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

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.
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.

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.

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.

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.
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 $2 \times 2$ 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 $2 \times 2$ 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. $\chi^2$ 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 ($I^2 \geq 75\%$) 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 exploratory bias correction.

**Figure 1**

Search result: n=590

Excluded based on title/abstract: n=531

Full papers further analyzed: n=59

Excluded after review:
- only malignant lesions: n=30
- not 1.5T or 3T: n=3
- <10 lesions: n=4
- redundant data: n=2

Included in the analysis: n=20

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Consecutive Recruitment?</th>
<th>QUADAS Score</th>
<th>No. of Patients</th>
<th>No. of Malignant Lesions</th>
<th>No. of Benign Lesions</th>
<th>MR Spectroscopy Sequence</th>
<th>BD Value</th>
<th>Repetition Time (msec)</th>
<th>Echo Time(s)</th>
<th>Spectroscopy after CM Administration?</th>
<th>Fat Saturation Used?</th>
<th>Method of Choline Analysis and Cutoff Value</th>
<th>FWHM(Hz)</th>
<th>Minimum Lesion Size (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roebuck et al 1998 (26)</td>
<td>NA</td>
<td>11</td>
<td>17</td>
<td>10</td>
<td>7</td>
<td>Single-voxel STEAM</td>
<td>1.5</td>
<td>2000</td>
<td>31,270</td>
<td>Yes</td>
<td>No</td>
<td>Visual</td>
<td>5.6–18</td>
<td>None</td>
</tr>
<tr>
<td>Tse et al 2003 (31)</td>
<td>No</td>
<td>12</td>
<td>46</td>
<td>19</td>
<td>27</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>38,135,270</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 2</td>
<td>NA</td>
<td>15</td>
</tr>
<tr>
<td>Kim et al 2003 (22)</td>
<td>NA</td>
<td>11</td>
<td>35</td>
<td>19</td>
<td>16</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2500</td>
<td>288</td>
<td>No</td>
<td>Yes</td>
<td>SNR</td>
<td>3–8</td>
<td>NA</td>
</tr>
<tr>
<td>Huang et al 2004 (20)</td>
<td>No</td>
<td>11</td>
<td>30</td>
<td>18</td>
<td>12</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>135</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Jacobs et al 2004 (21)</td>
<td>No</td>
<td>11</td>
<td>15</td>
<td>8</td>
<td>7</td>
<td>Chemical shift imaging</td>
<td>1.5</td>
<td>2000</td>
<td>272</td>
<td>Yes</td>
<td>Yes</td>
<td>SNR, 4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bartella et al 2006 (15)</td>
<td>Yes</td>
<td>12</td>
<td>56</td>
<td>31</td>
<td>26</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>135</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 2</td>
<td>&lt;20</td>
<td>10</td>
</tr>
<tr>
<td>Lu et al 2006 (25)</td>
<td>Yes</td>
<td>11</td>
<td>131</td>
<td>74</td>
<td>57</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>1500</td>
<td>110</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 3</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bartella et al 2007 (16)*</td>
<td>Yes</td>
<td>12</td>
<td>32</td>
<td>12</td>
<td>20</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>135</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 2</td>
<td>&lt;20</td>
<td>10</td>
</tr>
<tr>
<td>Baek et al 2008 (13)</td>
<td>Yes</td>
<td>11</td>
<td>36</td>
<td>27</td>
<td>9</td>
<td>Chemical shift imaging</td>
<td>1.5</td>
<td>1627</td>
<td>270</td>
<td>Yes</td>
<td>Yes</td>
<td>SNR, 3, 2</td>
<td>10–29</td>
<td>10</td>
</tr>
<tr>
<td>Sardanelli et al 2009 (28)</td>
<td>Yes</td>
<td>11</td>
<td>42</td>
<td>19</td>
<td>26</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>1500</td>
<td>136</td>
<td>Yes</td>
<td>Yes</td>
<td>Peak area/voxel</td>
<td>&lt;30</td>
<td>10</td>
</tr>
<tr>
<td>Tozaki and Fukuma 2009 (30)</td>
<td>Yes</td>
<td>11</td>
<td>165</td>
<td>91</td>
<td>80</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>1620</td>
<td>270</td>
<td>Yes</td>
<td>No</td>
<td>SNR, 2</td>
<td>10–20</td>
<td>NA</td>
</tr>
<tr>
<td>Balzer et al 2012 (14)</td>
<td>Yes</td>
<td>13</td>
<td>62</td>
<td>42</td>
<td>21</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>272</td>
<td>Yes*</td>
<td>No</td>
<td>SNR, 2</td>
<td>15.7</td>
<td>8</td>
</tr>
<tr>
<td>Thakur et al 2011 (29)</td>
<td>Yes</td>
<td>11</td>
<td>88</td>
<td>57</td>
<td>31</td>
<td>Single-voxel PRESS</td>
<td>1.5</td>
<td>2000</td>
<td>135</td>
<td>Yes</td>
<td>No</td>
<td>Internal reference, 0.1 mmol/kg</td>
<td>&lt;20</td>
<td>10</td>
</tr>
<tr>
<td>Gruber et al 2011 (19)</td>
<td>Yes</td>
<td>11</td>
<td>50</td>
<td>31</td>
<td>19</td>
<td>Chemical shift imaging</td>
<td>3</td>
<td>750</td>
<td>145</td>
<td>No</td>
<td>Yes</td>
<td>SNR, 2, 6</td>
<td>25–35</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 1 (continues)
### Key Parameters Extracted from the Investigated Studies

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

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.

---

**Spectroscopy: Acquisition Technique and Spectra Analysis**

In the majority (n = 18) of the eligible studies, a minimum lesion size as an inclusion criterion was defined in 11 studies (12–18,24,28,29) 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 (12–18,24,28,29) and suspicious findings at MR imaging in four investigations (15,16,20,23). In the majority of studies, MR spectroscopy was 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 postcontrast acquisitions, as sensitivity and specificity values were not clearly described in the text (22,23,26–27) (Table). A minimum lesion size of 10 mm was defined in 11 studies (12–18,24,28,29) 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 (12–18,24,28,29) and suspicious findings at MR imaging in four investigations (15,16,20,23).
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 max-
imum 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 re-
ported in 15 studies (12–16,18,19,22–
24,26–30) (Table).

Information on who planned spec-
troscopic 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) pre-
sented 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), me-
dian 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 con-
sidered was between 14 and 92 years.
Overall mean or median lesion siz-
es were described in five studies
(14,15,21,22,26) and ranged between
17 and 35 mm. Lesion sizes of be-
nign 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 be-
nign lesions. Minimum described be-
nign lesion size was 4 mm (30), and
minimum malignant lesion size was
2 mm (19).

Synthesis of General Diagnostic
Performance
Individual study results–weighted sum-
maries of sensitivity, specificity, and
DORs, together with their 95% confi-
dence intervals (CIs) are provided inFig-
ures 2–4. Pooled sensitivity, specificity,
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 ($I^2 = 84.5\%$, $P < .0001$) and DOR ($I^2 = 53\%$, $P < .0001$).

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 ($I^2 = 88.3\%$, $P < .0001$) and DOR ($I^2 = 65\%$, $P < .0001$). 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...
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 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 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.

**Figure 5**

Summary ROC curve of individual-study MR spectroscopy results. *SE* = standard error of the estimate.
EVIDENCE-BASED PRACTICE: Meta-Analysis of Diagnosis of Breast Lesions at Proton MR Spectroscopy

Baltzer and Dietzel

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 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.
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 studies. Further study is warranted to clarify especially the role of multichannel technology, higher field strength, and spectroscopic imaging in breast MR spectroscopy, as published previously.

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: 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.
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


