Dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted magnetic resonance imaging for predicting the response of locally advanced breast cancer to neoadjuvant therapy: a meta-analysis

John Virostko
Allison Hainline
Hakmook Kang
Lori R. Arlinghaus
Richard G. Abramson
Stephanie L. Barnes
Jeffrey D. Blume
Sarah Avery
Debra Patt
Boone Goodgame
Thomas E. Yankeelov
Anna G. Sorace

Dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted magnetic resonance imaging for predicting the response of locally advanced breast cancer to neoadjuvant therapy: a meta-analysis

John Virostko, Allison Hainline, Hakmook Kang, Lori R. Arlinghaus, Richard G. Abramson, Stephanie L. Barzies, Jeffrey D. Blume, Sarah Avery, Debra Patt, Boone Goodgame, Thomas E. Yankeelov, and Anna G. Sorace

University of Texas at Austin, Department of Medicine, Austin, Texas, United States
Texas Oncology, Austin, Texas, United States
Austin Radiological Association, Austin, Texas, United States
University of Texas at Austin, Department of Biomedical Engineering, Austin, Texas, United States
Vanderbilt University, Department of Biostatistics, Nashville, Tennessee, United States
Vanderbilt University, Center for Quantitative Sciences, Nashville, Tennessee, United States
Vanderbilt University Medical Center, Department of Radiology and Radiological Sciences, Nashville, Tennessee, United States
University of Texas at Austin, Institute for Computational and Engineering Sciences, Austin, Texas, United States
University of Texas at Austin, Department of Biomedical Engineering, Austin, Texas, United States
Texas Oncology, Austin, Texas, United States
Seton Hospital, Austin, Texas, United States
University of Texas at Austin, Department of Medicine, Austin, Texas, United States

Abstract. This meta-analysis assesses the prognostic value of quantitative dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and diffusion-weighted MRI (DW-MRI) performed during neoadjuvant therapy (NAT) of locally advanced breast cancer. A systematic literature search was conducted to identify studies of quantitative DCE-MRI and DW-MRI performed during breast cancer NAT that report the sensitivity and specificity for predicting pathological complete response (pCR). Details of the study population and imaging parameters were extracted from each study for subsequent meta-analysis. Meta-regression analysis, subgroup analysis, study heterogeneity, and publication bias were assessed. Across 10 studies that met the stringent inclusion criteria for this meta-analysis (out of 325 initially identified studies), we find that MRI had a pooled sensitivity of 0.91 [95% confidence interval (CI), 0.80 to 0.96] and specificity of 0.81 (95% CI, 0.68 to 0.89) when adjusted for covariates. Quantitative DCE-MRI exhibits greater specificity for predicting pCR than semi-quantitative DCE-MRI (p < 0.001). Quantitative DCE-MRI and DW-MRI are able to predict, early in the course of NAT, the eventual response of breast tumors, with a high level of specificity and sensitivity. However, there is a high degree of heterogeneity in published studies highlighting the lack of standardization in the field.

Keywords: response prediction; dynamic contrast-enhanced MRI; diffusion-weighted MRI; breast cancer; neoadjuvant; meta-analysis.

Paper 17205SSR received Jul. 7, 2017; accepted for publication Nov. 6, 2017; published online Nov. 24, 2017.

1 Introduction

Neoadjuvant therapy (NAT) is widely considered the standard of care for the treatment of locally advanced breast cancer. NAT increases the success rate for breast conservation surgery by reducing tumor burden and provides the opportunity to treat micrometastases at an earlier time point compared to adjuvant treatment. Importantly, patients who achieve a pathological complete response (pCR; i.e., complete absence of viable tumor cells in the breast or axilla at the time of surgery) in the neoadjuvant setting have increased survival compared with patients who have residual disease at the conclusion of NAT—should it be determined—early in the course of treatment. The ability to predict which breast cancer patients will eventually achieve pCR presents a formidable challenge. Conventional, tissue-based biomarkers of response require a biopsy, which can include sampling errors due to tumor heterogeneity. Conversely, imaging approaches assess the entire tumor, obviating sampling error. However, imaging protocols require infusion of a gadolinium contrast agent, which may be retained in the brain and must be optimized and validated for predicting tumor response to therapy. Furthermore, the...
optimal timing for image acquisition during NAT must be established to maximize the predictive ability of quantitative imaging. The most commonly used method for quantitatively assessing the response of a tumor to NAT is the response evaluation criteria in solid tumors (RECIST). RECIST tracks regression in tumor size and has been shown to correlate with survival in a number of different cancers. In its current version, RECIST 1.1 uses imaging to identify, measure, and sum the longest dimension of up to five lesions prior to treatment. These dimensions are summed and compared with similar measurements post-treatment. The resulting change over time is categorized as complete response (disappearance of all target lesions), partial response (≥30% decline in sum of dimensions), progressive disease (≥20% increase in dimensions or appearance of new lesions), or stable disease (none of the preceding conditions met). However, there are a number of limitations in both making RECIST measurements accurately in the setting of irregular and/or indiscernible tumor margins, as well as in capturing tumor complexity with a single measurement. Furthermore, standard tumor size-based methods of evaluating response (including RECIST) lag behind a tumor’s biological response to treatment, such as cellular and vascular alterations. Moreover, size-based techniques may underestimate early efficacy for targeted agents exhibiting predominantly cytostatic rather than cytotoxic effects. Fortunately, a number of MRI techniques have matured to the point where they can offer a quantitative description of tumor characteristics that have shown the ability to predict the response of locally advanced breast cancer to NAT. In this meta-analysis, we focus on the two MRI methods that have accumulated the largest body of data for predicting the response of locally advanced breast cancer to NAT: dynamic contrast-enhanced MRI (DCE-MRI) and diffusion-weighted MRI (DW-MRI).

2 Rationale and Objectives

This meta-analysis assesses the ability of DCE- and DW-MRI for predicting, early in the course of NAT, which breast cancer patients will achieve pCR at the conclusion of NAT. Previous meta-analyses and systematic reviews evaluating the use of MRI in the neoadjuvant setting for breast cancer have not assessed its predictive value. A 2010 meta-analysis, performed by Yuan et al., examined MRI performed following the completion of NAT. Similarly, a number of systematic reviews have investigated the ability of MRI to assess the response of breast cancer either immediately prior to and/or after the completion of NAT. This highlights a subtle, but important, distinction in that the present study focuses on MRI performed during the course of NAT, rather than the end of NAT, with the goal of predicting eventual response. A previous systematic review of MRI performed during NAT did not perform meta-analysis due to heterogeneity in MRI parameters and outcome definitions factors that also limited inclusion in the present study.

For the purpose of this meta-analysis, we have focused on the techniques of DCE-MRI and DW-MRI for predicting response. We further subdivide DCE-MRI into semi quantitative measurements, which generate measures lacking direct physiological correlates, and quantitative measurements, which do have a direct physiological correlate. The intricacies of these techniques are discussed further below. We also include multiparametric studies combining measurements from both DCE-MRI and DW-MRI.

2.1 Semiquantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging

DCE-MRI is an umbrella term used to describe a wide variety of dynamic MRI techniques and analytic approaches, including both qualitative and quantitative methods. Common to all approaches is the serial acquisition of heavily T₁-weighted images before and after the injection of the paramagnetic contrast agent. In the clinical setting, great emphasis is placed on obtaining DCE-MRI data at high spatial resolution which necessitates a lower temporal resolution, resulting in time series data that can only be analyzed qualitatively or semiquantitatively. Semiquantitative analysis involves the extraction of descriptive parameters describing this time series data, such as enhancing volume over time, general curve shape features, or the ratio between the lesion intensity before and after contrast. Figure 1 displays an example of a semiquantitative analysis of DCE-MRI data obtained from a patient who achieved pCR (top row) and a patient who did not achieve pCR (bottom row) performed before, after the first cycle, and at the conclusion of all NAT. Application of semiquantitative DCE-MRI in the NAT setting has demonstrated that the percentage of tumor voxels demonstrating progressive (i.e., increasing signal intensity with time), plateau (steady intensity), or washout (decreasing intensity) kinetics predicts response to therapy. Alternately, comparison of early and late enhancement measurements of peak signal enhancement and changes in DCE-MRI time course shape are similarly predictive of pathological response. Importantly, the parameters derived from semiquantitative DCE-MRI provide predictive indicators of NAT response that are independent of traditional measures such as tumor size. The prognostic value of DCE-MRI was established in the I-SPY 1 trial, which showed that multivariate models incorporating semiquantitative DCE-MRI, histopathology, and breast cancer subtype were the most predictive of therapeutic response.

2.2 Quantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging

In order to perform quantitative DCE-MRI, the following measurements are required: a precontrast T₁ map, high-temporal resolution dynamic T₁-weighted data, estimation of the time rate of change of the concentration of contrast agent in the blood plasma (i.e., the arterial input function, AIF), and a pharmacokinetic model to analyze the resulting data. By fitting the data to such a model (e.g., the Tofts model), one can extract rate constants that reflect the influx of contrast agent into tissue (ktrans, the volume transfer constant), efflux of contrast agent back into plasma (kep), fractional volume of the blood plasma (vpl), and fractional volume of the extracellular extravascular space (ve). This analysis can be conducted on a voxel-by-voxel basis allowing the construction of parametric maps. For example, Fig. illustrates an example of kep map overlaid on anatomical images of a patient who achieved pCR (top row) and a patient who did not achieve pCR (bottom row) performed before, after the first cycle, and at the conclusion of all NAT. Following only one cycle of NAT, quantitative DCE-MRI parameters have been shown to be excellent predictors of pCR, whereas changes in longest dimension as characterized by RECIST were poor predictors of pCR even at the midpoint of NAT.
2.3 Diffusion-Weighted Magnetic Resonance Imaging

The rate of water diffusion in cellular tissues can be described by an “apparent diffusion coefficient” (ADC), which depends, to a great extent, on the number and separation of barriers that a diffusing water molecule encounters. MRI methods have been developed to map the ADC and test–retest studies indicate that the ADC is highly repeatable and reproducible. Similar to DCE-MRI parametric mapping, the ADC can be computed...
on a voxel-by-voxel basis, yielding ADC parametric maps; for example, Fig. 3 presents representative ADC maps overlaid on anatomical images of a patient who achieved pCR (top row) and a patient who did not achieve pCR (bottom row) performed before and after the first cycle of NAT. As with DCE-MRI, DW-MRI has demonstrated the ability to predict response to therapy when performed during the course of NAT. Many publications have since appeared, validating the ability of DW-MRI to predict response in breast cancer. In particular, DW-MRI performed after the first cycle of NAT demonstrates that ADC increases in responders but not nonresponders and that changes in ADC correlate with decrease in tumor volume. Furthermore, changes in ADC may outperform measurements of tumor size for predicting tumor response to NAT. A multisite trial has confirmed that changes in ADC early during NAT are predictive of response with an area under the receiver operating curve of 0.825.

2.4 Multiparametric DCE-MRI and DW-MRI

DCE-MRI and DW-MRI have separately demonstrated clear prognostic value in breast cancer NAT, but their greatest predictive value may result from combining the two techniques. Multiparametric methods that combine quantitative parameters from both DCE- and DW-MRI outperform metrics derived from either technique individually for predicting pCR either early in the course of NAT or following the completion of NAT. One study of integrated DCE- and DW-MRI indicated that \( k_{\text{trans}} \) and ADC are the most sensitive metrics to changes in the tumor occurring between initiation of NAT and surgery, although the optimal combination of multiparametric measurements is under investigation. Other studies have included results from not only DCE- and DW-MRI but also added magnetic resonance spectroscopy (MRS) or susceptibility-weighted MRI.

We have included multiparametric studies comprising both DCE-MRI and DW-MRI in this meta-analysis in order to assess the benefit of combining DCE-MRI and DW-MRI measurements.

3 Methods

3.1 Identification of Eligible Studies

This meta-analysis was prospectively registered in the PROSPERO registry with the registration number CRD42016038770. The preferred reporting items for systematic reviews and meta-analyses checklist was followed when performing this meta-analysis. A comprehensive literature search was performed to identify studies reporting the sensitivity and specificity of DCE- and DW-MRI for predicting pCR in breast cancer patients receiving NAT as a component of their clinical care. The Pubmed and Cochrane library databases were searched from January 2001 through May 2017 using the following search terms: neoadjuvant and “breast cancer” and “contrast-enhanced” and MRI, preoperative and “breast cancer” and “contrast-enhanced” and MRI, neoadjuvant and “breast cancer” and diffusion and MRI, preoperative and “breast cancer” and diffusion and MRI. This search yielded a total of 325 studies eligible for meta-analysis. Duplicate studies were removed to yield 260 studies. From this list of 260 eligible studies, two reviewers (A.G.S. and J.V.) independently reviewed all studies according to the following inclusion criteria: must be reported in the English language, must report on human subjects, must include 10 or more subjects, peer-reviewed original article (no reviews, brief communications, or letters to the editor), sufficient data to determine specificity and sensitivity, the outcome measure must be pCR, MRI must be performed as a “predictive” measure (i.e., MRI must occur during the course of NAT, not after completion of NAT). Three studies met all inclusion criteria but included a receiver operator characteristic curve rather than providing sensitivity and specificity. For these studies, the corresponding author was contacted who then provided the sensitivity and specificity. Among reports with overlapping patient data, only the most recent publication was included in the meta-analysis. Studies were assessed for potential eligibility by first applying inclusion criteria on the abstract, followed by the full text if the abstract was deemed eligible or inconclusive. Data from eligible studies were extracted by two independent reviewers (A.G.S. and J.V.) with arbitration by a third reviewer (S.L.B.) in the case of disagreement. An overview of the study selection process is shown in Fig. Ten studies met the stringent inclusion criteria of which three studies performed quantitative DCE-MRI, three studies performed semiquantitative DCE-MRI, two studies performed DWI-MRI, and two multiparametric studies reported MRI metrics, which combined results from both DCE-MRI and DWI-MRI.

3.2 Data Extraction

For each report, the following data were extracted into a standardized entry form: first author, journal, year of publication, number of cases, patient age (mean and range), initial clinical stage, histological subtype, receptor subtype (estrogen receptor positive (ER+), progesterone receptor positive (PR+), human epidermal growth factor receptor 2-positive (HER2+), triple negative), preoperative therapeutic regimen, MRI time points used for prediction, pCR rate, magnetic field strength, contrast agent, contrast agent dose, DCE-MRI temporal resolution,
b-values for DW-MRI acquisition, and either sensitivity/specificity or sufficient data to construct a $2 \times 2$ contingency table. Additionally, the MRI type was categorized as either semi-quantitative DCE-MRI (in which the MRI parameter does not have direct physiological units), quantitative DCE-MRI (MRI parameter has physiological units), DW-MRI, or multiparametric (combination of DCE-MRI and DW-MRI). The MRI parameter used to define the sensitivity and specificity was also extracted. A summary of these parameters for studies included in this meta-analysis is given in Table 1.

The methodological quality of each study in this meta-analysis was assessed using the quality assessment of diagnostic accuracy studies (QUADAS) tool. Using the QUADAS tool, an overall quality score was calculated (maximum 14). The average number of patients in the studies in the meta-analysis was 64, and the minimum number of patients was 22.

3.3 Statistical Analysis

All analyses were performed using Stata version 14.1 (StataCorp, College Station, Texas) and R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria). The sensitivity, specificity, and log diagnostic odds ratios were calculated along with the associated 95% confidence intervals and overall adjusted estimates were pooled across all studies. The amount of heterogeneity across the studies was evaluated via the $I^2$ statistic, which represents the percentage of total variation in meta-analysis that can be attributed to between-study heterogeneity. This heterogeneity was taken into account in the evaluation of test performance with a hierarchical summary receiver operator curve (HSROC). The estimation of the HSROC uses a Bayesian hierarchical approach to account for both within- and between-study heterogeneity. Publication bias was examined graphically via a funnel plot in which the diagnostic odds ratio is plotted versus the inverse root of the effective sample size and was tested using Deek’s method. Also, note that any empty cells in the $2 \times 2$ table setup were replaced with 0.1 to align with requirements of statistical programming software as well as to allow for calculation of diagnostic odds ratios.

The large degree of heterogeneity in this meta-analysis implies that the data should not be pooled in a fixed-effect model. As a result, a random effects logistic regression model was fit to model sensitivity and specificity while accounting for differences between individual studies that are not due to study characteristics. The meta-regression analysis is given by the following equation:

$$
\logit(\Pr[MRI_{ij} = 1 | X_{ij}]) = (\beta_0 + b_{0j}) + \beta_1 \cdot pCR_{ij} + \beta_2 \cdot ER_i + \beta_3 \cdot PR_i + \beta_4 \cdot HER2_i + \beta_5 \cdot Age_i + \beta_6
$$

- TEMPORAL RESOLUTION, $X_{ij} = 1$ if MRI performed only before and/or after NAT, $X_{ij} = 0$ otherwise; $X_{ij}$ = 1 if NAT performed (n = 38); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if outcome not given as pCR (n = 30); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if review article/not peer reviewed (n = 26); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if cannot calculate sensitivity/specificity (n = 14); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if fewer than 10 patients (n = 11); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if imaging metric is not DCE-MRI or DW-MRI (n = 8); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if overlapping data (n = 5); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if other (n = 25).

The estimation of the HSROC uses a Bayesian hierarchical approach to account for both within- and between-study heterogeneity. Publication bias was examined graphically via a funnel plot in which the diagnostic odds ratio is plotted versus the inverse root of the effective sample size and was tested using Deek’s method. Also, note that any empty cells in the $2 \times 2$ table setup were replaced with 0.1 to align with requirements of statistical programming software as well as to allow for calculation of diagnostic odds ratios.

The large degree of heterogeneity in this meta-analysis implies that the data should not be pooled in a fixed-effect model. As a result, a random effects logistic regression model was fit to model sensitivity and specificity while accounting for differences between individual studies that are not due to study characteristics. The meta-regression analysis is given by the following equation:

$$
\logit(\Pr[MRI_{ij} = 1 | X_{ij}]) = (\beta_0 + b_{0j}) + \beta_1 \cdot pCR_{ij} + \beta_2 \cdot ER_i + \beta_3 \cdot PR_i + \beta_4 \cdot HER2_i + \beta_5 \cdot Age_i + \beta_6
$$

- TEMPORAL RESOLUTION, $X_{ij} = 1$ if MRI performed only before and/or after NAT, $X_{ij} = 0$ otherwise; $X_{ij}$ = 1 if NAT performed (n = 38); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if outcome not given as pCR (n = 30); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if review article/not peer reviewed (n = 26); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if cannot calculate sensitivity/specificity (n = 14); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if fewer than 10 patients (n = 11); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if imaging metric is not DCE-MRI or DW-MRI (n = 8); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if overlapping data (n = 5); $X_{ij}$ = 0 otherwise; $X_{ij}$ = 1 if other (n = 25).

The large degree of heterogeneity in this meta-analysis implies that the data should not be pooled in a fixed-effect model. As a result, a random effects logistic regression model was fit to model sensitivity and specificity while accounting for differences between individual studies that are not due to study characteristics. The meta-regression analysis is given by the following equation:

$$
\logit(\Pr[MRI_{ij} = 1 | X_{ij}]) = (\beta_0 + b_{0j}) + \beta_1 \cdot pCR_{ij} + \beta_2 \cdot ER_i + \beta_3 \cdot PR_i + \beta_4 \cdot HER2_i + \beta_5 \cdot Age_i + \beta_6
$$

Where $i$ denotes the study ($i = 1, \ldots, 10$) and $j$ denotes the individual within study $i$. MRI$_{ij}$ represents the MRI prediction of pCR for the $j$th patient within study $i$; MRI$_{ij} = 1$ represents an MRI prediction of pCR; and MRI$_{ij} = 0$ represents an MRI prediction of a non-pCR (i.e., residual tumor remaining at time of surgery after NAT). Similarly, a patient with pCR$_{ij} = 1$ achieved a pCR and pCR$_{ij} = 0$ indicates non-pCR. ER, PR, and HER2 are percentages of each study population that exhibited overexpression of each receptor, respectively. Age is mean age within each study. Temporal resolution and the number of NAT cycles (NATcycle) between start of NAT and MRI were constant within each study. MRI type was constant within each study and took values of either quantitative DCE, semi-quantitative DCE, multiparametric, or DWI. Note that the inclusion of pCR$_{ij}$ in the model in combination with the outcome of pCR$_{ij}$ in the model in combination with the outcome of
Table 1 Description of studies included in the meta-analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>No. of patients</th>
<th>Patient age, mean (range)</th>
<th>Initial clinical stage</th>
<th>Histological subtype</th>
<th>Receptor status</th>
<th>Preoperative therapy regimen</th>
<th>Time between start of NAT and MRI</th>
<th>pCr rate</th>
<th>MRI magnetic field strength</th>
<th>Contrast agent</th>
<th>DCE- MRI temporal resolution (s)</th>
<th>DCE-MRI b values</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semiquantitative DCE-MRI studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parikh3</td>
<td>2014</td>
<td>36</td>
<td>49.8 (24–67)</td>
<td>I–III</td>
<td>35 IDC, 1 ILC</td>
<td>21 ER+, 6 HER2+</td>
<td>3°FEC → 4<em>T, 4</em>EC → 4*T HER2+ patients after 2010 received trastuzumab</td>
<td>3 cycles NAT</td>
<td>22.22%</td>
<td>1.5T</td>
<td>Dotarem (20 mL)</td>
<td>92</td>
<td>87.5%</td>
<td>82.1%</td>
<td></td>
</tr>
<tr>
<td>Rigter4</td>
<td>2013</td>
<td>246</td>
<td>48 (18–68)</td>
<td>I–IV</td>
<td>190 IDC, 33 ILC, 6 IDC/ILC</td>
<td>190 ER+, 246 PR+</td>
<td>3<em>AC → 3</em>AC, 3<em>AC → 3</em>TCa</td>
<td>3 cycles NAT</td>
<td>1.63%</td>
<td>1.5T</td>
<td>Gadolinium CA (0.1 mmol/kg)</td>
<td>90</td>
<td>75.0%</td>
<td>33.2%</td>
<td></td>
</tr>
<tr>
<td>Yang3</td>
<td>2012</td>
<td>22</td>
<td>45 (35–67)</td>
<td>II–III</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Antiangiogenic-cytotoxic combination therapy</td>
<td>2 cycles NAT</td>
<td>27.27%</td>
<td>1.5T</td>
<td>Magnevist (0.1 mmol/kg)</td>
<td>70</td>
<td>78.0%</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>Quantitative DCE-MRI studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tateishi49</td>
<td>2012</td>
<td>143</td>
<td>57 (43–72)</td>
<td>I–III</td>
<td>131 IDC, 11 ILC</td>
<td>100 ER+, 82 PR+, 111 HER2+</td>
<td>4<em>FEC → 12</em>T, 4<em>AC → 12</em>T, HER2+ patients received trastuzumab</td>
<td>2 cycles NAT</td>
<td>16.90%</td>
<td>3T</td>
<td>Magnevist (0.1 mmol/kg)</td>
<td>10</td>
<td>51.7%</td>
<td>92.0%</td>
<td></td>
</tr>
<tr>
<td>Ah-See50</td>
<td>2008</td>
<td>28</td>
<td>46 (29–70)</td>
<td>II–III</td>
<td>21 IDC, 3 ILC</td>
<td>Not reported</td>
<td>6*FEC</td>
<td>2 cycles NAT</td>
<td>39.28%</td>
<td>1.5T</td>
<td>Magnevist (0.1 mmol/kg)</td>
<td>12</td>
<td>94.0%</td>
<td>82.0%</td>
<td></td>
</tr>
<tr>
<td>DWI-MRI studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fangberget51</td>
<td>2011</td>
<td>31</td>
<td>50.7 (37–72)</td>
<td>Not reported</td>
<td>24 IDC, 7 ILC</td>
<td>21 ER+, 18 PR+, 11 HER2+, 5 TNBC</td>
<td>4<em>FEC → 2</em>FEC, 4<em>EC→ 2</em>T, HER2+ patients received trastuzumab</td>
<td>4 cycles NAT</td>
<td>35.48%</td>
<td>1.5T</td>
<td>100, 250, 800</td>
<td>100</td>
<td>88.0%</td>
<td>80.0%</td>
<td></td>
</tr>
<tr>
<td>Minarikova44</td>
<td>2017</td>
<td>42</td>
<td>52 (29–74)</td>
<td>I–III</td>
<td>41 ID, 1 ILC</td>
<td>27 ER+, 13 PR+, 5 HER2+</td>
<td>ACT<em>6, ACT</em>8, AC → T</td>
<td>2 cycles NAT</td>
<td>16.67%</td>
<td>3.0T</td>
<td>0, 850</td>
<td>86.67%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiparametric MRI studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li39</td>
<td>2015</td>
<td>37</td>
<td>45 (28–67)</td>
<td>II–III</td>
<td>Not reported</td>
<td>16 ER+, 16 PR+, 12 HER2+, 12 TNBC</td>
<td>4<em>AC → 4</em>T, 4<em>AC → 12</em>T, HER2+ patients received trastuzumab or other treatments</td>
<td>1 cycle NAT</td>
<td>35.14%</td>
<td>3T</td>
<td>Magnevist (0.1 mmol/kg)</td>
<td>16</td>
<td>0, 500</td>
<td>92.0%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Wu45</td>
<td>2015</td>
<td>31</td>
<td>48.4 (33–62)</td>
<td>II–III</td>
<td>Not reported</td>
<td>24 ER+, 23 PR+, 16 HER2+</td>
<td>FEC, FEC → T, T, or other treatments</td>
<td>1 cycle NAT</td>
<td>9.70%</td>
<td>3.0T</td>
<td>Gadovist (0.1 mmol/kg)</td>
<td>40–50</td>
<td>50, 600, 1000</td>
<td>90.9%</td>
<td>83.8%</td>
</tr>
</tbody>
</table>

Note: IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; TNBC, triple negative breast cancer; A, doxorubicin; C, cyclophosphamide; T, taxane; F, fluorouracil; E, epirubicin; D, docetaxel; Ca = capecitabine.
MRI$_{ij}$ allows for our calculation of sensitivity, specificity, and diagnostic odds ratio.

To fit the regression model, an expanded dataset was generated, as previously described\cite{note1}. In the expanded dataset, each study was given a row for each study subject, where each subject was classified as true positive, false positive, false negative, or true negative. Study-level covariates remained constant for all subjects while MRI test result (i.e., positive or negative) and pCR status were pseudoindividual specific. This expanded dataset allows for analysis to be conducted at the pseudoindividual level. Individual study characteristics included in the regression included: percent of the study population that were ER+, percent of the study population that were PR+, percent of the study population that were HER2+, mean age, number of NAT cycles between start of NAT and MRI, DCE temporal resolution, and magnetic field strength (1.5T or 3T). MRI type was highly correlated with magnetic field strength (1.5T or 3T) and magnetic field strength was thus dropped from the final model.

Interactions between each variable and pCR status were considered and included if deemed of scientific interest; i.e., the effect of HER2 status on the association between MRI test result and pCR status. We also assessed interactions between pCR and the other hormone status, but as these were not statistically significant, they were not included in our model. The interaction terms were assessed for their effects on both sensitivity and specificity by likelihood ratio tests while adjusting for the effects of other variables in the model. Not all studies reported temporal resolution, ER, PR, and HER2 rates for their individual study populations, thus, multiple imputation with chained equations was utilized to allow for the inclusion of these studies without loss of information that would result from a complete-case analysis\cite{note2}. In this process, 10 imputed datasets were created, where pCR rate for each study was used to predict the temporal resolution, ER, PR, and HER2 rates that were missing in the original data. These imputed datasets are used to calculate parameter values and variances for temporal resolution, ER, PR, and HER2 in the metaregression analysis.

The logistic regression model provides adjusted estimates of sensitivity and specificity for each study that can be averaged to calculate average adjusted estimates for both sensitivity and specificity. Subgroup analysis was performed according to the type of MRI performed (semiquantitative versus quantitative DCE-MRI). In this process, information was pooled across semiquantitative studies and quantitative studies, respectively, resulting in two independent $2 \times 2$ tables. A simple test of the difference between two independent means was conducted for each of the three outcomes of interest: log diagnostic odds ratio, sensitivity, and specificity.

### 4 Results

#### 4.1 Meta-analysis Outcome Measures

The funnel plot of diagnostic odds ratio versus the inverse root of the effective sample size revealed a lack of asymmetry and, consequently, no evidence of publication bias ($p = 0.34$). The selection of only 10 studies reveals possible publication bias, though the small sample size prevents a clear conclusion. The high $I^2$ statistic calculated in this meta-analysis (90%) suggests a high degree of heterogeneity present among studies. Additionally, the unadjusted pooled analysis revealed a significant degree of heterogeneity among the 10 included studies in both sensitivity (78.8%) and specificity (99.5%). Pooled, unadjusted estimates of sensitivity and specificity for MRI in predicting pCR are given in Table \ref{tab:meta-analysis-outcome}. Due to the high amount of heterogeneity, we chose to fit a random effects metaregression model to help account for differences between individual studies.

Test performance of the MRI methodologies across all studies is summarized in the HSROC, where individual studies are shown alongside the pooled estimate (Fig. \ref{fig:hsroc}). This figure was created via the hierarchical model from Rutter and Gatsonis\cite{note3}. The area under the HSROC curve was 0.92 (95% CI, 0.89–0.94). The width of the 95% confidence contour demonstrates the amount of heterogeneity present between the 10 studies. Of note, 4 of the 10 studies cluster closely to this receiver operating curve. The two largest studies are seen to have considerable influence over the width of the prediction contour. The width of

\[ \begin{array}{l|c|c}
   & \text{Unadjusted-mean (95% CI)} & \text{Adjusted-mean (95% CI)} \\
\hline
   \text{Sensitivity} & 0.85 (0.70, 0.93) & 0.91 (0.80, 0.96) \\
   \text{Specificity} & 0.86 (0.72, 0.94) & 0.81 (0.68, 0.89) \\
   \text{Diagnostic odds ratio} & 35 (11, 107) & 42.81 (13.67, 134.06) \\
\end{array} \]

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{hsroc.png}
\caption{The HSROC curve displays test performance of individual studies as well as the pooled average estimate. The summary operating point and large area under the curve demonstrate that DCE- and DW-MRI can achieve high sensitivity and specificity for predicting pCR in the neoadjuvant setting for breast cancer. The width of the confidence contour demonstrates the high amount of heterogeneity present in the included selection of studies. Each circle represents a study and the size of the circle refers to the size of the study.}
\end{figure}
the prediction contour demonstrates the uncertainty with which we may predict the results of a future study. The variation among studies included in this meta-analysis does not allow for a targeted prediction of the performance of an unobserved future study.

The studies had QUADAS scores ranging from 10 to 12. Due to the nature of the MRI being performed early in all studies, the time between the MRI and pathological evaluation was not short enough to be reasonably sure that the target condition did not change between MRI and pathology. Furthermore, it was not clear that the MRI result was evaluated without knowledge of the pathology results. Some studies did not explain study withdrawals or uninterpretable test results sufficiently.

4.2 Metaregression Analysis

The random effects model described in Sec. 3.3 was fit to account for possible heterogeneity between studies. In this technique, each study has allowed its own unique baseline effect while the covariate effects were assumed to be similar across studies. This regression model allows for estimates of both sensitivity and specificity to be adjusted for study-specific covariate
values. These adjusted estimates are shown in Table 2 along with similar estimates that have not been adjusted for covariates. Note that adjusting for covariates resulted in a small decrease in estimated specificity (0.81 adjusted versus 0.86 unadjusted), whereas the sensitivity (0.91 adjusted versus 0.85 unadjusted) and diagnostic odds ratio (42.81 adjusted versus 35 unadjusted) were increased by adjustment for covariates. A forest plot of individual study estimates of sensitivity, specificity, and log diagnostic odds as well as overall adjusted estimates is given by Fig. 3. The forest plot visual demonstrates the heterogeneity among studies as well as the variation that is present within each individual study.

Using our random effects model, we found that both MRI type and the number of cycles of NAT before the MRI had significant effects on the sensitivity and specificity simultaneously, at a 95% confidence level. MRI performed earlier after the start of NAT (fewer NAT cycles prior to MRI) led to increased specificity but decreased sensitivity compared with MRI performed later in the course of treatment. For both analyses, the models took into account the effect of other covariates in the model; i.e., ER, PR, temporal resolution, and magnetic field strength. The parameter estimates as well as standard errors and p-values for the metaregression analysis are shown in Table 3.

### 4.3 Subgroup Analysis

A subgroup analysis was performed to determine if semiquantitative and quantitative DCE-MRI were statistically different in terms of sensitivity, specificity, and diagnostic odds ratios. Three quantitative DCE-MRI (199 total subjects) and three semiquantitative DCE-MRI (304 total subjects) studies were included in this subgroup analysis. T-test analysis indicates that quantitative techniques resulted in higher specificity and log diagnostic odds ratio (p < 0.001 and p = 0.014, respectively) at a 5% significance level. Semiquantitative techniques resulted in higher sensitivity than quantitative techniques, though this difference is not statistically significant at a 5% level (p = 0.13).

### 5 Discussion

This meta-analysis identified 10 studies performing DCE- or DW-MRI during the course of NAT for breast cancer that reported the sensitivity and specificity for predicting pathological response. The pooled sensitivity, specificity, and diagnostic odds demonstrate the potential of MRI for early prediction of response to NAT. However, there was a high degree of heterogeneity between these studies, as expected, given the variation in MRI methods and patient populations. A random effects model examining differences between studies in terms of patient population (patient age, percent of ER, PR, and HER2+) and imaging protocol (magnetic field strength, temporal resolution of DCE-MRI, cycle of NAT) found that the type of MRI performed and the time between the start of NAT and MRI significantly influence sensitivity and specificity simultaneously. MRI performed earlier after the start of NAT had increased specificity but decreased sensitivity for predicting pCR. This highlights the need to identify the best time point for predicting response. While this study may be underpowered to detect the influence of other covariates due to the small number of studies, it suggests that accuracy of prediction of response is not influenced by differences in patient populations or other imaging parameters included in the random effects model.

Although this study sought to compare quantitative DCE-MRI, semiquantitative DCE-MRI, DW-MRI, and multiparametric approaches incorporating both DCE- and DW-MRI, the limited number of studies reduced our ability to compare these groups. There were only two multiparametric studies and two DW-MRI studies, which met the inclusion criteria for this meta-analysis. Subgroup analysis comparing quantitative and semiquantitative DCE-MRI indicated that quantitative techniques had a higher specificity and log diagnostic odds ratio. Further studies are needed to validate whether the added rigor of quantitative DCE-MRI translates into improved predictive power for MRI in the neoadjuvant setting. Studies comparing quantitative and semiquantitative DCE-MRI performed on the same data would be especially useful for this evaluation.

The number of studies included in this meta-analysis was limited by heterogeneity in the reported metrics on patient outcome. The success of NAT for breast cancer is commonly measured either in terms of patient survival or pathological response at resection. We confined this meta-analysis to studies reporting patient outcome as pCR, indicating a lack of positive tumor margins or nodal involvement at the time of resection. However, our study design excluded a number of excellent studies, which reported NAT success in terms of patient survival. Survival and pCR correlate, as demonstrated by pooled analysis of nearly 12,000 women with breast cancer; however, this study noted heterogeneity in this relationship as a function of receptor subtype. The correlation between recurrence free survival and pCR is demonstrably improved when stratified by receptor subtype. Given this dependence on receptor subtype, we were unable to combine studies reporting pCR and survival in this meta-analysis. Even among studies reporting outcome in terms of survival, some outcomes were reported as overall survival whereas others were given as disease-free survival. Similarly, studies reporting pathological response commonly
The lack of standardization in outcome metrics confounds meta-analyses and hampers direct comparison of different imaging techniques.

Lack of standardization is similarly of concern in both image acquisition and processing. There are a number of parameters that can be adjusted in MRI acquisition, and there is currently no consensus on the optimal settings for acquiring images. To the best of our knowledge, there are no multisite, multivendor studies validating the quantitative MRI techniques currently in use. The quantitative imaging network (QIN) is attempting to address this shortcoming by standardizing methods for image acquisition, analysis, and sharing. Additionally, the QIN seeks to cross-calibrate imaging results obtained at different centers and build reference datasets for development and validation of new quantitative imaging methods. There is also a paucity of clinical results assessing the repeatability through test–retest studies and reproducibility, through multisite trials, of quantitative MRI metrics performed in the breast. In order to be adopted as routine biomarkers, the repeatability of quantitative MRI metrics must be established and validated with multisite studies.

Image processing of quantitative and semiquantitative MRI is an area of active research. One burgeoning field, known as radiomics, uses computer-aided image processing to extract a large number of quantitative parameters from each image. Radiomics has recently been applied to imaging in breast cancer NAT. Indeed, one of the studies included in this meta-analysis employs a radiomic approach to describe the irregularity of contrast-enhanced MRI. Texture analysis of breast cancer DCE-MRI demonstrates differences between responders and nonresponders. Radiomic analysis can be performed on both quantitative and semiquantitative DCE-MRI and comparisons of these approaches are needed. There are also a number of predictive models under development that integrate multiple parameters extracted from quantitative MRI. For example, mathematical models based on tissue mechanical properties and constrained by DW-MRI data may be useful for predicting pCR. Alternately, predictive models incorporating DCE- and DW-MRI to estimate tumor cell proliferation can be used to predict pCR.

There are a number of other imaging techniques being applied to image the breast, which were not included in this meta-analysis. One such technique is the use of MRS to characterize the concentration of certain tumor metabolites. For instance, MRS can detect choline, a marker of high cellular turnover, which is elevated in certain tumors and inflammatory processes. MRS following NAT demonstrates a decline in choline concentration, although this may not predict pCR as well as DW-MRI. Alternately, MRS has been used to measure the ratio of water signal to fat signal early in the course of NAT and found that this ratio may have predictive power. Positron emission tomography (PET) is another promising method for detecting molecular changes in tumors in response to therapy. A multimodal study combining DCE-MRI and fluorodeoxyglucose (FDG)-PET found that both techniques were able to independently predict disease free survival, but that the combination of MRI and PET gave the best prognostic value. A similar result was found when combining DW-MRI and FDG-PET after the completion of NAT, wherein the combination of MRI and PET metrics improved the specificity of predicting response. The fusion of quantitative multimodal imaging techniques, such as MRI and PET, can provide a more complete characterization of the tumor and provide a voxel-level fusion of complementary imaging data that can be coregistered longitudinally over the course of NAT. A third imaging technique, currently in its early stages, is the application of near-infrared light to perform diffuse optical tomography of the breast. A pilot study of diffuse optical tomography demonstrates that the technique can detect vascular alterations in breast cancer early in the course of NAT. A direct comparison between DCE-MRI and diffuse optical tomography demonstrated that optical tomography differentiates responders and nonresponders as early as the first cycle of treatment and is equally effective in predicting response as DCE-MRI.

6 Conclusion

This comprehensive meta-analysis demonstrates that DCE- and DW-MRI performed during breast cancer NAT can predict pathological response across a range of studies, an exciting and important finding with potential clinical implications. However, the present study also highlights a high degree of heterogeneity (both in patient population and image acquisition/analysis) in the field. Moreover, it reveals a lack of studies that have simultaneously investigated DCE-MRI and DW-MRI as predictive methods to evaluate response to treatment early in the course of NAT. Further development and evaluation of new and established MRI techniques in breast cancer NAT would benefit from standardization in study design, patient population, and reported metrics of prediction accuracy. Additionally, the type of MRI analysis performed may also influence predictive accuracy, as suggested by the improved specificity of quantitative dynamic contrast-enhanced versus semiquantitative techniques.

Disclosures

Dr. Yankeelov reports grants from the National Cancer Institute and the Cancer Prevention and Research Institute of Texas.

Acknowledgments

We thank the National Cancer Institute for support through U01CA174706 and U01CA142565. We thank the Cancer Prevention and Research Institute of Texas (CPRIT) for funding through RR160005. T.E.Y. is a CPRIT scholar of cancer research.

References


37. K. Pinker et al., “Improved diagnostic accuracy with multiparametric magnetic resonance imaging of the breast using dynamic contrast-enhanced magnetic resonance imaging, diffusion-weighted imaging, and 3-dimensional proton magnetic resonance spectroscopic imaging,” PLoS Med. 7(4), e1000136 (2010).


70. N. C. Atuegwu et al., “Parameterizing the logistic model of tumor growth by DW-MRI and DCE-MRI data to predict treatment response and changes in breast cancer cellularity during neoadjuvant chemotherapy,” Breast Oncol. 6(3), 256–264 (2013).

John Virostko is an assistant professor of diagnostic medicine at Dell Medical School at the University of Texas at Austin. He received his master of science and PhD in biomedical engineering from Vanderbilt University. He also received his master of science in clinical investigation from Vanderbilt University. He is interested in developing multimodal imaging techniques and integrating them into computational models of cancer development and progression.

Allison Hainline is a PhD candidate in the biostatistics department at Vanderbilt University. She received her BS in statistics from Baylor University.

Hakmook Kang received his PhD in biostatistics from Brown University in 2011. He joined the faculty at Vanderbilt University in the fall of 2011 and is an assistant professor of biostatistics. He is the core director of biostatistics and bioinformatics core, Vanderbilt Kennedy Center. His research interests are in spatiotemporal modeling and multiple comparisons in biomedical imaging data analysis.

Lori R. Arlinghaus is an imaging research scientist at Vanderbilt University Institute of Imaging Science. She received her master of science and PhD in biomedical engineering from Vanderbilt University.

Richard G. Abramson graduated magna cum laude from Harvard College, where he was elected to Phi Beta Kappa and was a Rhodes scholarship finalist. He obtained his medical degree from Harvard Medical School. He is an associate professor of radiology at Vanderbilt University Medical Center, where he was recently named vice chair for innovation. His research interests span both translational cancer imaging and health care policy and management.

Stephanie L. Barnes is an imaging research scientist at University of Texas at Austin in the Department of Biomedical Engineering with over 10 years of experience in MRI and PET preclinical and clinical acquisition and analysis.

Jeffery D. Blume is an associate professor in the Department of Biostatistics at Vanderbilt University School of Medicine. He is the founding director of graduate studies in biostatistics at Vanderbilt. He received his PhD in biostatistics from John Hopkins School of Public Health. He is the leading expert in likelihood methods for modeling and statistical inference, and he published on the foundations of statistical inference.

Sarah Avery is a practicing radiologist at Austin Radiological Association. She received her medical degrees from Medical College of Texas at Austin in the Department of Biomedical Engineering with over 10 years of experience in MRI and PET preclinical and clinical acquisition and analysis.

Hakmook Kang received his PhD in biostatistics from Brown University in 2011. He joined the faculty at Vanderbilt University in the fall of 2011 and is an assistant professor of biostatistics. He is the core director of biostatistics and bioinformatics core, Vanderbilt Kennedy Center. His research interests are in spatiotemporal modeling and multiple comparisons in biomedical imaging data analysis.

Lori R. Arlinghaus is an imaging research scientist at Vanderbilt University Institute of Imaging Science. She received her master of science and PhD in biomedical engineering from Vanderbilt University.

Richard G. Abramson graduated magna cum laude from Harvard College, where he was elected to Phi Beta Kappa and was a Rhodes scholarship finalist. He obtained his medical degree from Harvard Medical School. He is an associate professor of radiology at Vanderbilt University Medical Center, where he was recently named vice chair for innovation. His research interests span both translational cancer imaging and health care policy and management.

Stephanie L. Barnes is an imaging research scientist at University of Texas at Austin in the Department of Biomedical Engineering with over 10 years of experience in MRI and PET preclinical and clinical acquisition and analysis.

Jeffery D. Blume is an associate professor in the Department of Biostatistics at Vanderbilt University School of Medicine. He is the founding director of graduate studies in biostatistics at Vanderbilt. He received his PhD in biostatistics from John Hopkins School of Public Health. He is the leading expert in likelihood methods for measuring statistical inference, and he published on the foundations of statistical inference.

Sarah Avery is a practicing radiologist at Austin Radiological Association. She received her medical degrees from Medical College of Texas at Austin in the Department of Biomedical Engineering with over 10 years of experience in MRI and PET preclinical and clinical acquisition and analysis.
Georgia and completed a residency in diagnostic radiology at the Robert C. Byrd Health Sciences Center of West Virginia University. She also completed a fellowship in breast/body imaging at Memorial Sloan-Kettering Cancer Center.

Debra Patt is a practicing oncologist and breast cancer specialist in Austin, Texas, and a vice president of Texas Oncology. She is an active leader in breast cancer research, serves on the US Oncology Research breast cancer committee, and chairs the breast cancer subsection of the pathways task force for the US Oncology Network. She received her medical degree and residency training in internal medicine and pediatrics from the Baylor College of Medicine. She completed her training in hematology and medical oncology at the University of Texas MD Anderson Cancer Center.

Boone Goodgame is the medical director of the Shivers Cancer Center as well as the medical director for oncology: Seton Healthcare Family. He is also an assistant professor of medicine in the Dell Medical School. He received his medical degree from Baylor College of Medicine. He completed his residency and fellowship at Barnes Jewish Hospital and Washington University School of Medicine.

Thomas E. Yankeelov is the W.A. “Tex” Moncrief professor of computational oncology and professor of biomedical engineering and diagnostic medicine at the University of Texas at Austin. He also serves as the director of the Center for Computational Oncology in the Institute for Computational and Engineering Sciences. The overall goal of his research is to develop tumor forecasting methods by integrating advanced imaging technologies with predictive, multiscale models of tumor growth to optimize therapy.

Anna G. Sorace is an assistant professor of diagnostic medicine at Dell Medical School at the University of Texas at Austin. Her research interests include developing and utilizing translational advanced multimodality imaging techniques to provide insight into the pathology of diseases, identifying imaging biomarkers for early response to neoadjuvant and adjuvant cancer treatment, and quantitative imaging to guide and enhance drug efficacy.