroscopic technique, and whether spectroscopy was performed before or after contrast medium injection. Furthermore, data on how spectra were analyzed and the number and experience of observers were collected. Overall numbers of true-positive, true-negative, false-positive, and false-negative findings were extracted, and, if available, were stratified according to mass and nonmass subgroups. Study quality was assessed by both independent observers by using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool for scoring studies. This tool provides a checklist of items regarding the representativeness and methodologic quality of investigated studies. Positive scores of 14 items are added up and can vary from 0 to 14 (34). If present, disagreement was solved by a consensus rereading of unclear points.

Statistical Analysis

All analyses were performed by using Meta-DiSc (35) and Stata, version 11.0 (Stata, College Station, Tex). Spectroscopic classification results were tabulated against the reference standard by using 2 × 2 contingency tables. These raw data were further analyzed as described below.

Control for data inhomogeneity.—A random-effects model as proposed by DerSimonian and Laird was applied to control for differences in reported data (eg, patient characteristics and methods used). It represents a classic, noniterative method to account for interstudy heterogeneity. χ^2 and I^2 statistics were computed. I^2 values were interpreted according to the proposal of Higgins and Thompson (36) as showing low ($I^2 \leq 25\%$), medium ($I^2 \leq 50\%$), or high ($I^2 \leq 75\%$) heterogeneity.

Pooled diagnostic accuracy.—Threshold analysis was implemented to assess whether the diagnostic odds ratio (DOR) was constant. A symmetric summary receiver operating characteristic (ROC) curve was fitted by using the Moses constant of linear model (weighted regression, inverse variance). Measures for the analysis of summary ROC included the area under the ROC curve (AUC) and the Q* index. Being invariant to heterogeneity, Q* is defined as the limit case, where specificity equals sensitivity (37).

Influence factors on diagnostic accuracy were assessed by means of formal meta-regression analysis (least squares weighted by inverse variance) (38). The parameters listed in the Data Collection and Quality Assessment section were used as covariates. $P < .05$ was considered to indicate a significant difference.

"Publication bias" describes a discrepancy between what is likely to be published among available results. Studies showing significant results have a higher probability of being published compared with studies showing little or nonsignificant effects. Publication bias was assessed by using a funnel plot with each study's log DOR plotted against its standard error of the estimate. Quantitative analysis for possible publication bias was performed by using the methods proposed by Begg and Mazumdar (39) and Egger et al (40), with $P < .05$ regarded as indicative of significant publication bias. The trim and fill method proposed by Duval and Tweedie (41) was used for exploratory bias correction.

Results

Study Design

Twenty eligible studies were identified (12–31). A flowchart summarizing the selection process of the finally included studies is shown in Figure 1. Study design was described as prospective in 19 studies and retrospective in one (30) study. In one study (27), the retrospective or prospective character of the study could not be identified. It should be mentioned that spectroscopic measurements in general have to be planned and acquired prospectively. "Retrospective" in this context can only refer to a later time point of spectra analysis or patient subgroup selection. Patient recruitment was consecutive in 10 studies (13–16,18,19,25,28–30). Six reports described nonconsecutive (case-control) patient recruitment (12,17,20,21,24,31), and in another four studies, patient recruitment was

Figure 1

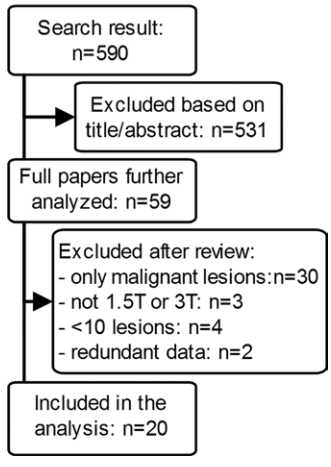

Figure 1: Flowchart summarizes the selection process toward the final group of studies analyzed. Of 20 included studies, 19 were used for general data pooling, and one was used only for nonmass subgroup analysis.

Key Parameters Extracted from the Investigated Studies

Study	Consecutive Recruitment?	QUADAS Score	No. of Patients	No. of Malignant Lesions	No. of Benign Lesions	MR Spectroscopy Sequence	B0 Value	Repetition Time (msec)	Echo Time(s) (msec)	Spectroscopy after CM Administration?	Fat Saturation Used?	Method of Choline Analysis and Cutoff Value	FWHM(Hz)	Minimum Lesion Size (mm)
Roebuck et al 1998 (26)	NA	11	17	10	7	Single-voxel STEAM	1.5	2000	31, 270	Yes	No	Visual	5.6–18	None
Kvistad et al 1999 (24)	No	11	22	11	11	Single-voxel PRESS	1.5	2000	135, 350, 450	No	No	Visual	12–14	10
Cecil et al 2001 (17)	No	12	38	23	15	Single-voxel STEAM	1.5	2000	270	Mixed	No	Visual	NA	10
Tse et al 2003 (31)	No	12	46	19	27	Single-voxel PRESS	1.5	2000	38, 135, 270	Yes	No	SNR, 2	NA	15
Kim et al 2003 (22)	NA	11	35	19	16	Single-voxel PRESS	1.5	2500	288	No	Yes	SNR	3–8	NA
Huang et al 2004 (20)	No	11	30	18	12	Single-voxel PRESS	1.5	2000	135	Yes	No	SNR, 2	NA	NA
Jacobs et al 2004 (21)	No	11	15	8	7	Chemical shift imaging	1.5	2000	272	Yes	Yes	SNR, 4	NA	NA
Bartella et al 2006 (15)	Yes	12	56	31	26	Single-voxel PRESS	1.5	2000	135	Yes	No	SNR, 2	<20	10
Lu et al 2006 (25)	Yes	11	131	74	57	Single-voxel PRESS	1.5	1500	110	Yes	No	SNR, 3	NA	NA
Bartella et al 2007 (16)*	Yes	12	32	12	20	Single-voxel PRESS	1.5	2000	135	Yes	No	SNR, 2	<20	10
Baek et al 2008 (13)	Yes	11	36	27	9	Chemical shift imaging	1.5	1627	270	Yes	Yes	SNR, 3.2	10–29	10
Sardanelli et al 2009 (28)	Yes	11	42	19	26	Single-voxel PRESS	1.5	1500	136	Yes	Yes	Peak area/voxel	<30	10
Tozaki and Fukuma 2009 (30)	Yes	11	165	91	80	Single-voxel PRESS	1.5	1620	270	Yes	No	SNR, 2	10–20	NA
Baltzer et al 2012 (14)	Yes	13	62	42	21	Single-voxel PRESS	1.5	2000	272	Yes*	No	SNR, 2	15.7	8
Thakur et al 2011 (29)	Yes	11	88	57	31	Single-voxel PRESS	1.5	2000	135	Yes	No	Internal reference, 0.1 mmol/kg	<20	10
Gruber et al 2011 (19)	Yes	11	50	31	19	Chemical shift imaging	3	750	145	No	Yes	SNR, 2.6	25–35	NA

Table 1 (continues)

(continued)

Key Parameters Extracted from the Investigated Studies

Study	Consecutive Recruitment?	QUADAS Score	No. of Patients	No. of Malignant Lesions	No. of Benign Lesions	MR Spectroscopy Sequence	B0 Value	Repetition Time (msec)	Echo Time(s) (msec)	Spectroscopy after CM Administration?	Fat Saturation Used?	Method of Choline Analysis and Cutoff Value	FWHM(Hz)	Minimum Lesion Size (mm)
Dorrius et al 2012 (18)	Yes	11	25	15	11	Chemical shift imaging	1.5	1500	135	No	Yes	Internal reference, 1.5 mmol/kg	6–9	10
Baek 2012 (12)	No	11	112	99	13	Single-voxel PRESS	1.5	2000	270	Yes	Yes	SNR, 2	8–17	10
Kousi et al 2012 (23)	NA	11	24	14	13	Single-voxel PRESS	3	2000	155	Yes	Yes	Visual	<25	NA
Sah et al 2012 (27)	NA	8	189	151	38	Single-voxel PRESS	1.5	1500	100	Yes	Yes	Internal reference, 2.54 mmol/kg	10–20	NA

Note.—CM = contrast medium, FWHM = full width at half maximum, NA = not available, PRESS = point-resolved spatially localized spectroscopy, SNR = signal-to-noise ratio, STEAM = stimulated-echo acquisition mode.
* Used only for nonmass subgroup analysis.

not clearly described in the text (22,23,26,27) (Table). A minimum lesion size as an inclusion criterion was defined in 11 studies (12–18,24,28,29,31) and ranged between 8 and 15 mm, with a median of 10 mm. Inclusion criteria were suspicious or unclear findings at conventional imaging in 16 studies (12–14,17–28,31) and suspicious findings at MR imaging in four investigations (15,16,29,30).

Spectroscopy: Acquisition Technique and Spectra Analysis

In the majority (n = 18) of the eligible studies, spectroscopic data were acquired with 1.5-T magnets, and only two studies (19,23) examined MR spectroscopy at 3.0 T. Single-voxel spectroscopy was used in 16 studies, with 14 of these applying the point-resolved spatially localized spectroscopy (PRESS) sequence (12,14–16,20,22–25,27–31) and two applying the stimulated-echo acquisition mode technique (17,26) (Table). Spectroscopic imaging with the PRESS technique was applied in four studies (13,18,19,21). Water saturation was applied in all studies, and fat saturation was used in nine studies (12,13,18,19,21–23,27,28). The tCho resonance was examined by means of subjective visual inspection in four studies (17,23,24,26), while 12 studies (12–16,19–22,25,30,31) used the signal-to-noise ratio of the tCho peak as compared with baseline noise to examine the tCho resonance, with typical thresholds greater than 2 (range, 2–4). A fully relaxed spectrum was used as an internal reference for tCho quantitation in three studies (18,27,29), with cutoff values for malignancy varying between 0.1 and 2.54 mmol/kg, while one study (28) used a normalization of the tCho peak integral by the size of the interrogated voxel. In four studies (18,19,22,24), spectra were acquired before contrast medium administration, while in one study (23), spectra were acquired twice—before and after contrast medium administration. Here we considered only postcontrast acquisitions, as sensitivity and specificity values were substantially higher. Another study (17) included a mix of pre- and

Figure 2

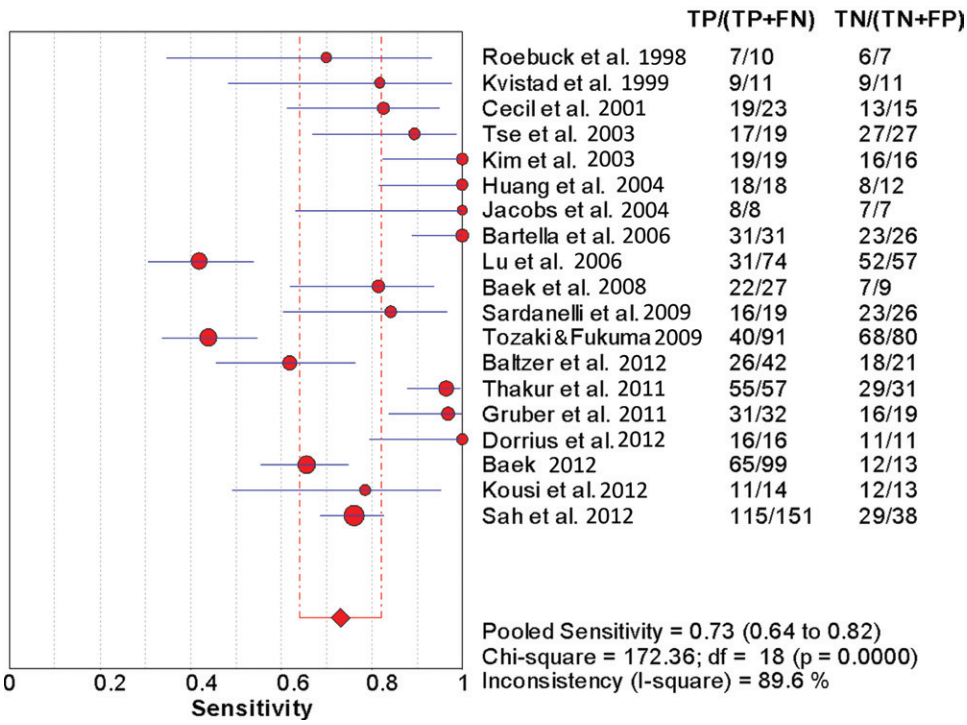

Figure 2: Forest plot of sensitivity of MR spectroscopy. Circles = individual study point estimates. Circle size, indicating relative individual contribution to data pooling, is proportional to 1/(within-study variance + between-study variance). Horizontal lines = 95% CIs. Rhomb and dashed vertical lines = pooled sensitivity and its 95% CI. *FN* = false-negative, *FP* = false-positive, *TN* = true-negative, *TP* = true-positive.

postcontrast spectra, and in 14 studies (12–16,20,21,25–31), only postcontrast spectroscopic data were evaluated.

Three studies (25,27,31) did not provide information on spectroscopic voxel size, while in the other studies, the mean voxel size at single-voxel MR spectroscopy ranged from 2.2 to 6.3 mL, with reported minimum and maximum values of between 0.82 and 25.2 mL. Results of B0 homogenization by shimming expressed by the FWHM of the water resonance at 4.7 ppm was reported in 15 studies (12–16,18,19,22–24,26–30) (Table).

Information on who planned spectroscopic voxel placement was given in three studies (14,21,29), while the number and experience of observers interpreting acquired spectra were provided by six reports (14–17,19,30). Spectra interpretation was blinded in five investigations (14–17,31), and no information regarding blinded reading was given in the remaining articles. The mean assigned QUADAS score was 11.1 (median, 11; range, 8–13) (Table).

Synthesis of Individual Studies: Demographic Data and Lesions

Of 20 studies, 19 were used for general data pooling, as one study (16) presented a subgroup analysis (of nonmass lesions) with overlap to another study (15). In these 19 included studies, a total of 1183 patients (range, 15–189) and 1198 lesions (range, 15–189), of which 773 were malignant (range, eight to 151) and 452 were benign (range, seven to 80), were included. Age distribution was heterogeneously reported; mean age was reported in 15 studies (12–14,16,18–23,25,26,28–30) and ranged between 44.8 and 60 years. Subgroup age demographics only were given in three studies (24,27,31), median age and range were given in one study (15), and raw data on age only were given in one study (17). The reported age range for all studies considered was between 14 and 92 years. Overall mean or median lesion sizes were described in five studies (14,15,21,22,26) and ranged between 17 and 35 mm. Lesion sizes of benign and malignant subgroups were given in 10 studies (12,13,19,22,24–26,28,30,31) and ranged between 16.9 and 33 mm for malignant lesions and between 8 and 30.5 mm for benign lesions. Minimum described benign lesion size was 4 mm (30), and minimum malignant lesion size was 2 mm (19).

Synthesis of General Diagnostic Performance

Individual study results–weighted summaries of sensitivity, specificity, and DORs, together with their 95% confidence intervals (CIs) are provided in Figures 2–4. Pooled sensitivity, specificity,

Figure 3

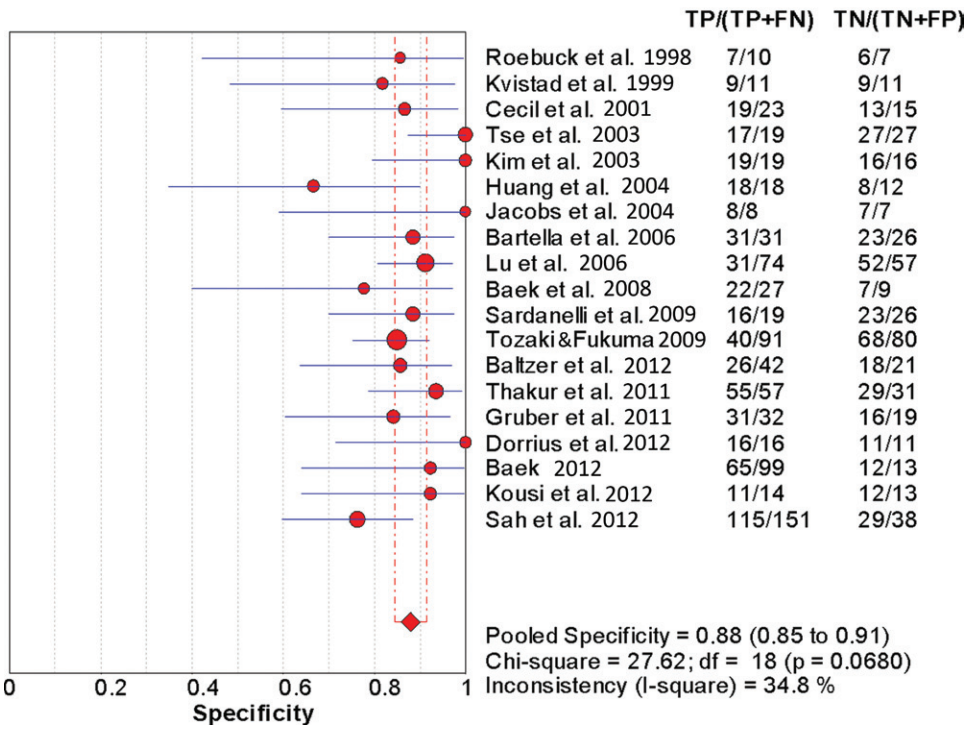

Figure 3: Forest plot of specificity of MR spectroscopy. Circles = individual study point estimates. Circle size, indicating relative individual contribution to data pooling, is proportional to 1/(within-study variance + between-study variance). Horizontal lines = 95% CIs. Rhomb and dashed vertical lines = pooled specificity and its 95% CI. *FN* = false-negative, *FP* = false-positive, *TN* = true-negative, *TP* = true-positive.

and DOR were 73% (556 of 761), 88% (386 of 439), and 34.3, respectively. There was strong evidence of between-study heterogeneity for sensitivity (I^2 = 89.6%, $P < .0001$) and DOR (I^2 = 65%, $P < .0001$).

A summary ROC curve with AUC and Q* index data is shown in Figure 5. The AUC and Q* index were 0.88 and 0.81, respectively.

Synthesis of Diagnostic Performance in Mass and Nonmass Subgroups

Five studies (13,14,17,19,30) included contingency table data on mass lesions and another six studies (13,14,16,17,19,30) included data on nonmass lesion subgroups. Of the latter, only five studies were used for data pooling, as the nonmass lesion subgroup in one study (17) consisted of benign lesions only. Pooled estimates for sensitivity and specificity (Fig 6) in mass lesions were 68% (115 of 170) and 88% (87 of 99), respectively. There was strong evidence of between-study heterogeneity for sensitivity (I^2 = 84.5%, $P < .0001$).

Figure 7 shows forest plots that include pooled sensitivity and specificity in nonmass lesions, revealing pooled sensitivity of 62% (34 of 55) and pooled specificity of 79% (50 of 63). There was strong evidence of between-study heterogeneity for sensitivity (I^2 = 88.3%, $P < .0001$). As shown in Figures 6 and 7, sensitivity and especially specificity seemed to be lower in nonmass lesions.

Factors Influencing the Diagnostic Performance of MR Spectroscopy

Meta-regression analysis identified number of patients investigated as the only significant predictor of diagnostic performance (coefficient = −0.01; standard error of the estimate = 0.0032; P = .0058). The (confounder-corrected) correlation coefficient was 0.587 between sensitivity and number of patients (P = .017). No correlation between number of patients and specificity was observed. Studies in which MR spectroscopy was performed before contrast medium application showed relatively high sensitivity, of between 82% (nine of 11) (24) and 100% (16 of 16) (18), without reaching statistical significance ($P > .05$). All other investigated covariates did not show a significant influence on the diagnostic performance of MR spectroscopy.

Assessment of Publication Bias

To address publication bias, a funnel plot of the log DOR against the standard error of the estimate of the log DOR was constructed (Fig 8). As can be seen in the funnel plot, studies of small sample size have a higher DOR than studies of a larger sample size. The Egger test confirmed the presence of publication bias ($P < .0001$). Use of the trim and fill method for bias correction revealed

Figure 4

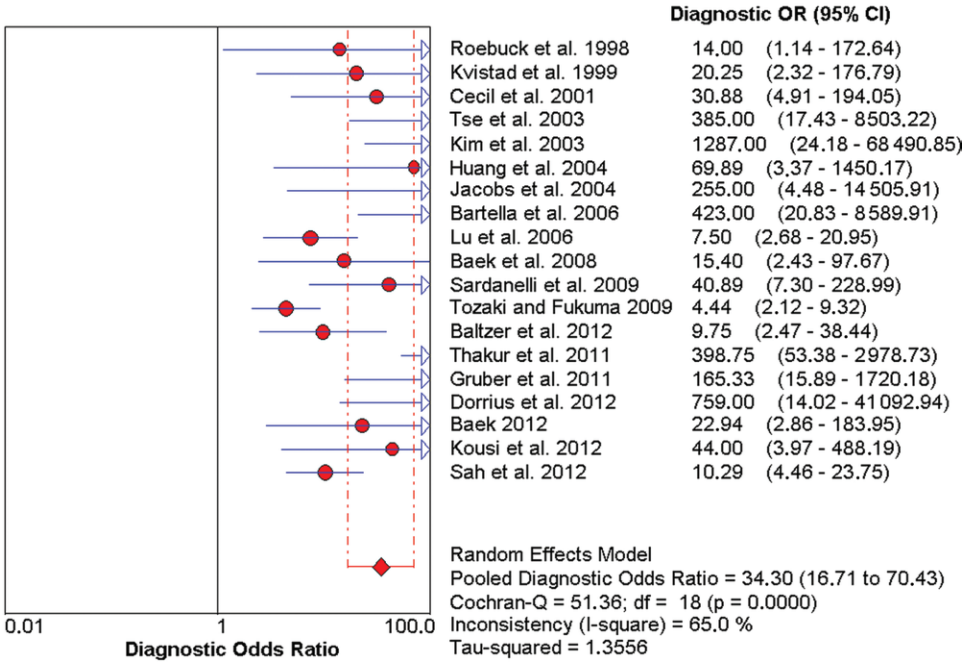

Figure 4: Forest plot of DOR of MR spectroscopy. Circles = individual study point estimates. Circle size, indicating relative individual contribution to data pooling, is proportional to 1/(within study variance + between study variance). Horizontal lines = 95% CIs. Rhomb and dashed vertical lines = pooled DOR and its 95% CI.

a log[DOR] of 2.445 (95% CI: 1.692, 3.199), corresponding to a bias-corrected pooled DOR estimate of 11.53 (95% CI: 5.43, 24.51).

Discussion

The present meta-analysis investigated the diagnostic performance of proton MR spectroscopy for differentiation between benign and malignant neoplastic lesions of the breast. Several quality-related issues were identified: Only 50% (10 of 20) of all studies recruited patients in a consecutive manner. Although the spectroscopic technique was described sufficiently in all articles, only 75% (15 of 20) reported assessment of basic spectroscopic quality criteria in terms of FWHM. Because peak height and FWHM are correlated, high FWHM corresponding to low B0 field homogeneity implies false-negative choline findings at spectroscopy. This is why FWHM details should be provided in any MR spectroscopy study. Furthermore, only

Figure 5

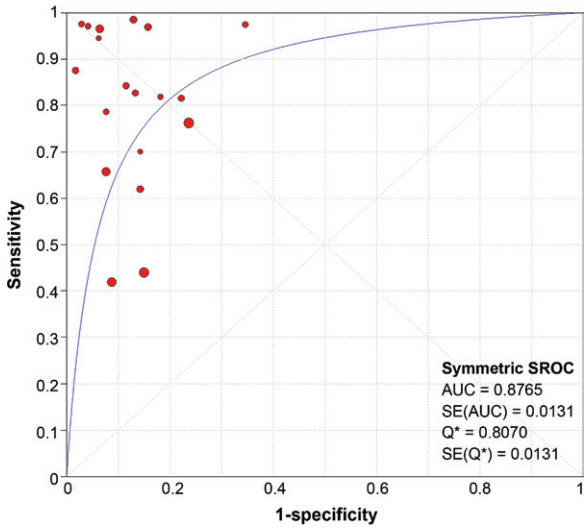

Figure 5: Summary ROC curve of individual-study MR spectroscopy results. *SE* = standard error of the estimate.

25% (five of 20) of all reports described a blinded analysis of spectroscopic data. Three (15%) of 20 studies reported on who planned MR spectroscopy, and six (30%) of 20 reports provided information on the persons analyzing the spectra. Reproducibility of spectroscopic voxel placement and spectra analysis

Figure 6

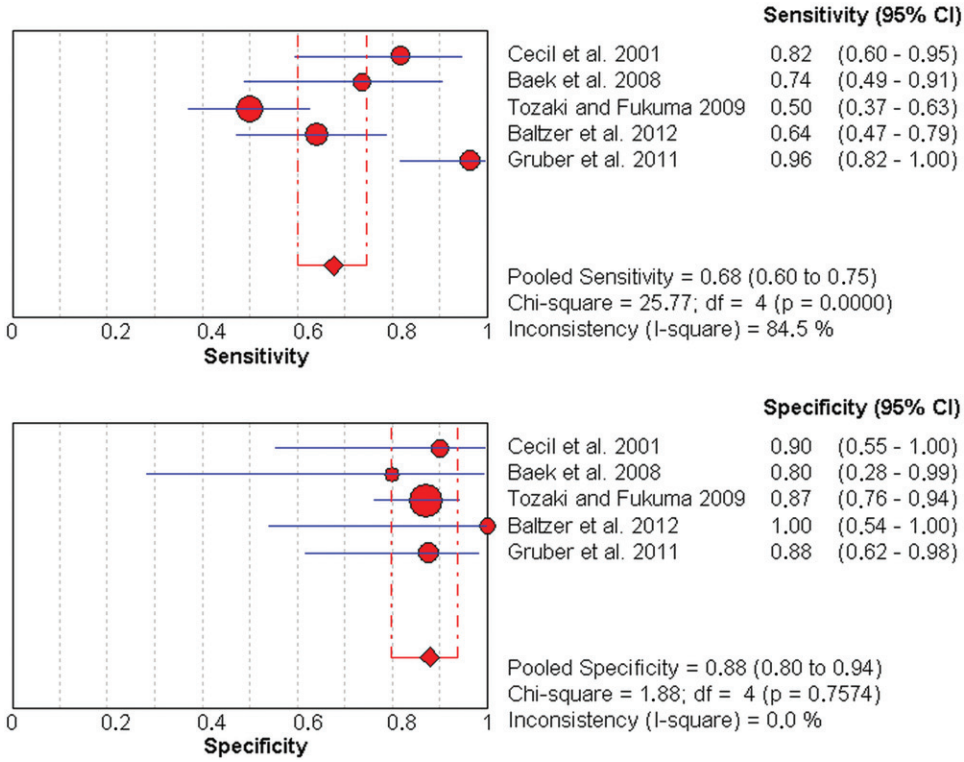

Figure 6: Forest plots of sensitivity and specificity of MR spectroscopy in mass lesions. Circles = individual study point estimates. Circle size, indicating relative individual contribution to data pooling, is proportional to 1/(within-study variance + between-study variance). Horizontal lines = 95% CIs. Rhomb and dashed vertical lines = pooled sensitivity and specificity and corresponding 95% CIs.

were not investigated in any study. The mean QUADAS score was rather high at 11.1, meaning that the investigated studies fulfilled most quality criteria. However, the issues listed above, first of all the missing statements regarding blinded spectra interpretation, have to be considered as substantial limitations.

Analysis of 1198 lesions demonstrated a consistent high pooled specificity of 88% (386 of 439) and a lower sensitivity of 73% (556 of 761). The latter showed substantial heterogeneity and varied between 42% (31 of 74) (25) and 100% (15,18,20–22). As threshold effects were not significant in our analysis, further reasons seem to influence heterogeneity. Meta-regression identified a negative correlation between study size and diagnostic performance (P = .0058). Publication bias toward higher diagnostic performance in terms of diagnostic odds ratio was found, especially in small studies. It has to be assumed that cancers investigated in small studies differed (eg, in terms of size, grade) from those in studies in larger numbers of subjects. However, this is not reflected in lesion-related data provided in the text. As no effects of technical parameters or study design on diagnostic performance of MR spectroscopy were identified, either retention of negative results or bias in malignant lesion recruitment has to be assumed. This is of interest, as it has previously been assumed that variations in methodology are a confounding factor for breast MR spectroscopy diagnostic performance (11). Using formal meta-analysis, we can reject the hypothesis of significant influences on MR spectroscopy diagnostic performance by the different methodologies used. The most important consequence of the presence of publication bias for the present meta-analysis is an overestimation of the diagnostic performance of MR spectroscopy. Bias correction by the trim and fill method showed a corrected DOR of 11.53, which was lower than the original estimate of 34.3. However, even after bias correction, breast MR spectroscopy shows significant discriminatory power as a diagnostic test.

A variety of spectroscopic techniques at different field strengths were used. Higher field strengths provide higher signal-to-noise ratios, and spectroscopic imaging allows spatially resolved examinations with small voxels. Although a diagnostic benefit may thus be expected, comparison with studies at 1.5 T did not reveal such an advantage yet. A diagnostic advantage of MR spectroscopy related to improvements in coil

Figure 7

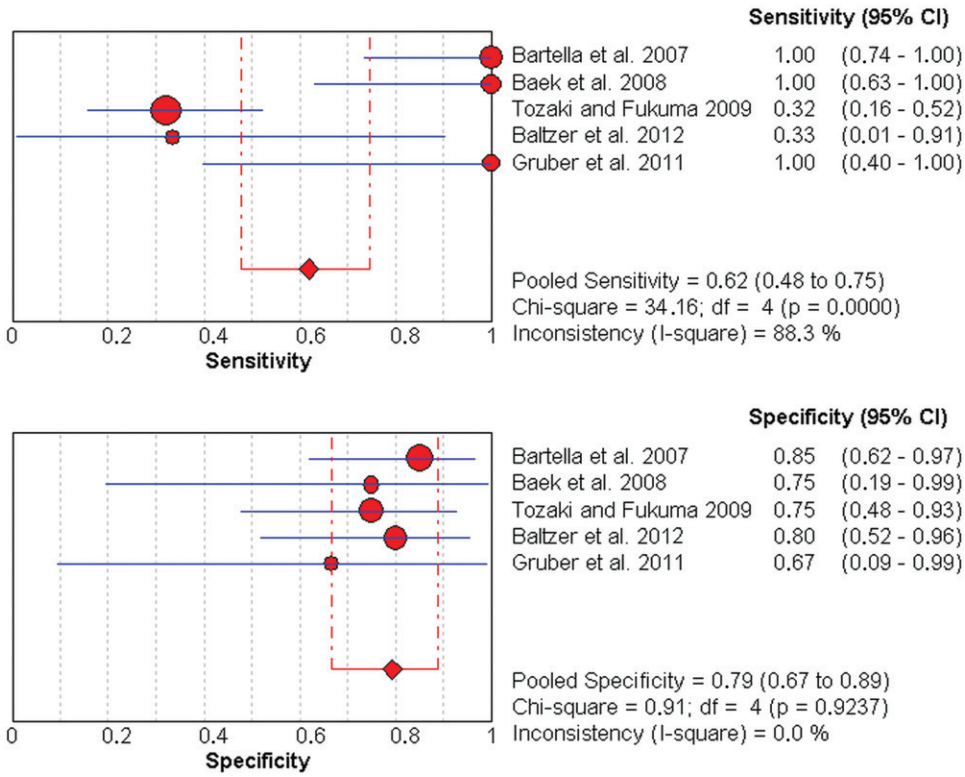

Figure 7: Forest plots of sensitivity and specificity of MR spectroscopy in nonmass lesions. Circles = individual study point estimates. Circle size, indicating relative individual contribution to data pooling, is proportional to 1/(within-study variance + between-study variance). Horizontal lines = 95% CIs. Rhomb and dashed vertical lines = pooled sensitivity and specificity and corresponding 95% CIs.

Figure 8

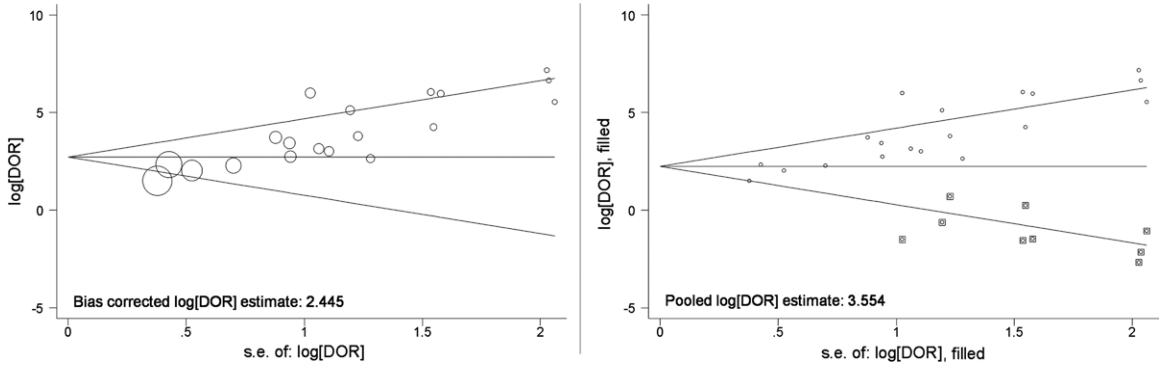

Figure 8: Left: Funnel plot of MR spectroscopy results with pseudo 95% CI. Right: Funnel plot filled according to the trim and fill method. Rectangles = filled hypothetical studies. *s.e.* = Standard error of the estimate.

architecture and MR imaging systems was not identified: Year of publication showed no influence on diagnostic performance. However, further study is warranted to clarify especially the role of multichannel technology, higher field strength, and spectroscopic imaging in breast MR spectroscopy, as published empirical data were found to be very limited. A detrimental effect of contrast agents on MR spectroscopy has been described in both experimental and

clinical settings (42–44). Although visual inspection of forest plots showed a relatively high sensitivity of precontrast MR spectroscopy, no statistical significance was reached, hinting at a relatively small effect regarding lesion differentiation at MR spectroscopy. However, altered tCho signal intensities after contrast medium injection have to be considered if absolute tCho quantification is performed.

Potentially, absolute tCho quantification could lead to a more standardized reading of breast spectra, enabling transfer of thresholds between institutions. The internal reference approach used by three investigated studies eliminates the influence of tumor-to-voxel ratio, voxel size, and voxel position (45). However, water content in breast tissue is variable and may thus bias quantification results (46,47). In our meta-analysis, quantitative approaches did not show higher accuracy compared with qualitative spectra inspection.

It should be kept in mind that MR spectroscopy has a generally low signal-to-noise ratio. Although large lesions were investigated, with 50% (10 of 20) of all studies omitting lesions smaller than 10 mm, sensitivity showed the variations described above. This fact limits the applicability of MR spectroscopy in the diagnosis of early breast cancer and generally small lesions. Although the pooled specificity of MR spectroscopy was found to be high and of little heterogeneity, the limited sensitivity of this method as identified in this meta-analysis may be detrimental for decreasing the number of false-positive findings at contrast-enhanced breast MR imaging by using MR spectroscopy, as suggested by some authors (15,18,20). False-positive findings at contrast enhanced breast MR imaging are commonly encountered in small and nonmass lesions (6,7). Subgroup meta-analysis showed a similar picture, demonstrating heterogeneous diagnostic performance of MR spectroscopy in nonmass lesions. In particular, specificity was lower as compared with that for mass lesions. Although breast cancer might be detected by means of spectroscopic imaging only, single-voxel spectroscopy cannot be used for lesion detection. Consequently, MR spectroscopy does depend on further MR imaging–based imaging techniques, further limiting its use in the diagnostic setting.

In conclusion, the present meta-analysis shows high specificity (88% [386 of 439]) and a lower, very variable sensitivity (73% [556 of 761]) of breast MR spectroscopy in a lesion-based differentiation task, independent from variations in methodology. A diagnostic advantage of 3.0 T as compared with 1.5 T, precontrast MR spectroscopy as compared with postcontrast MR spectroscopy, or quantitative MR spectroscopy as compared with qualitative MR spectroscopy could not be identified. Publication bias toward higher diagnostic performance was identified, hinting at a possible overestimation of pooled diagnostic parameters. Reporting of MR spectroscopy studies could be improved regarding study design and patient recruitment, as well as spectra acquisition and reading conditions. Data on reliability were insufficient. Standardized prospective multicenter trials providing patient-based comparisons with standard imaging procedures are warranted to clarify the use of MR spectroscopy for differential diagnosis of breast lesions.

Disclosures of Conflicts of Interest: P.A.T.B. No relevant conflicts of interest to disclose. **M.D.** No relevant conflicts of interest to disclose.

References

1. Houssami N, Ciatto S, Macaskill P, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. J Clin Oncol 2008;26(19):3248–3258.
2. Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med 2008;148(9):671–679.
3. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology 1999;211(1):101–110.
4. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. Radiology 2006;238(1):42–53.
5. Ikeda DM, Hylton NM, Kinkel K, et al. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. J Magn Reson Imaging 2001;13(6):889–895.
6. 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(6):1658–1663.
7. 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(4):994–1000.
8. Liberman L, Mason G, Morris EA, Dershaw DD. Does size matter? Positive predictive value of MRI-detected breast lesions as a function of lesion size. AJR Am J Roentgenol 2006;186(2):426–430.
9. Begley JK, Redpath TW, Bolan PJ, Gilbert FJ. In vivo proton magnetic resonance spectroscopy of breast cancer: a review of the literature. Breast Cancer Res 2012;14(2):207.
10. Mountford C, Ramadan S, Stanwell P, Malycha P. Proton MRS of the breast in the clinical setting. NMR Biomed 2009;22(1):54–64.
11. Haddadin IS, McIntosh A, Meisamy S, et al. Metabolite quantification and high-field MRS in breast cancer. NMR Biomed 2009;22(1):65–76.
12. Baek HM. Diagnostic value of breast proton magnetic resonance spectroscopy at 1.5T in different histopathological types. ScientificWorldJournal 2012;2012:508295.
13. Baek HM, Chen JH, Yu HJ, Mehta R, Nalcioglu O, Su MY. Detection of choline signal in human breast lesions with chemical-shift imaging. J Magn Reson Imaging 2008;27(5):1114–1121.
14. Baltzer PA, Gussew A, Dietzel M, et al. Effect of contrast agent on the results of in vivo ^{1}H MRS of breast tumors: is it clinically significant? NMR Biomed 2012;25(1):67–74.
15. Bartella L, Morris EA, Dershaw DD, et al. Proton MR spectroscopy with choline peak as malignancy marker improves positive predictive value for breast cancer diagnosis: preliminary study. Radiology 2006;239(3):686–692.
16. Bartella L, Thakur SB, Morris EA, et al. Enhancing nonmass lesions in the breast: evaluation with proton (1H) MR spectroscopy. Radiology 2007;245(1):80–87.
17. Cecil KM, Schnall MD, Siegelman ES, Lenkinski RE. The evaluation of human breast lesions with magnetic resonance imaging and proton magnetic resonance spectroscopy. Breast Cancer Res Treat 2001;68(1):45–54.

18. Dorrius MD, Pijnappel RM, van der Weide Jansen MC, et al. The added value of quantitative multi-voxel MR spectroscopy in breast magnetic resonance imaging. Eur Radiol 2012;22(4):915–922.

19. Gruber S, Debski BK, Pinker K, et al. Three-dimensional proton MR spectroscopic imaging at 3 T for the differentiation of benign and malignant breast lesions. Radiology 2011;261(3):752–761.

20. Huang W, Fisher PR, Dulaimy K, Tudorica LA, O'Hea B, Button TM. Detection of breast malignancy: diagnostic MR protocol for improved specificity. Radiology 2004;232(2):585–591.

21. Jacobs MA, Barker PB, Bottomley PA, Bhujwalla Z, Bluemke DA. Proton magnetic resonance spectroscopic imaging of human breast cancer: a preliminary study. J Magn Reson Imaging 2004;19(1):68–75.

22. Kim JK, Park SH, Lee HM, et al. In vivo 1H-MRS evaluation of malignant and benign breast diseases. Breast 2003;12(3):179–182.

23. Kousi E, Tsougos I, Vasiou K, et al. Magnetic resonance spectroscopy of the breast at 3T: pre- and post-contrast evaluation for breast lesion characterization. ScientificWorldJournal 2012;2012:754380.

24. Kvistad KA, Bakken IJ, Gribbestad IS, et al. Characterization of neoplastic and normal human breast tissues with in vivo (1) H MR spectroscopy. J Magn Reson Imaging 1999;10(2):159–164.

25. Lu H, Liu PF, Bao RX, Sun F. Evaluation of spectral selected press sequence in breast lesion characterization. Chin Med Sci J 2006;21(4):265–269.

26. Roebuck JR, Cecil KM, Schnall MD, Lenkinski RE. Human breast lesions: characterization with proton MR spectroscopy. Radiology 1998;209(1):269–275.

27. Sah RG, Sharma U, Parshad R, Seenu V, Mathur SR, Jagannathan NR. Association of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 status with total choline concentration and tumor volume in breast cancer patients: an MRI and in vivo proton MRS study. Magn Reson Med 2012;68(4):1039–1047.

28. Sardanelli F, Fausto A, Di Leo G, de Nijs R, Vorbuchner M, Podo F. In vivo proton MR spectroscopy of the breast using the total choline peak integral as a marker of malignancy. AJR Am J Roentgenol 2009;192(6):1608–1617.

29. Thakur SB, Brennan SB, Ishill NM, et al. Diagnostic usefulness of water-to-fat ratio and choline concentration in malignant and benign breast lesions and normal breast parenchyma: an in vivo (1) H MRS study. J Magn Reson Imaging 2011;33(4):855–863.

30. Tozaki M, Fukuma E. 1H MR spectroscopy and diffusion-weighted imaging of the breast: are they useful tools for characterizing breast lesions before biopsy? AJR Am J Roentgenol 2009;193(3):840–849.

31. Tse GM, Cheung HS, Pang LM, et al. Characterization of lesions of the breast with proton MR spectroscopy: comparison of carcinomas, benign lesions, and phyllodes tumors. AJR Am J Roentgenol 2003;181(5):1267–1272.

32. Yeung DK, Cheung HS, Tse GM. Human breast lesions: characterization with contrast-enhanced in vivo proton MR spectroscopy—initial results. Radiology 2001;220(1):40–46.

33. Katz-Brull R, Lavin PT, Lenkinski RE. Clinical utility of proton magnetic resonance spectroscopy in characterizing breast lesions. J Natl Cancer Inst 2002;94(16):1197–1203.

34. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

35. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31.

36. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539–1558.

37. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12(14):1293–1316.

38. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21(11):1525–1537.

39. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50(4):1088–1101.

40. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629–634.

41. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56(2):455–463.

42. Lenkinski RE, Wang X, Elian M, Goldberg SN. Interaction of gadolinium-based MR contrast agents with choline: implications for MR spectroscopy (MRS) of the breast. Magn Reson Med 2009;61(6):1286–1292.

43. Madhu B, Robinson SP, Howe FA, Griffiths JR. Effect of Gd-DTPA-BMA on choline signals of HT29 tumors detected by in vivo 1H MRS. J Magn Reson Imaging 2008;28(5):1201–1208.

44. Murphy PS, Leach MO, Rowland IJ. The effects of paramagnetic contrast agents on metabolite protons in aqueous solution. Phys Med Biol 2002;47(6):N53–N59.

45. Baik HM, Su MY, Yu H, Mehta R, Nalcioglu O. Quantification of choline-containing compounds in malignant breast tumors by 1H MR spectroscopy using water as an internal reference at 1.5 T. MAGMA 2006;19(2):96–104.

46. Jagannathan NR, Singh M, Govindaraju V, et al. Volume localized in vivo proton MR spectroscopy of breast carcinoma: variation of water-fat ratio in patients receiving chemotherapy. NMR Biomed 1998;11(8):414–422.

47. Sharma U, Kumar M, Sah RG, Jagannathan NR. Study of normal breast tissue by in vivo volume localized proton MR spectroscopy: variation of water-fat ratio in relation to the heterogeneity of the breast and the menstrual cycle. Magn Reson Imaging 2009;27(6):785–791.