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

SPIE.

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,[a,b,*] Allison Hainline,[c] Hakmook Kang,[c,d] Lori R. Arlinghaus,[e] Richard G. Abramson,[d,e] Stephanie L. Barnes,[f,g] Jeffrey D. Blume,[c] Sarah Avery,[h] Debra Patt,[i] Boone Goodgame,[j,k] Thomas E. Yankeelov,[a,b,f,g] and Anna G. Sorace[a,b,*]

[a]University of Texas at Austin, Department of Diagnostics, Austin, Texas, United States
[b]University of Texas at Austin, Livestrong Cancer Institutes, Austin, Texas, United States
[c]Vanderbilt University, Department of Biostatistics, Nashville, Tennessee, United States
[d]Vanderbilt University, Center for Quantitative Sciences, Nashville, Tennessee, United States
[e]Vanderbilt University Medical Center, Department of Radiology and Radiological Sciences, Nashville, Tennessee, United States
[f]University of Texas at Austin, Institute for Computational and Engineering Sciences, Austin, Texas, United States
[g]University of Texas at Austin, Department of Biomedical Engineering, Austin, Texas, United States
[h]Austin Radiological Association, Austin, Texas, United States
[i]Texas Oncology, Austin, Texas, United States
[j]Seton Hospital, Austin, Texas, United States
[k]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. Metaregression 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. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JMI.5.1.011011]

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

1 Introduction

Neoadjuvant therapy (NAT) is widely considered the standard of care for the treatment of locally advanced breast cancer.[1,2] 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.[3–7] If it could be determined—early in the course of

NAT—that a particular therapeutic regimen is unlikely to achieve a pCR, the treating physician could discontinue an ineffective treatment and substitute with an alternative regimen that may be more effective. With the numerous options for NAT that have become available, development of a method to predict response early in the course of therapy is especially needed.

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,[8] and must be optimized and validated for predicting tumor response to therapy. Furthermore, the

*Address all correspondence to: John Virostko, E-mail: jack.virostko@austin.utexas.edu; Anna G. Sorace, E-mail: anna.sorace@austin.utexas.edu

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).[9] RECIST tracks regression in tumor size and has been shown to correlate with survival in a number of different cancers.[10–12] 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 ($\geq$30% decline in sum of dimensions), progressive disease ($\geq$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.[13–15] Moreover, size-based techniques may underestimate early efficacy for targeted agents exhibiting predominantly cytostatic rather than cytotoxic effects.[13,14,16] 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.[17] 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.,[18] 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.[19–21] 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,[22] 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 semiquantitative 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.[23] Common to all approaches is the serial acquisition of heavily T_1-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,[24] 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,[24] or the ratio between the lesion intensity before and after contrast.[25] 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.[26] Alternately, comparison of early and late enhancement,[27] measurements of peak signal enhancement,[28] and changes in DCE-MRI time course shape[29] 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.[30] 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.[31]

2.2 Quantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging

In order to perform quantitative DCE-MRI, the following measurements are required: a precontrast T_1 map, high-temporal resolution dynamic T_1-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[32]), one can extract rate constants that reflect the influx of contrast agent into tissue (K^{trans}, the volume transfer constant), efflux of contrast agent back into plasma (k_{ep}), fractional volume of the blood plasma (v_{p}), and fractional volume of the extracellular extravascular space (v_{e}). This analysis can be conducted on a voxel-by-voxel basis allowing the construction of parametric maps. For example, Fig. 2 illustrates an example of k_{ep} 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.[33]
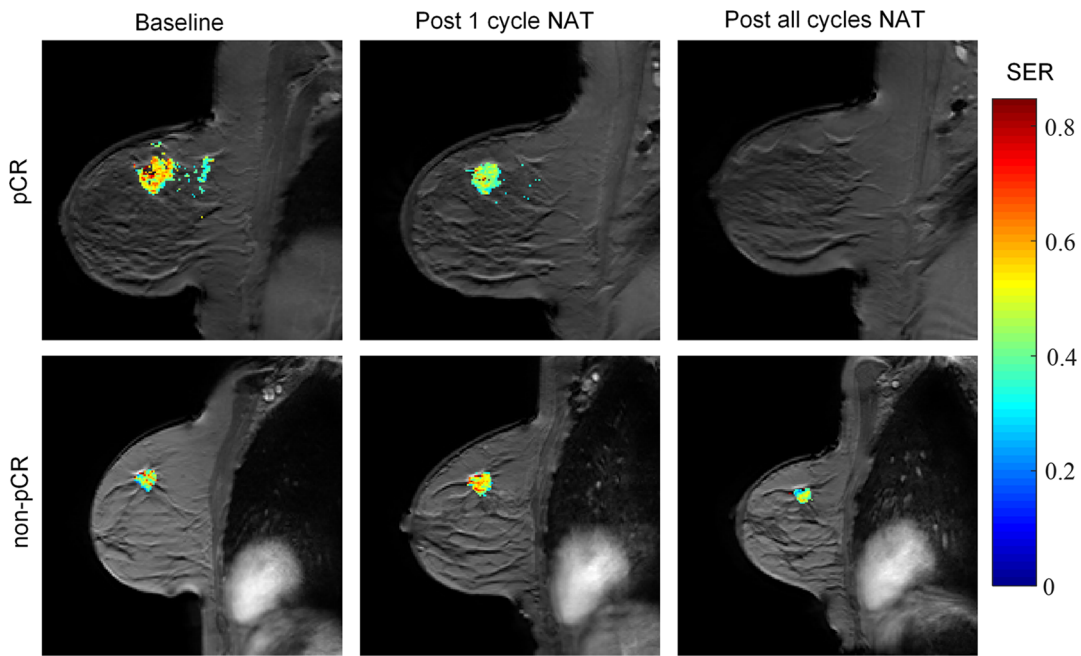
Fig. 1 An example of semiquantitative DCE-MRI of a patient achieving pCR (top row) and a patient not achieving pCR (bottom row) before (first column), after the first cycle (second column), and at conclusion of all NAT (third column). The signal enhancement ratio (SER) overlaid on a high resolution anatomical MRI scan is shown. The high intensity regions-of-interest indicate gross tumor burden. The patient who ultimately achieved pCR demonstrates reduced tumor burden after one cycle of NAT.
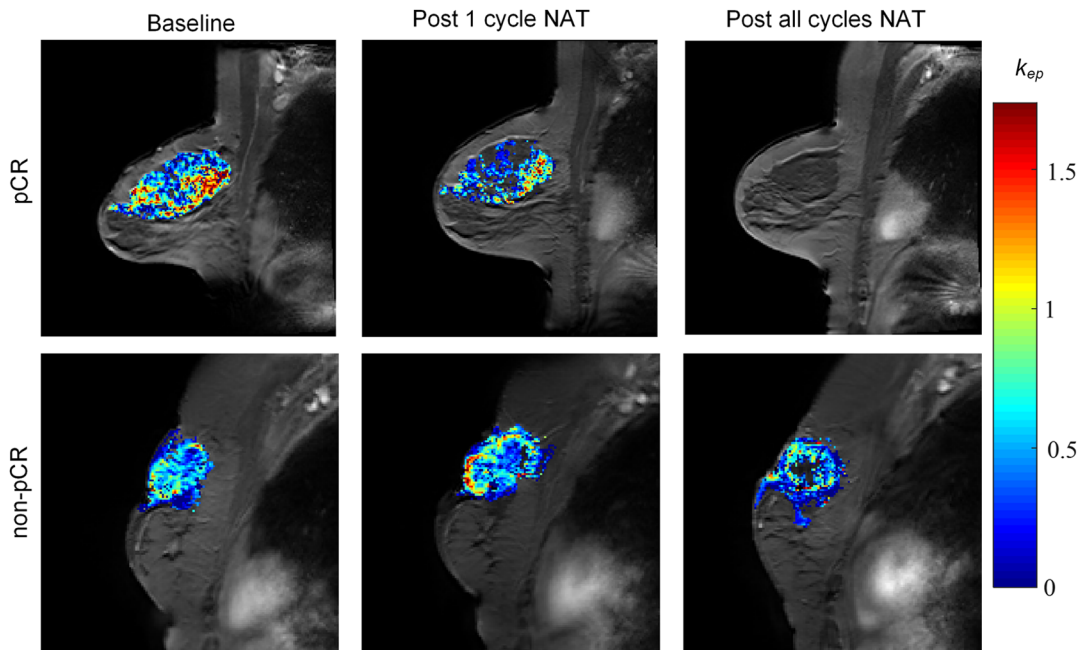

Fig. 2 An example of quantitative analysis of DCE-MRI data of a patient achieving pCR (top row) and a patient not achieving pCR (bottom row) before (first column), after the first cycle (second column), and at conclusion of all NAT (third column). The parametric map of k_{ep} overlaid on a high resolution anatomical MRI scan is shown. Note the diminution of k_{ep} exhibited in the patient achieving pCR after one cycle of NAT, but the enhanced k_{ep} in the patient not achieving pCR.

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.[34] Similar to DCE-MRI parametric mapping, the ADC can be computed
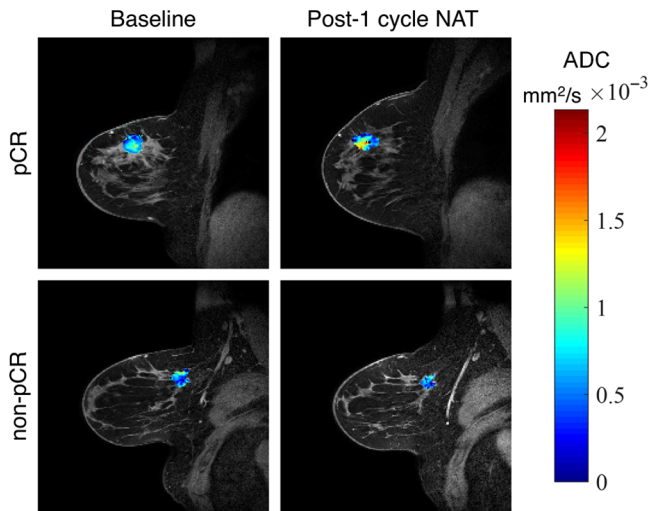
Fig. 3 An example of DW-MRI data from a patient achieving pCR (top row) and a patient not achieving pCR (bottom row). ADC maps from the tumor region are overlaid on a high resolution anatomical image before (left column) and after the first cycle of NAT (right column). The patient who ultimately achieved pCR exhibits increased at ADC early time points within therapy, whereas the patient who did not achieve pCR does not.

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.[35,36] 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.[37] Furthermore, changes in ADC may outperform measurements of tumor size for predicting tumor response to NAT.[38] 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.[34]

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[39] or following the completion of NAT.[40] One study of integrated DCE- and DW-MRI indicated that K^{trans} and ADC are the most sensitive metrics to changes in the tumor occurring between initiation of NAT and surgery,[36] 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)[41] or susceptibility-weighted MRI.[42] 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.[43] 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 operative characteristic curve rather than providing sensitivity and specificity.[44–46] 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. 4. 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,
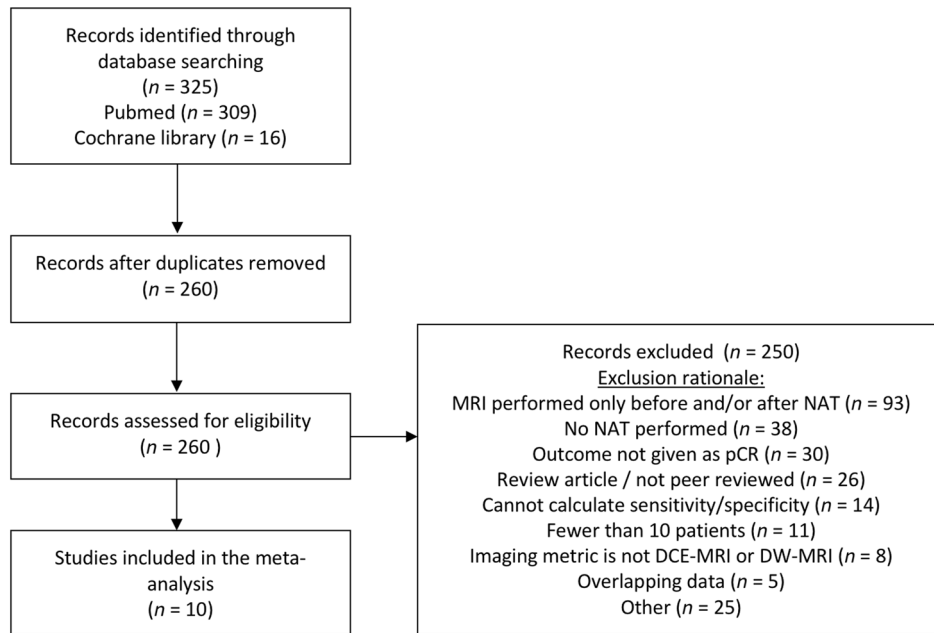
```
┌─────────────────────────┐
│ Records identified through │
│    database searching    │
│        (n = 325)         │
│    Pubmed (n = 309)      │
│  Cochrane library (n = 16) │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Records after duplicates removed │
│        (n = 260)         │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐              ┌──────────────────────────────────────────┐
│ Records assessed for eligibility │────▶│     Records excluded  (n = 250)            │
│        (n = 260 )        │              │        Exclusion rationale:                │
└─────────────────────────┘              │ MRI performed only before and/or after NAT (n = 93) │
            │                            │       No NAT performed (n = 38)            │
            ▼                            │    Outcome not given as pCR (n = 30)       │
┌─────────────────────────┐              │  Review article / not peer reviewed (n = 26) │
│ Studies included in the meta- │        │ Cannot calculate sensitivity/specificity (n = 14) │
│        analysis         │              │      Fewer than 10 patients (n = 11)       │
│        (n = 10)         │              │ Imaging metric is not DCE-MRI or DW-MRI (n = 8) │
└─────────────────────────┘              │       Overlapping data (n = 5)             │
                                         │           Other (n = 25)                   │
                                         └──────────────────────────────────────────┘
```

Fig. 4 Overview of study selection. A total of 325 studies are identified from the Pubmed and Cochrane library databases. After removal of duplicates, 260 records were assessed for eligibility in this meta-analysis. After exclusion of 250 studies for reasons detailed in the flow chart, the remaining 10 studies were analyzed.

b-values for DW-MRI acquisition, and either sensitivity/ specificity or sufficient data to construct a 2×2 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.[52] 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).[53] 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 2×2 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:

$$\text{logit}[\Pr(\text{MRI}_{ij} = 1 | X_{ij})] = (\beta_0 + b_{0i}) + \beta_1 \cdot \text{pCR}_{ij} + \beta_2$$

$$\cdot \text{ER}_i + \beta_3 \cdot \text{PR}_i + \beta_4 \cdot \text{HER2}_i + \beta_5 \cdot \text{Age}_i + \beta_6$$

$$\cdot \text{Temporal Resolution}_i + \beta_7 \cdot \text{MRI Type} + \beta_8$$

$$\cdot \text{NAT cycle} + \beta_9 \cdot \text{pCR}_{ij} \cdot \text{HER2}_i + \beta_{10} \cdot \text{pCR}_{ij} \cdot \text{Age}_i,$$

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; $\text{MRI}_{ij} = 1$ represents an MRI prediction of pCR; and $\text{MRI}_{ij} = 0$ represents an MRI prediction of a non-pCR (i.e., residual tumor remaining at time of surgery after NAT). Similarly, a patient with $\text{pCR}_{ij} = 1$ achieved a pCR and $\text{pCR}_{ij} = 0$ 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

Table 1 Description of studies included in the meta-analysis.

Study	Year	No. of patients	Patient age, mean (range)	Initial clinical stage	Histological subtype	Receptor status	Preoperative therapy regimen	Time between start of NAT and MRI	pCr rate	MRI magnetic field strength	Contrast agent	DCE temporal resolution (s)	DW-MRI b values	Sensitivity	Specificity
Semiquantitative DCE-MRI studies															
Parikh[47]	2014	36	49.8 (24–67)	I–III	35 IDC, 1 ILC	21 ER+, 6 HER2+	3*FEC → 4*T, 4*EC → 4*T HER2+ patients after 2010 received trastuzumab	3 cycles NAT	22.22%	1.5T	Dotarem (20 mL)	92		87.5%	82.1%
Rigter[48]	2013	246	48 (18–68)	I–IV	190 IDC, 33 ILC, 6 IDC/ILC	190 ER+, 246 PR+	3*AC → 3*AC, 3*AC → 3*TCa	3 cycles NAT	1.63%	1.5T	Gadolinium CA (0.1 mmol/kg)	90		75.0%	33.2%
Yang[46]	2012	22	45 (35–67)	II–III	Not reported	Not reported	Antiangiogenic-cytotoxic combination therapy	2 cycles NAT	27.27%	1.5T	Magnevist (0.1 mmol/kg)	70		78.0%	100.0%
Quantitative DCE-MRI studies															
Tateishi[49]	2012	143	57 (43–72)	I–III	131 IDC, 11 ILC	100 ER+, 82 PR+, 111 HER2+	4*FEC → 12*T, 4*AC → 12*T, HER2+ patients received trastuzumab	2 cycles NAT	16.90%	3T	Magnevist (0.1 mmol/kg)	10		51.7%	92.0%
Ah-See[50]	2008	28	46 (29–70)	II–III	21 IDC, 3 ILC	Not reported	6*FEC	2 cycles NAT	39.28%	1.5T	Magnevist (0.1 mmol/kg)	12		94.0%	82.0%
DWI-MRI studies															
Fangberget[51]	2011	31	50.7 (37–72)	Not reported	24 IDC, 7 ILC	21 ER+, 18 PR+, 11 HER2+, 5 TNBC	4*FEC → 2*FEC, 4*FEC-> 2*T, HER2+ patients received trastuzumab	4 cycles NAT	35.48%	1.5T			100, 250, 800	88.0%	80.0%
Minarikova[44]	2017	42	52 (29–74)	I–III	41 ID, 1 ILC	27 ER+, 13 PR+, 5 HER2+	ACT*6, ACT*8, AC → T	2 cycles NAT	16.67%	3.0T			0, 850	66.67%	100%
Multiparametric MRI studies															
Li[39]	2015	37	45 (28–67)	II–III	Not reported	16 ER+, 16 PR+, 12 HER2+, 12 TNBC	4*AC → 4*T, 4*AC → 12*T, HER2+ patients received trastuzumab or other treatments	1 cycle NAT	35.14%	3T	Magnevist (0.1 mmol/kg)	16	0, 500	92.0%	75.0%
Wu[45]	2015	31	48.4 (33–62)	II–III	Not Reported	24 ER+, 23 PR+, 16 HER2+	FEC, FEC → T, T, or other treatments	1 cycle NAT	9.70%	3.0T	Gadovist (0.1 mmol/kg)	40–50	50, 600, 1000	90.9%	83.8%

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.[54] 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.[55] 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 2×2 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 indicated 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

Table 2 Pooled sensitivity, specificity, and diagnostic odds ratios for the 10 studies in this meta-analysis. The unadjusted values were calculated without adjusting for covariates, while the adjusted values were adjusted for covariates according to the logistic regression model. Both sets of values take into account possible heterogeneity between studies via a random effects model. These values, along with the individual sensitivities, specificities, and diagnostic odds ratios for individual studies, are displayed visually in the forest plots in Fig. 6.

	Unadjusted-mean (95% CI)	Adjusted-mean (95% CI)
Sensitivity	0.85 (0.70, 0.93)	0.91 (0.80, 0.96)
Specificity	0.86 (0.72, 0.94)	0.81 (0.68, 0.89)
Diagnostic odds ratio	35 (11, 107)	42.81 (13.67, 134.06)

(99.5%). Pooled, unadjusted estimates of sensitivity and specificity for MRI in predicting pCR are given in Table 2. 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. 5). This figure was created via the hierarchical model from Rutter and Gatsonis.[53] 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
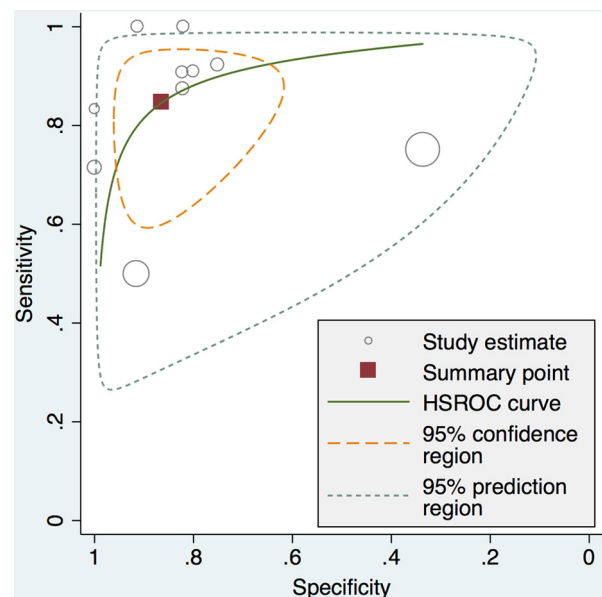

Fig. 5 The HSROC curve displays test performance of individual studies as well as the pooled average estimate. The summary operating point and large area under the curve demonstrate that DCE- and DW-MRI can achieve high sensitivity and specificity for predicting pCR in the neoadjuvant setting for breast 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.
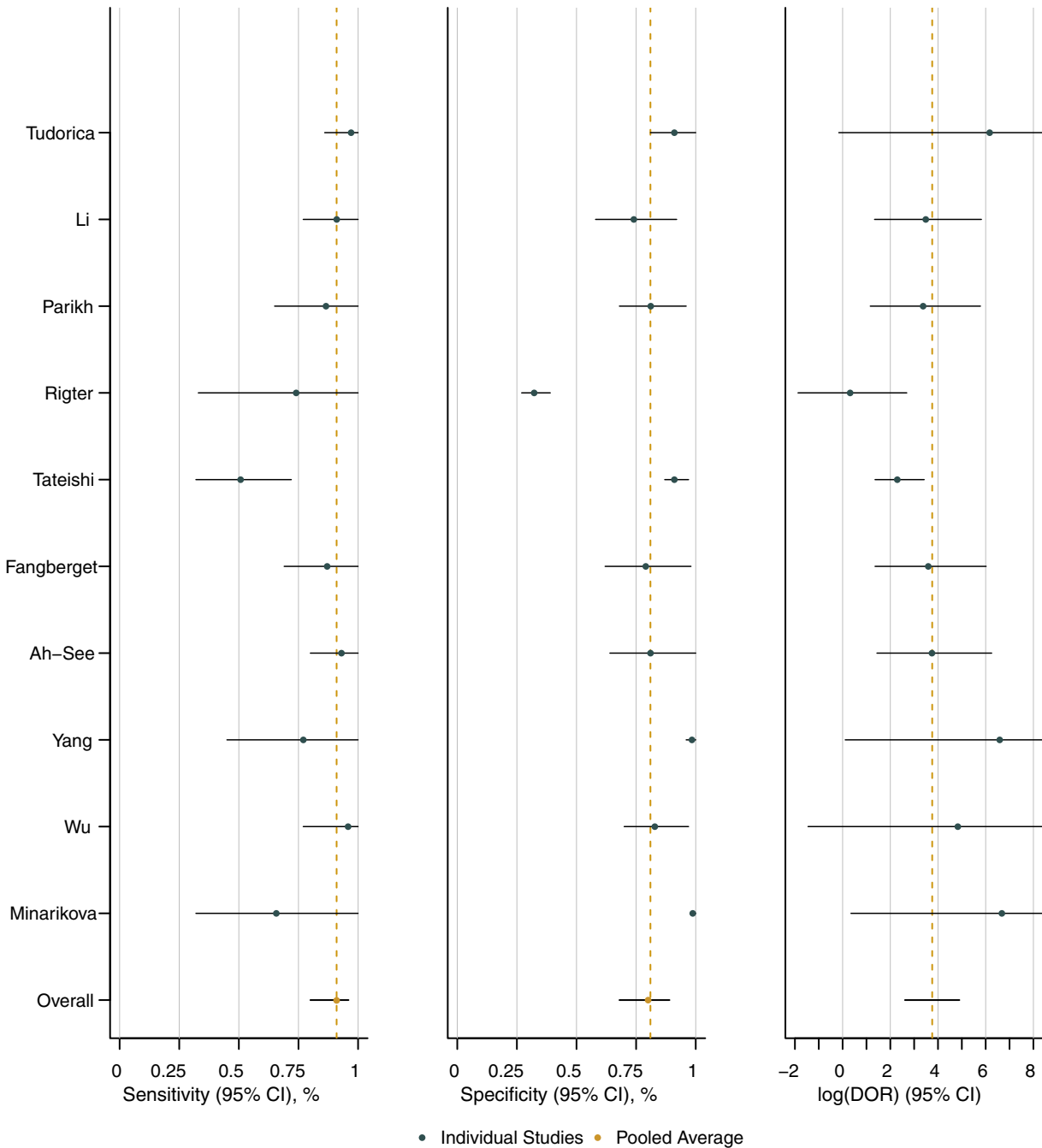
Fig. 6 Individual study estimates for sensitivity, specificity, and log diagnostic odds ratio are displayed in the forest plot along with the overall average adjusted estimates (from the random effects model) and 95% confidence intervals for each. The pooled average demonstrates that DCE- and DW-MRI have high sensitivity, specificity, and log diagnostic odds ratio for predicting response to NAT in breast cancer.

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 average adjusted estimates is given by Fig. 6. 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

Table 3 The parameter estimates as well as standard errors and p-values for the meta-regression analysis are shown. In this analysis, MRI type and MRI time were both statistically significant at a 95% confidence level. Note that, while pCR has a statistically significant coefficient, this parameter is used in the calculation of sensitivity, specificity, and diagnostic odds ratio, rather than having an interpretation of its own.

	Coefficient	Standard error	p-value
Log (ER)	−1.25	17.75	0.945
Log (PR)	0.97	7.53	0.899
Log (HER2)	1.61	9.27	0.863
pCR	8.90	4.40	0.043
pCR*log (HER2)	−1.96	2.04	0.336
Mean (age)	−0.12	0.08	0.130
pCR*mean (age)	−0.09	0.10	0.375
Temporal resolution	0.02	0.08	0.757
MRI type	1.59	0.56	0.005
NAT cycle	1.65	0.54	0.002
Intercept	−3.12	4.29	0.467

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.[30,31,56–60] 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.[61] The correlation between recurrence free survival and pCR is demonstrably improved when stratified by receptor subtype.[62] 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

grouped complete and partial responses.[34,63] 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.[64] The quantitative imaging network (QIN) is attempting to address this shortcoming by standardizing methods for image acquisition, analysis, and sharing.[65] 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.[66,67] 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.[47] Texture analysis of breast cancer DCE-MRI demonstrates differences between responders and nonresponders.[68] 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.[69] Alternately, predictive models incorporating DCE- and DW-MRI to estimate tumor cell proliferation can be used to predict pCR.[70]

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.[71] 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.[72] 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.[57] 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.[73] 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[74] that can be coregistered longitudinally over the course of NAT.[75] A third imaging technique, currently in its early stages, is the application of near-infrared light to perform diffuse optical tomography of the breast.[76] A pilot study of diffuse optical tomography demonstrates that the technique can detect vascular alterations in breast cancer early in the course of NAT.[77] 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.[78]

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.

References

1. A. DeMichele et al., "The neoadjuvant model is still the future for drug development in breast cancer," *Clin. Cancer Res.* **21**(13), 2911–2915 (2015).
2. S. V. Liu et al., "Neoadjuvant therapy for breast cancer," *J. Surg. Oncol.* **101**(4), 283–291 (2010).
3. H. D. Bear et al., "The effect on tumor response of adding sequential preoperative docetaxel to preoperative doxorubicin and cyclophosphamide: preliminary results from national surgical adjuvant breast and bowel project protocol B-27," *J. Clin. Oncol.* **21**(22), 4165–4174 (2003).
4. L. A. Carey et al., "The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes," *Clin. Cancer Res.* **13**(8), 2329–2334 (2007).
5. C. Liedtke et al., "Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer," *J. Clin. Oncol.* **26**(8), 1275–1281 (2008).
6. G. von Minckwitz, "Neoadjuvant chemotherapy in breast cancer-insights from the German experience," *Breast Cancer* **19**(4), 282–288 (2012).

7. A. S. Hamy-Petit et al., "Pathological complete response and prognosis after neoadjuvant chemotherapy for HER2-positive breast cancers before and after trastuzumab era: results from a real-life cohort," *Br. J. Cancer* **114**(1), 44–52 (2016).

8. T. Kanda et al., "Gadolinium-based contrast agent accumulates in the brain even in subjects without severe renal dysfunction: evaluation of autopsy brain specimens with inductively coupled plasma mass spectroscopy," *Radiology* **276**(1), 228–232 (2015).

9. E. A. Eisenhauer et al., "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *Eur. J. Cancer* **45**(2), 228–247 (2009).

10. R. K. Jain et al., "Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies," *J. Clin. Oncol* **30**(21), 2684–2690 (2012).

11. M. Buyse et al., "Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer," *Lancet* **356**(9227), 373–378 (2000).

12. M. Paesmans et al., "Response to chemotherapy has predictive value for further survival of patients with advanced non-small cell lung cancer: 10 years experience of the European Lung Cancer Working Party," *Eur. J. Cancer* **33**(14), 2326–2332 (1997).

13. M. J. Ratain and S. G. Eckhardt, "Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST," *J. Clin. Oncol.* **22**(22), 4442–4445 (2004).

14. R. S. Tuma, "Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints," *J. Natl. Cancer Inst.* **98**(18), 1272–1274 (2006).

15. I. R. Kamel et al., "Unresectable hepatocellular carcinoma: serial early vascular and cellular changes after transarterial chemoembolization as detected with MR imaging," *Radiology* **250**(2), 466–473 (2009).

16. J. D. Wolchok et al., "Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria," *Clin. Cancer Res.* **15**(23), 7412–7420 (2009).

17. R. G. Abramson et al., "Current and emerging quantitative magnetic resonance imaging methods for assessing and predicting the response of breast cancer to neoadjuvant therapy," *Breast Cancer* **2012**(4), 139–154 (2012).

18. Y. Yuan et al., "Accuracy of MRI in prediction of pathologic complete remission in breast cancer after preoperative therapy: a meta-analysis," *Am. J. Roentgenol.* **195**(1), 260–268 (2010).

19. K. M. Leong, P. Lau, and S. Ramadan, "Utilisation of MR spectroscopy and diffusion weighted imaging in predicting and monitoring of breast cancer response to chemotherapy," *J. Med. Imaging Radiat. Oncol.* **59**(3), 268–277 (2015).

20. M. B. Lobbes et al., "The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: a systematic review," *Insights Imaging* **4**(2), 163–175 (2013).

21. L. M. Wu et al., "Can diffusion-weighted MR imaging and contrast-enhanced MR imaging precisely evaluate and predict pathological response to neoadjuvant chemotherapy in patients with breast cancer?" *Breast Cancer Res. Treat.* **135**(1), 17–28 (2012).

22. M. L. Marinovich et al., "Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI," *Breast* **21**(5), 669–677 (2012).

23. T. E. Yankeelov and J. C. Gore, "Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples," *Curr. Med. Imaging Rev.* **3**(2), 91–107 (2007).

24. C. K. Kuhl, H. H. Schild, and N. Morakkabati, "Dynamic bilateral contrast-enhanced MR imaging of the breast: trade-off between spatial and temporal resolution," *Radiology* **236**(3), 789–800 (2005).

25. K. L. Li et al., "Invasive breast cancer: predicting disease recurrence by using high-spatial-resolution signal enhancement ratio imaging," *Radiology* **248**(1), 79–87 (2008).

26. R. G. Abramson et al., "Early assessment of breast cancer response to neoadjuvant chemotherapy by semi-quantitative analysis of high-temporal resolution DCE-MRI: preliminary results," *Magn. Reson. Imaging* **31**(9), 1457–1464 (2013).

27. C. E. Loo et al., "Dynamic contrast-enhanced MRI for prediction of breast cancer response to neoadjuvant chemotherapy: initial results," *Am. J. Roentgenol.* **191**(5), 1331–1338 (2008).

28. N. M. Hylton et al., "Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy—results from ACRIN 6657/I-SPY TRIAL," *Radiology* **263**(3), 663–672 (2012).

29. D. K. Woolf et al., "Assessing response in breast cancer with dynamic contrast-enhanced magnetic resonance imaging: are signal intensity-time curves adequate?" *Breast Cancer Res. Treat.* **147**(2), 335–343 (2014).

30. M. D. Pickles et al., "Prognostic value of DCE-MRI in breast cancer patients undergoing neoadjuvant chemotherapy: a comparison with traditional survival indicators," *Eur. Radiol.* **25**(4), 1097–1106 (2015).

31. N. M. Hylton et al., "Neoadjuvant chemotherapy for breast cancer: functional tumor volume by MR imaging predicts recurrence-free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL," *Radiology* **279**(1), 44–55 (2016).

32. P. S. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging," *J. Magn. Reson. Imaging* **7**(1), 91–101 (1997).

33. A. Tudorica et al., "Early prediction and evaluation of breast cancer response to neoadjuvant chemotherapy using quantitative DCE-MRI," *Transl. Oncol.* **9**(1), 8–17 (2016).

34. C. J. Galban et al., "Multi-site clinical evaluation of DW-MRI as a treatment response metric for breast cancer patients undergoing neoadjuvant chemotherapy," *PLoS One* **10**(3), e0122151 (2015).

35. M. D. Pickles et al, "Diffusion changes precede size reduction in neoadjuvant treatment of breast cancer," *Magn. Reson. Imaging* **24**(7), 843–847 (2006).

36. T. E. Yankeelov et al., "Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results," *Magn. Reson. Imaging* **25**(1), 1–13 (2007).

37. X. R. Li et al., "DW-MRI ADC values can predict treatment response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy," *Med. Oncol.* **29**(2), 425–431 (2012).

38. U. Sharma et al., "Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy," *NMR Biomed.* **22**(1), 104–113 (2009).

39. X. Li et al., "Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer," *Invest. Radiol.* **50**(4), 195–204 (2015).

40. S. Y. Hahn et al., "Role of diffusion-weighted imaging as an adjunct to contrast-enhanced breast MRI in evaluating residual breast cancer following neoadjuvant chemotherapy," *Eur. J. Radiol.* **83**(2), 283–288 (2014).

41. 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," *Invest. Radiol.* **49**(6), 421–430 (2014).

42. E. A. O'Flynn et al., "Multi-parametric MRI in the early prediction of response to neo-adjuvant chemotherapy in breast cancer: value of non-modelled parameters," *Eur. J. Radiol.* **85**(4), 837–842 (2016).

43. D. Moher et al., "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS Med.* **6**(7), e1000097 (2009).

44. L. Minarikova et al., "Investigating the prediction value of multiparametric magnetic resonance imaging at 3 T in response to neoadjuvant chemotherapy in breast cancer," *Eur. Radiol.* **27**(5), 1901–1911 (2017).

45. L. A. Wu et al., "Evaluation of the treatment response to neoadjuvant chemotherapy in locally advanced breast cancer using combined magnetic resonance vascular maps and apparent diffusion coefficient," *J. Magn. Reson. Imaging* **42**(5), 1407–1420 (2015).

46. X. Yang et al., "Microcirculatory fraction (MCF$_f$) as a potential imaging marker for tumor heterogeneity in breast cancer," *Magn. Reson. Imaging* **30**(8), 1059–1067 (2012).

47. J. Parikh et al., "Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy," *Radiology* **272**(1), 100–112 (2014).

48. L. S. Rigter et al., "Neoadjuvant chemotherapy adaptation and serial MRI response monitoring in ER-positive HER2-negative breast cancer," *Br. J. Cancer* **109**(12), 2965–2972 (2013).

49. U. Tateishi et al., "Neoadjuvant chemotherapy in breast cancer: prediction of pathologic response with PET/CT and dynamic contrast-enhanced

MR imaging—prospective assessment," *Radiology* **263**(1), 53–63 (2012).

50. M. L. Ah-See et al., "Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer," *Clin. Cancer Res.* **14**(20), 6580–6589 (2008).

51. A. Fangberget et al., "Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging," *Eur. Radiol.* **21**(6), 1188–1199 (2011).

52. P. Whiting et al., "The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews," *BMC Med. Res. Methodol.* **3**, 25 (2003).

53. C. M. Rutter and C. A. Gatsonis, "A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations," *Stat. Med.* **20**(19), 2865–2884 (2001).

54. S. A. Deppen et al., "Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease a meta-analysis," *J. Am. Med. Assoc.* **312**(12), 1227–1236 (2014).

55. F. E. Harrell, *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer, New York (2001).

56. E. F. Jones et al., "MRI enhancement in stromal tissue surrounding breast tumors: association with recurrence free survival following neoadjuvant chemotherapy," *PLoS One* **8**(5), e61969 (2013).

57. I. Lim et al., "The combination of FDG PET and dynamic contrast-enhanced MRI improves the prediction of disease-free survival in patients with advanced breast cancer after the first cycle of neoadjuvant chemotherapy," *Eur. J. Nucl. Med. Mol. Imaging* **41**(10), 1852–1860 (2014).

58. W. C. Lo et al., "Effect of imaging parameter thresholds on MRI prediction of neoadjuvant chemotherapy response in breast cancer subtypes," *PLoS One* **11**(2), e0142047 (2016).

59. D. C. Newitt et al., "Real-time measurement of functional tumor volume by MRI to assess treatment response in breast cancer neoadjuvant clinical trials: validation of the aegis SER software platform," *Transl. Oncol.* **7**(1), 94–100 (2014).

60. V. Y. Park et al., "Breast parenchymal signal enhancement ratio at preoperative magnetic resonance imaging: association with early recurrence in triple-negative breast cancer patients," *Acta Radiol.* **57**(7), 802–808 (2015).

61. P. Cortazar et al., "Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis," *Lancet* **384**(9938), 164–172 (2014).

62. L. J. Esserman et al., "Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL—CALGB 150007/150012, ACRIN 6657," *J. Clin. Oncol.* **30**(26), 3242–3249 (2012).

63. M. M. Nadrljanski et al., "MRI in the evaluation of breast cancer patient response to neoadjuvant chemotherapy: predictive factors for breast conservative surgery," *Diagn. Interventional Radiol.* **19**(6), 463–470 (2013).

64. T. E. Yankeelov et al., "Quantitative imaging in cancer clinical trials," *Clin. Cancer Res.* **22**(2), 284–290 (2016).

65. J. M. Mountz et al., "Letter to cancer center directors: predict and/or measure tumor response in cancer therapy trials," *J. Clin. Oncol.* **32**(19), 2115–2116 (2014).

66. B. Morgan et al., "A simple, reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging," *Br. J. Cancer* **94**(10), 1420–1427 (2006).

67. A. R. Padhani et al., "Reproducibility of quantitative dynamic MRI of normal human tissues," *NMR Biomed.* **15**(2), 143–153 (2002).

68. A. Ahmed et al., "Texture analysis in assessment and prediction of chemotherapy response in breast cancer," *J. Magn. Reson. Imaging* **38**(1), 89–101 (2013).

69. J. A. Weis et al., "Predicting the response of breast cancer to neoadjuvant therapy using a mechanically coupled reaction-diffusion model," *Cancer Res.* **75**(22), 4697–4707 (2015).

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," *Transl. Oncol.* **6**(3), 256–264 (2013).

71. H. J. Shin et al., "Prediction of pathologic response to neoadjuvant chemotherapy in patients with breast cancer using diffusion-weighted imaging and MRS," *NMR Biomed.* **25**(12), 1349–1359 (2012).

72. D. J. Manton et al., "Neoadjuvant chemotherapy in breast cancer: early response prediction with quantitative MR imaging and spectroscopy," *Br. J. Cancer* **94**(3), 427–435 (2006).

73. S. H. Park et al., "Comparison of diffusion-weighted MR imaging and FDG PET/CT to predict pathological complete response to neoadjuvant chemotherapy in patients with breast cancer," *Eur. Radiol.* **22**(1), 18–25 (2012).

74. T. E. Yankeelov, R. G. Abramson, and C. C. Quarles, "Quantitative multimodality imaging in cancer research and therapy," *Nat. Rev. Clin. Oncol.* **11**(11), 670–680 (2014).

75. N. C. Atuegwu et al., "Longitudinal, intermodality registration of quantitative breast PET and MRI data acquired before and during neoadjuvant chemotherapy: preliminary results," *Med. Phys.* **41**(5), 052302 (2014).

76. B. J. Tromberg et al., "Assessing the future of diffuse optical imaging technologies for breast cancer management," *Med. Phys.* **35**(6), 2443–2451 (2008).

77. Q. Zhu et al., "Noninvasive monitoring of breast cancer during neoadjuvant chemotherapy using optical tomography with ultrasound localization," *Neoplasia* **10**(10), 1028–1040 (2008).

78. B. E. Schaafsma et al., "Optical mammography using diffuse optical spectroscopy for monitoring tumor response to neoadjuvant chemotherapy in women with locally advanced breast cancer," *Clin. Cancer Res.* **21**(3), 577–584 (2015).

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 multi-modal 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 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

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.