Impact of MRI Protocol Adherence on Prediction of Pathological Complete Response in the I-SPY 2 Neoadjuvant Breast Cancer Trial

Natsuko Onishi[1], Wen Li[1], Jessica Gibbs[1], Lisa J. Wilmes[1], Alex Nguyen[1], Ella F. Jones[1], Vignesh Arasu[1], John Kornak[2], Bonnie N. Joe[1], Laura J. Esserman[3], David C. Newitt[1], and Nola M. Hylton[1]

Departments of [1]Radiology & Biomedical Imaging; [2]Epidemiology and Biostatistics; and [3]Surgery, University of California, San Francisco, San Francisco, CA

Corresponding Author:
Nola Hylton, PhD
1600 Divisadero St. Room C255, San Francisco, CA 94115;
E-mail: Nola.Hylton@ucsf.edu

We investigated the impact of magnetic resonance imaging (MRI) protocol adherence on the ability of functional tumor volume (FTV), a quantitative measure of tumor burden measured from dynamic contrast-enhanced MRI, to predict response to neoadjuvant chemotherapy. We retrospectively reviewed dynamic contrast-enhanced breast MRIs for 990 patients enrolled in the multicenter I-SPY 2 TRIAL. During neoadjuvant chemotherapy, each patient had 4 MRI visits (pretreatment [T0], early-treatment [T1], inter-regimen [T2], and presurgery [T3]). Protocol adherence was rated for 7 image quality factors at T0–T2. Image quality factors confirmed by DICOM header (acquisition duration, early phase timing, field of view, and spatial resolution) were adherent if the scan parameters followed the standardized imaging protocol, and changes from T0 for a single patient's visits were limited to defined ranges. Other image quality factors (contralateral image quality, patient motion, and contrast administration error) were considered adherent if imaging issues were absent or minimal. The area under the receiver operating characteristic curve (AUC) was used to measure the performance of FTV change (percent change of FTV from T0 to T1 and T2) in predicting pathological complete response. FTV changes with adherent image quality in all factors had higher estimated AUC than those with non-adherent image quality, although the differences did not reach statistical significance (T1, 0.71 vs. 0.66; T2, 0.72 vs. 0.68). These data highlight the importance of MRI protocol adherence to predefined scan parameters and the impact of data quality on the predictive performance of FTV in the breast cancer neoadjuvant setting.

INTRODUCTION

Neoadjuvant chemotherapy (NAC) plays an important role in the treatment of locally advanced breast cancer (1, 2). NAC enables down-sizing of the tumor to allow for breast conserving surgery and shows equivalent disease-free and overall survival when compared with adjuvant therapy (2). Breast magnetic resonance imaging (MRI) during NAC facilitates the evaluation of disease extent in the breast and the monitoring of tumor response to systemic therapy (3–8). Functional tumor volume (FTV) is a quantitative measure of tumor burden derived from dynamic contrast-enhanced (DCE) breast MRI. The I-SPY 1 (ACRIN 6657) TRIAL showed strong associations between FTV and pathological complete response (pCR) and recurrence-free survival after NAC (6, 7).

The ability of FTV to predict pCR has the potential to advance personalized medicine by informing treatment redirection or deescalation based on patient response during NAC.

The I-SPY 2 TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response through Imaging and Molecular Analysis 2) is an ongoing phase II trial with adaptive randomization to multiple therapy arms. The aim of I-SPY 2 is to assess the efficacy of experimental agents alone or in combination with standard NAC in patients with breast cancer who are at a high risk for early recurrence (9–11). Patients have 4 MRI visits during neoadjuvant treatment, and FTV derived from each MRI examination is used to adjust patient randomization and estimate predictive probabilities for pCR. This multicenter clinical trial

includes a total of 25 participating sites and >1600 patients randomized to treatment over 9 years. While there is a standardized MRI protocol distributed to all participating sites with instruction to use the same scan parameters for all sequential MRI visits for a single patient, adherence to protocol-specified scan parameters can be logistically and technically challenging given different magnetic resonance scanner vendors, models, configurations, and sequences across all participating sites and all sequential visits.

This study investigated the impact of MRI protocol adherence on the ability of FTV to predict pCR in the I-SPY 2 TRIAL.

METHODOLOGY

Study Cohort

This retrospective study is based on the review of MRI data from 990 patients with breast cancer randomized to 1 of 9 experimental drug arms completed by November 2016 or to a control arm (standard of care) in the I-SPY 2 TRIAL. We conducted this study in compliance with the Health Insurance Portability and Accountability Act. All participating sites received approval from their institutional review board, and all patients provided written informed consent to participate in the I-SPY 2 TRIAL.

Eligibility to enroll in I-SPY 2 was as follows: 1) women 18 years or older 2) diagnosed with locally advanced breast cancer (tumor size ≥ 2.5 cm) without distant metastasis. Eligible patients were classified according to breast cancer subtype, defined by their hormone (estrogen/progesterone) receptor (HR), human epidermal growth factor receptor 2 (HER2) status, and MammaPrint 70-gene signature (MammaPrint, Agendia, Amsterdam, The Netherlands). Patients with HR+/HER2− tumors that were assessed as low risk by the MammaPrint 70-gene signature were screened out from the trial. As neoadjuvant therapy, participants received 12 cycles of weekly paclitaxel alone or in combination with 1 of the 9 experimental agents, followed by 4 cycles of anthracycline–cyclophosphamide prior to surgery. Patients with HER2+ cancer also received trastuzumab. Each patient had 4 MRI examination visits: pretreatment (T0), early-treatment (3 weeks after treatment initiation, T1), inter-regimens (T2), and presurgery (T3). Figure 1 shows the study schema.

MRI Acquisition Protocol

Each site performed MRI examinations including DCE series by using a bilateral 3-dimensional, T1-weighted sequence with fat suppression on a 3.0 T or 1.5 T MRI scanner with dedicated breast coil. The standardized acquisition protocol in the ISPY-2 TRIAL, distributed to sites prospectively before study initiation, was as follows: repetition time of 4–10 milliseconds with minimum echo time; flip angle = 10–20°; field of view (FOV) = 26–36 cm; acquired frequency or read matrix = 384–512; acquired phase encoding matrix ≥ 256; in-plane resolution ≤ 1.4×1.4 mm; thickness ≤ 2.5 mm; temporal resolution = 80–100 second; axial orientation; and prone position. The standardized contrast injection rate was 2 cc/s with a 20-cc saline flush. DCE MRI was performed once before and multiple times after contrast injection using identical sequences, with scanning to continue for at least 8 minutes after contrast injection. Early and delayed postcontrast phases were selected from the postcontrast series at the time of analysis based on temporal sampling of the center of k-space

closest to 2 minutes 30 seconds for early phase and 7 minutes 30 seconds for late phase. All sites submitted test cases showing their ability to comply with the standardized imaging protocol before study accrual began.

Protocol Adherence of Image Quality Factors

We defined image quality factors to investigate the impact of protocol adherence on the prediction of pCR. Image quality factors were categorized as confirmable or not confirmable based on whether adherence could be confirmed by metadata stored in the DICOM header.

DICOM confirmable factors (DC factors) included acquisition duration, early phase timing, FOV, and spatial resolution (in-plane resolution and slice thickness). Each DC factor was defined as adherent if it fulfilled the standardized acquisition protocol described above and remained within the defined ranges for change from baseline: acquisition duration ≤3-second change; early phase timing ≤5-second change; FOV ≤50-mm change; in-plane resolution ≤10% change; slice thickness ≤10% change (Table 1). The allowable ranges for change were not included in the standardized acquisition protocol, and thus these were defined based on the observation of the data distribution and clinical perspectives in this study.

Not DICOM-confirmable factors (nDC factors) included contralateral image quality (signal homogeneity, adequacy of fat suppression, presence of signal flare near breast coil or other artifacts), patient motion during scanning, and contrast administration errors (off-protocol saline flush volume or injection rate). The contralateral image quality was visually assessed and ranked by consensus of 3 trained readers in the I-SPY 2 Imaging Core Lab as part of the measurement of background parenchymal enhancement in a separate study (12). Readers reviewed 3 axial slices selected using an automated computer algorithm corresponding to the first, third, and fifth of the central 5 slices of the whole breast. Image quality of each examination was ranked using a 3-point scoring system (2, good; 1, adequate; 0, poor). The contralateral image quality was defined as adherent if the score was 2 or 1 and non-adherent if the score was 0. Patient motion was noted during the FTV calculation process, and cases with minimal or no motion were defined as adherent. Contrast administration errors were documented on case report forms completed by sites when the examinations were submitted to the Core Lab for FTV analysis. Examinations without reported errors were defined as adherent.

Functional Tumor Volume (FTV)

In this study, we focused on FTV at T0, T1, and T2 and disregarded T3, because prediction of pCR at T3, the presurgery time point after completion of NAC, does not influence treatment redirection during NAC. Using in-house software developed with IDL (Exelis Visual Information Solutions, Boulder, CO), FTV was calculated by summing all voxels with early percent enhancement (~2.5 minutes post contrast injection) above 70% and signal enhancement ratio above zero within a manually delineated 3-dimensional region of interest as described in the literature (13). We defined FTV at T0, T1, and T2 as FTV0, FTV1, and FTV2, respectively. Percent change of FTV from baseline (T0, pretreatment) to T1 and T2 was defined as %ΔFTV0_1 and %ΔFTV0_2,
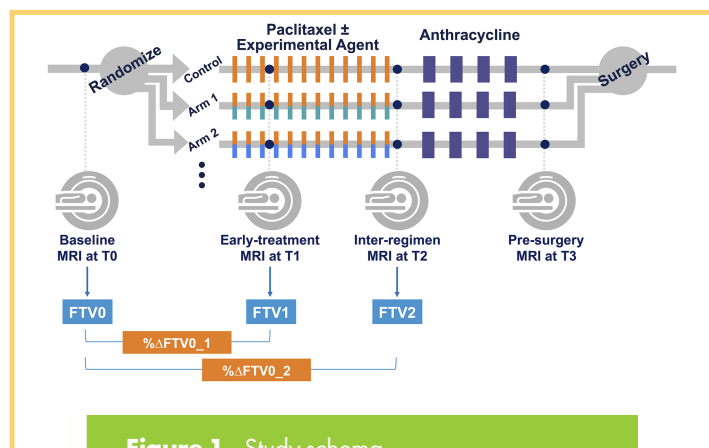
Figure 1. Study schema.

respectively. FTV changes (%ΔFTV0_1 and %ΔFTV0_2) were then stratified into adherent or non-adherernt subsets. For FTV change to be stratified as adherent at a given time point, both examinations had to be adherent. For example, for %ΔFTV0_1 to be adherent, FTV0 and FTV1 both had to be adherent. This stratification was done for each image quality factor (DICOM-confirmable and not DICOM-confirmable) at each treatment time point. FTV changes (%ΔFTV0_1 and %ΔFTV0_2) that were adherent for all image quality factors were stratified as adherent in "combined factors."

Pathological Outcome

In the I-SPY 2 TRIAL, pCR was defined as the absence of residual invasive disease in the breast and axillary lymph nodes on surgical specimen after NAC.

Statistical Analysis

Statistical analyses were performed using R Version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) along with the R packages "pROC" (14) and "boot" (15, 16). P values <.05 were considered statistically significant.

The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the performance of FTV change for predicting pCR. For each image quality factor, the AUC was estimated independently for adherent and nonadherent FTV change subsets at early treatment (T1) and inter-regimen (T2) time points. In addition, associated 95% confidence intervals (CIs) for the AUCs were calculated based on 2000 bootstrap

replications. For the statistical comparison of the AUCs between the adherent and nonadherent subsets for each image quality factor and each time point, the 95% CI and P value for the difference between 2 AUCs (AUC of the adherent subset minus AUC of the nonadherent subset) were estimated based on 2000 bootstrap replications.

RESULTS

Patient Characteristics

Patient characteristics including age, menopausal status, race, molecular subtype, assigned chemotherapy, and treatment response for the 990 patients are listed in Table 2. The mean age was 48.8 years in the range of 23–77 years except for 1 patient with no record available. The whole cohort included 464 patients (47%) with premenopausal status, 33 patients (3%) with perimenopausal status, 291 patients (29%) with postmenopausal status, and 202 patients (20%) with no record available. The cohort comprised of 784 (79%) white, 121 (12%) African American, 68 (7%) Asian, 5 (1%) Native Hawaiian or Pacific Islander, 4 (0.4%) American Indian or Alaska Native, 7 (1%) patients with mixed race, and 1 (0.1%) patient with no record available. Of the 990 breast cancers observed in this study, 380 (38%) were HR+ HER2− subtype, 156 (16%) were HR+ HER2+ subtype, 89 (9%) were HR− HER2+ subtype, 363 (37%) were HR− HER2−, and 2 (0.2%) had no record available. Standard NAC was assigned for 211 (21%), experimental drug arms were assigned to 777 (78%) and no record was available for 2 (0.2%). In total, 324 (33%) of the 990 patients achieved pCR.

Data Set

Of a possible total of 2970 MRI exams ($3 exams \times 990 patients$), 113 exams were not completed or were rejected by the I-SPY2 TRIAL owing to errors in case report forms (n = 1), patient withdrawal of treatment consent (n = 43), patient illness (n = 13), MRI scanner or contrast injection issues (n = 4), early surgery after early discontinuation of NAC (n = 3), or missed patient appointments for unknown reasons (n = 49). In addition, 18 exams were excluded by the I-SPY 2 Imaging Core Lab owing to severe imaging issues such as inconsistent scanner being used between visits, patient scanned in the sagittal or coronal direction, repositioning after contrast injection, and corrupted data. In total, 2839 MRI examinations (T0, 989; T1, 952; T2, 898) were analyzed, and FTV and percent changes of FTV were

Table 1. MRI Protocol Adherence Criteria for DICOM-Confirmable Factor

DICOM-Confirmable Factor	Protocol	Allowable Change from Baseline (T0)
Acquisition Duration	80–100 seconds	≤3-second change
Early Phase Timing	120–150 seconds post contrast injection	≤5-second change
Field of View	260–360 mm	≤50-mm change
In-Plane Resolution	≤1.4 mm	≤10% change
Slice Thickness	≤2.5 mm	≤10% change

Table 2. Patient Characteristics

Patient Characteristics	Total, n = 990
Age	
Mean ± SD	48.8 ± 10.6 y
Range	23–77 y
Menopausal Status	
Premenopause	464 (47)
Perimenopause	33 (3)
Postmenopause	291 (29)
Not available	202 (20)
Race	
White	784 (79)
African American	121 (12)
Asian	68 (7)
Native Hawaiian or Pacific Islander	5 (1)
American Indian or Alaska Native	4 (0.4)
Mixed Race	7 (1)
Not Available	1 (0.1)
Molecular Subtype	
HR+, HER2−	380 (38)
HR+, HER2+	156 (16)
HR−, HER2+	89 (9)
HR−, HER2−	363 (37)
Not available	2 (0.2)
Assigned Chemotherapy	
Standard Neoadjuvant Chemotherapy	211 (21)
Experimental Drug Arms	777 (78)
Not Available	2 (0.2)
Treatment Response	
pCR	324 (33)
Non-pCR	666 (67)

Unless otherwise specified, data represent the number of patients and data in parentheses are percentages.
Abbreviations: SD, standard deviation; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; pCR, pathological complete response.
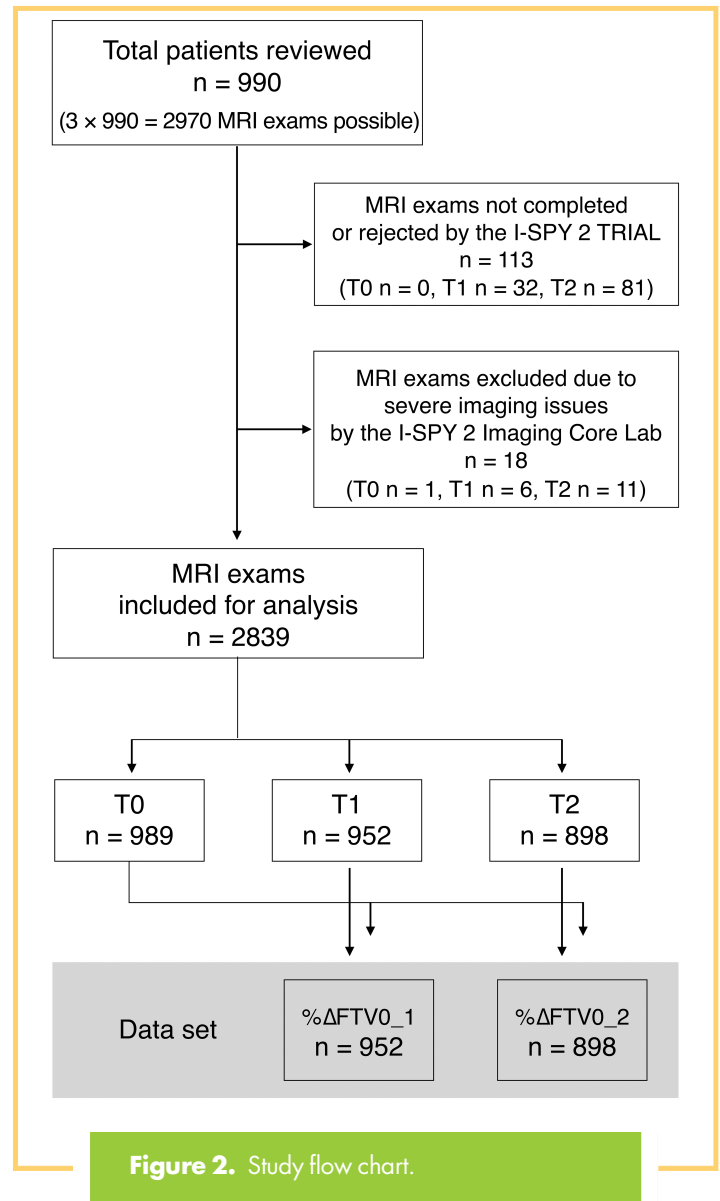

Figure 2. Study flow chart.

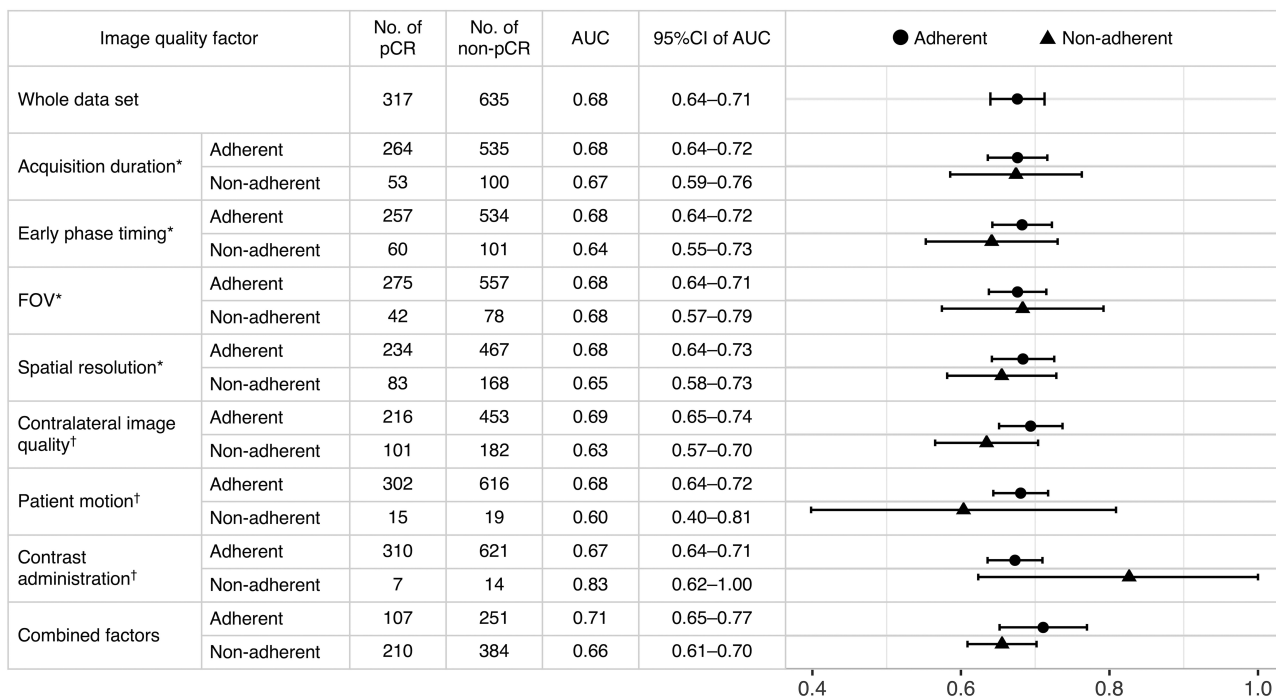calculated: %ΔFTV0_1, 952; %ΔFTV0_2, 898. Figure 2 shows the study flow chart.

Predictive Performance of pCR and Image Quality Factors

In the whole data set, FTV change (%ΔFTV0_1 and %ΔFTV0_2) led to estimated AUC values of 0.68 and 0.69, respectively (Figure 3, A and B). For each image quality factor, the AUC and associated 95% CI for adherent and non-adherent subsets of FTV change at T1 and T2 are shown with the number of pCR or non-pCR patients in Figure 3A and Figure 3B, respectively. Table 3 shows the 95% CI of the AUC difference between adherent and non-adherent subsets and the associated *P* value based on the bootstrap test.

DICOM-Confirmable Factors (DC Factors). FTV changes with adherent image quality factors tended to have higher estimated AUC values than those with non-adherent image quality factors, although the difference between the AUCs did not reach statistical significance (Figure 3 and Table 3): AUC values for adherent vs. non-adherent subsets at T1 (acquisition duration, 0.68 vs 0.67; early phase timing, 0.68 vs 0.64; spatial resolution, 0.68 vs 0.65) and at T2 (acquisition duration, 0.70 vs 0.66; early phase timing, 0.70 vs 0.66; FOV, 0.70 vs 0.64; spatial resolution, 0.70 vs 0.68).

Not DICOM-Confirmable Factors (nDC Factors). In terms of contralateral image quality, FTV changes with adherent image quality factor had higher estimated AUC values than those with non-adherent image quality factor, although the difference between the AUCs did not reach statistical significance (Figure 3 and Table 3): AUC values for adherent vs non-adherent subsets were 0.69 vs 0.63 at T1 and 0.70 vs 0.67 at T2. In terms of other nDC factors (patient motion and contrast administration), the non-adherent subsets had a relatively small sample size (T1

A FTV change at T1 (%ΔFTV0_1)

Image quality factor		No. of pCR	No. of non-pCR	AUC	95%CI of AUC	● Adherent ▲ Non-adherent
Whole data set		317	635	0.68	0.64–0.71	
Acquisition duration*	Adherent	264	535	0.68	0.64–0.72	
	Non-adherent	53	100	0.67	0.59–0.76	
Early phase timing*	Adherent	257	534	0.68	0.64–0.72	
	Non-adherent	60	101	0.64	0.55–0.73	
FOV*	Adherent	275	557	0.68	0.64–0.71	
	Non-adherent	42	78	0.68	0.57–0.79	
Spatial resolution*	Adherent	234	467	0.68	0.64–0.73	
	Non-adherent	83	168	0.65	0.58–0.73	
Contralateral image quality†	Adherent	216	453	0.69	0.65–0.74	
	Non-adherent	101	182	0.63	0.57–0.70	
Patient motion†	Adherent	302	616	0.68	0.64–0.72	
	Non-adherent	15	19	0.60	0.40–0.81	
Contrast administration†	Adherent	310	621	0.67	0.64–0.71	
	Non-adherent	7	14	0.83	0.62–1.00	
Combined factors	Adherent	107	251	0.71	0.65–0.77	
	Non-adherent	210	384	0.66	0.61–0.70	

0.4   0.6   0.8   1.0

B FTV change at T2 (%ΔFTV0_2)

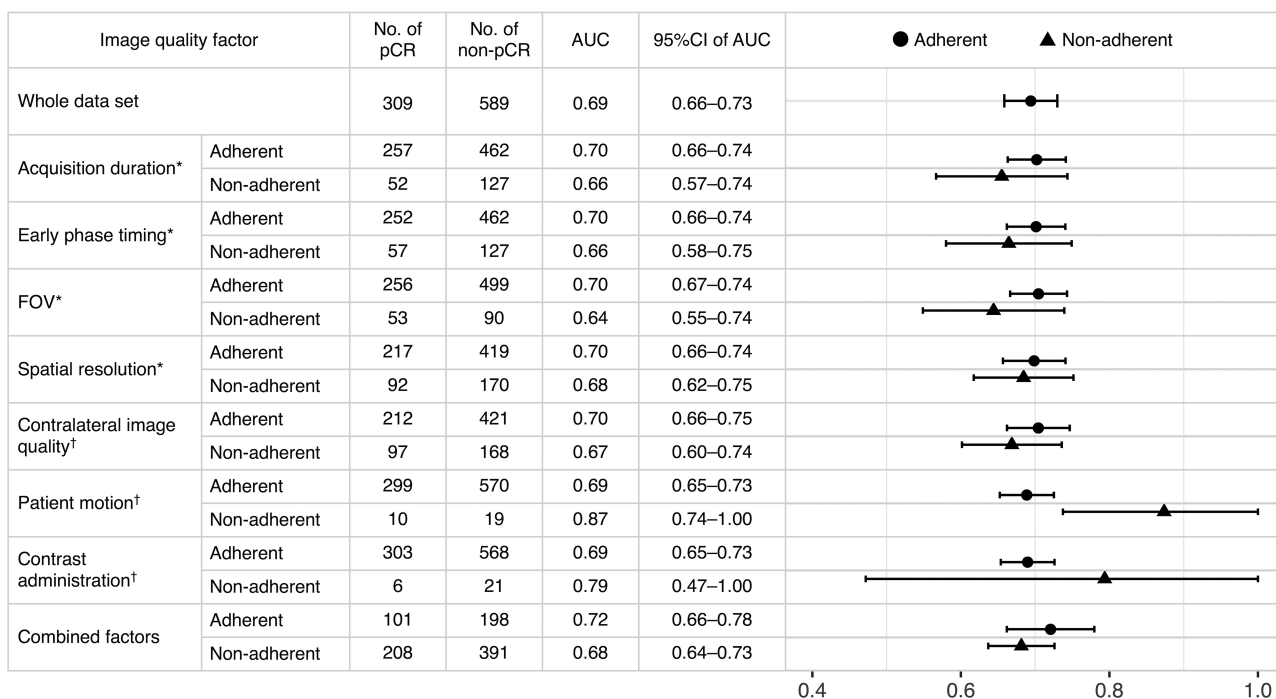| Image quality factor		No. of pCR	No. of non-pCR	AUC	95%CI of AUC	● Adherent ▲ Non-adherent
Whole data set		309	589	0.69	0.66–0.73	
Acquisition duration*	Adherent	257	462	0.70	0.66–0.74	
	Non-adherent	52	127	0.66	0.57–0.74	
Early phase timing*	Adherent	252	462	0.70	0.66–0.74	
	Non-adherent	57	127	0.66	0.58–0.75	
FOV*	Adherent	256	499	0.70	0.67–0.74	
	Non-adherent	53	90	0.64	0.55–0.74	
Spatial resolution*	Adherent	217	419	0.70	0.66–0.74	
	Non-adherent	92	170	0.68	0.62–0.75	
Contralateral image quality†	Adherent	212	421	0.70	0.66–0.75	
	Non-adherent	97	168	0.67	0.60–0.74	
Patient motion†	Adherent	299	570	0.69	0.65–0.73	
	Non-adherent	10	19	0.87	0.74–1.00	
Contrast administration†	Adherent	303	568	0.69	0.65–0.73	
	Non-adherent	6	21	0.79	0.47–1.00	
Combined factors	Adherent	101	198	0.72	0.66–0.78	
	Non-adherent	208	391	0.68	0.64–0.73	

0.4   0.6   0.8   1.0

Figure 3. Predictive performance of pCR and image quality factors. Area under the receiver operating characteristic curves (AUCs) and associated 95% confidence intervals (CI) for adherent and non-adherent functional tumor volume (FTV) change subsets at (A) T1 and (B) T2 are shown. Data for the whole data set are shown in the top followed by data for subsets for each image quality factor (DICOM-confirmable factors*, not DICOM-confirmable factors†, and combined factors). In the plot column, dots, or triangles show the estimated AUCs, and bars show the range of 95% CIs for adherent and non-adherent subsets.

Table 3. Difference of AUCs between Adherent and Nonadherent FTV Change Subsets

Image Quality Factor	Difference of AUCs for FTV Change at T1			Difference of AUCs for FTV Change at T2		
	Estimate	95%CI (LL, UL)	P Value	Estimate	95%CI (LL, UL)	P Value
Acquisition Duration	0.00	(−0.09, 0.10)	.966	0.05	(−0.05, 0.15)	.347
Early Phase Timing	0.04	(−0.05, 0.15)	.407	0.04	(−0.05, 0.13)	.444
FOV	−0.01	(−0.12, 0.11)	.904	0.06	(−0.05, 0.17)	.251
Resolution	0.03	(−0.06, 0.11)	.505	0.01	(−0.06, 0.10)	.725
Contralateral Image Quality	0.06	(−0.02, 0.14)	.146	0.04	(−0.04, 0.11)	.389
Patient Motion	0.08	(−0.12, 0.25)	.462	−0.18	(−0.30, −0.01)	.008 *
Contrast Administration	−0.15	(−0.31, 0.10)	.135	−0.10	(−0.32, 0.18)	.493
Combined Factors	0.06	(−0.02, 0.13)	.149	0.04	(−0.04, 0.11)	.295

Abbreviations: AUC, area under the curve; CI, confidence interval; FOV, field of view; FTV, functional tumor volume; LL, lower limit; UL, upper limit.
Difference of AUCs between adherent and nonadherent subsets was calculated as AUC of adherent subset minus AUC of nonadherent subset.
* $P < .05$.

patient motion, n = 34; contrast administration, n = 21; T2 patient motion n = 29; contrast administration, n = 27) and AUC with a wide range of 95% CI (Table 3). In these 2 factors, the majority of the FTV changes with adherent image quality factors had lower estimated AUC values than those with non-adherent image quality factors, although the differences between the AUCs did not reach statistical significance except for patient motion at T2 (Figure 3 and Table 3): AUC values for adherent vs non-adherent subsets were 0.67 vs 0.83 for contrast administration at T1, 0.69 vs 0.87 for patient motion at T2, and 0.69 vs 0.79 for contrast administration at T2.

Combined Factors. FTV changes that were adherent for all image quality factors ("combined factors") showed higher AUC values than those that were non-adherent (T1, 0.71 vs 0.66; T2, 0.72 vs 0.68), although the difference between the AUCs did not reach statistical significance (Figure 3 and Table 3). Figure 4 shows the ROC curves of adherent or non-adherent FTV change subsets for combined factors.

DISCUSSION

In this study, we showed that MRI protocol adherence has an impact on the performance of FTV as a predictor of neoadjuvant treatment response. Importantly, these effects were found in the set of MRI examinations already meeting acceptance criteria for I-SPY2. The current MRI acceptance rate in I-SPY2 is >95%. These study results suggest that stricter protocol adherence requirements have the potential to increase predictive performance.

Neoadjuvant chemotherapy (NAC) enables the monitoring of tumor response in vivo, in addition to the downsizing of tumors before surgery. Because pCR is strongly associated with better outcomes, especially in more aggressive subtypes (17, 18), the ability to reliably predict non-pCR at the early-treatment (T1) and inter-regimen (T2) time points could impact treatment decision-making and allow for personalized redirection to more effective therapy. In addition, accurate early prediction of pCR could facilitate deescalation in patients who are responding well to treatment.

Breast MRI has shown greater accuracy than clinical examination or conventional imaging in predicting residual disease after NAC (19–25). A previous analysis of patients with breast cancer undergoing NAC also showed that FTV derived from DCE MRI was a stronger predictor of pCR after NAC than clinical assessment (6). Thus, MRI would be a reliable and clinically relevant modality to monitor breast cancer tumor response. Our study suggests that MRI protocol adherence is important to improve the prediction of pCR at the early-treatment (T1) and inter-regimen (T2) time points, when treatment redirection is possible. These analyses are continuing and will be used to refine quantitative imaging requirements in the ongoing I-SPY 2 TRIAL.

In multicenter trials involving MR quantitative imaging, there are many challenges to standardizing imaging sequences for different scanner manufacturers, field strengths, and coil configurations. In I-SPY 2, patients have 4 MRI examinations over the course of NAC, and using similar scan parameters for a single patient's visits over a 6-month period is logistically and technically challenging. Although a standardized MR imaging protocol with specific image acquisition parameter ranges was prospectively distributed to all participating sites, differences in clinical workflow and institutional infrastructure made protocol adherence challenging across sites. Also, while similar scan parameters for each patient's 4 MRI visits are crucial for accurate FTV calculation and assessment of FTV change from baseline, it was not always feasible to obtain adherence to baseline values across all 4 MRI visits. Our study showed that FTV change with adherent image quality in all factors showed higher AUC value than that with non-adherent image quality, although the difference did not reach statistical significance. Conversely, the nonsignificant difference may suggest that our FTV calculation method is somewhat robust in the context of minor protocol non-adherence. Future versions of the MRI protocol and standard operating procedure documents will emphasize the importance of protocol adherence for accurate calculation of FTV, and real-time remediation for sites with non-adherent data will be implemented.

Figure 4. Receiver operating characteristic curves of adherent or non-adherent FTV change subsets at (A) T1 and (B) T2 in combined factors. FTV change with adherent image quality in combined factors had higher AUC value than that with nonadherent image quality (T1, 0.71 vs 0.66; T2, 0.72 vs 0.68), although the differences did not reach statistical significance.
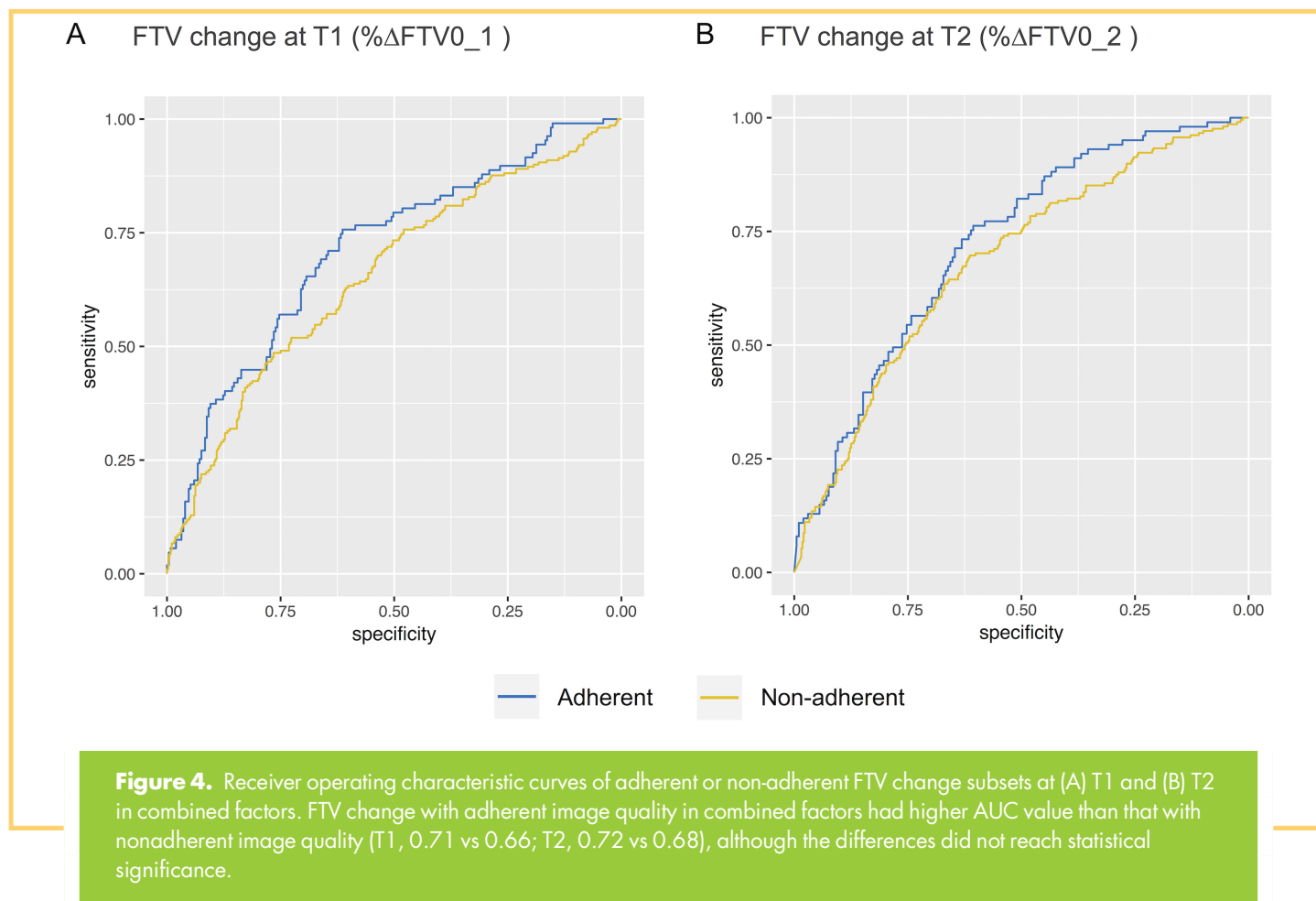
DICOM-confirmable factors included DCE parameters related to scan timing and voxel size. FTV calculation is based on predefined criteria for signal enhancement, which is dynamic and sensitive to timing in DCE MRI (6). Thus, it is reasonable that FTV change had higher estimated AUC in the adherent subset than in the non-adherent subset in relation to acquisition duration and early phase timing, albeit the differences did not reach statistical significance. FTV calculation is also dependent on voxel size because it is calculated by multiplying the voxel size by the number of pixels fulfilling the criteria for signal enhancement. Therefore, voxel size–related parameters including FOV and spatial resolution may affect FTV calculation and its predictive performance.

Although DICOM-confirmable factors were relatively easy to verify, the not DICOM-confirmable factors were much more challenging to assess. We analyzed contralateral image quality, which was originally assessed in the process of analyzing contralateral background parenchymal enhancement in a separate study (12), as one of the not DICOM-confirmable factors. An automated method was used to select and review 3 representative slices, but it was still a time-consuming process to review and score all 2970 examinations. Thus, we hypothesized that the contralateral image quality might reflect the overall imaging quality of MRI. Although a scoring method reviewing just 3 slices of the contralateral breast is one limitation of this study, our results

show that contralateral image quality impacts the performance of FTV change for predicting pCR. This result suggests the importance of verifying quality factors such as magnetic shimming, breast position, fat suppression, and artifacts in both breasts at the time of scanning (26). It may also indicate the technical challenges of FTV calculation for ptotic or large breasts, which are difficult to optimally position in the coil and prone to poor fat suppression or signal flare near the breast coil.

Other not DICOM confirmable factors, patient motion, and contrast administration had counterintuitive estimated results, with lower AUC for the adherent subset than the nonadherent subset. We can think of 2 possible explanations for these results. First, owing to the challenges in assessing and documenting these image quality factors retrospectively during FTV calculation, the adherent subsets might have included unknown or undetected patient motion or contrast administration errors. Identification of patient motion was subjective, and the presence of motion was not recorded accurately in real time. To the best of our knowledge, there is also no reliable method to quantify and correct for motion on breast MRI, particularly in the neoadjuvant biomarker setting. Changes to Imaging Core Lab postprocessing routines for image registration and recent modifications to the study database used to document the presence of motion will hopefully improve the available data on patient motion in the future.

Verification of protocol-adherent contrast administration rate and saline flush volume relied on self-report from participating sites, and data entry errors compromised the completeness and reliability of this information. Recent improvements to the contrast injection case report form submitted by sites and modifications to the database used to store this information have made tracking of this important factor more robust.

Second, the small sample size in the non-adherent subsets leads to considerable uncertainty in the estimated AUCs. It is unclear which, if any, of these issues are responsible for the counterintuitive results. Given the limitations of the data collected on motion and contrast administration error and the small sample size in the nonadherent subsets, definitive statements regarding the causes of the counterintuitive results, or whether they are in fact genuine, cannot be made at this time.

This study has limitations. First, it was a retrospective study, and collecting data for not DICOM confirmable factors was particularly challenging. Second, confounding factors might affect our results. Possible confounding factors include 4 tumor subtypes, 10 drug arms, and 25 participating sites. Considering the statistical uncertainty owing to the small sample size when the whole cohort was divided into each subcohort, we could not adequately assess the impact of these confounding factors.

In conclusion, our results highlighted that MRI protocol adherence has an impact on the performance of FTV for predicting pathological complete response. This was a retrospective study of 2970 MRI examinations that were accepted for FTV calculation in the I-SPY2 TRIAL and met a quality standard of the prospective acceptance standards. Thus, our results reflect incremental improvements. The results of this study will be used to inform higher standards for personalized strategies.

REFERENCES

1. Kaufmann M, Hortobagyi GN, Goldhirsch A, Scholl S, Makris A, Valagussa P, Blohmer J-U, Eiermann W, Jackesz R, Jonat W, Lebeau A, Loibl S, Miller W, Seeber S, Semiglazov V, Smith R, Souchon R, Stearns V, Untch M, von Minckwitz G. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update. J Clin Oncol. 2006;24:1940–1949.
2. Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, Margolese RG, Hoehn JL, Vogel VG, Dakhil SR, Tamkus D, King KM, Pajon ER, Wright MJ, Robert J, Paik S, Mamounas EP, Wolmark N. Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. J Clin Oncol. 2008;26:778–785.
3. Cho N, Im SA, Park IA, Lee KH, Li M, Han W, Noh DY, Moon WK. Breast cancer: early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging. Radiology. 2014;272:385–396.
4. Heldahl MG, Lundgren S, Jensen LR, Gribbestad IS, Bathen TF. Monitoring neoadjuvant chemotherapy in breast cancer patients: improved MR assessment at 3 T? J Magn Reson Imaging. 2011;34:547–556.
5. Cheung Y-C, Chen S-C, Su M-Y, See L-C, Hsueh S, Chang H-K, Lin Y-C, Tsai C-S. Monitoring the size and response of locally advanced breast cancers to neoadjuvant chemotherapy (weekly paclitaxel and epirubicin) with serial enhanced MRI. Breast Cancer Res Treat. 2003;78:51–58.
6. Hylton NM, Blume JD, Bernreuter WK, Pisano ED, Rosen MA, Morris EA, Weatherall PT, Lehman CD, Newstead GM, Polin S, Marques HS, Esserman LJ, Schnall MD; ACRIN 6657 Trial Team and I-SPY 1 TRIAL Investigators. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy–results from ACRIN 6657/I-SPY TRIAL. Radiology. 2012;263:663–672.
7. Hylton NM, Gatsonis CA, Rosen MA, Lehman CD, Newitt DC, Partridge SC, Bernreuter WK, Pisano ED, Morris EA, Weatherall PT, Polin SM, Newstead GM, Marques HS, Esserman LJ, Schnall MD; ACRIN 6657 Trial Team and I-SPY 1 TRIAL Investigators. 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. 2016;279:44–55.
8. Partridge SC, Gibbs JE, Lu Y, Esserman LJ, Tripathy D, Wolverton DS, Rugo HS, Hwang ES, Ewing CA, Hylton NM. MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. AJR Am J Roentgenol. 2005;184:1774–1781.
9. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther. 2009;86:97–100.
10. Rugo HS, Olopade OI, DeMichele A, Yau C, van 'T Veer LJ, Buxton MB, Hogarth M, Hylton NM, Paoloni M, Perlmutter J, Symmans WF, Yee D, Chien AJ, Wallace AM, Kaplan HG, Boughey JC, Haddad TC, Albain KS, Liu MC, Isaacs C, Khan QJ, Lang JE, Viscusi RK, Pusztai L, Moulder SL, Chui SY, Kemmer KA, Elias AD, Edmiston KK, Euhus DM, Haley BB, Nanda R, Northfelt DW, Tripathy D, Wood WC, Ewing C, Schwab R, Lyandres J, Davis SE, Hirst GL, Sanil A, Berry DA, Esserman LJ; I-SPY2 Investigators. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. N Engl J Med. 2016;375:23–34.
11. Park JW, Liu MC, Yee D, Yau C, van 'T Veer LJ, Symmans WF, Paoloni M, Perlmutter J, Hylton NM, Hogarth M, DeMichele A, Buxton MB, Chien AJ, Wallace AM, Boughey JC, Haddad TC, Chui SY, Kemmer KA, Kaplan HG, Isaacs C, Nanda R, Tripathy D, Albain KS, Edmiston KK, Elias AD, Northfelt DW, Pusztai L, Moulder SL, Lang JE, Viscusi RK, Euhus DM, Haley BB, Khan QJ, Wood WC, Melisko M, Schwab R, Helsten T, Lyandres J, Davis SE, Hirst GL, Sanil A, Esserman LJ, Berry DA; I-SPY2 Investigators. Adaptive randomization of neratinib in early breast cancer. N Engl J Med. 2016;375:11–22.
12. Strand F, Arasu V, Nguyen A, Jones E, Harnish R, Newitt D, Hylton N. A comparison of visual, automated and semi-manual approaches in assessing background parenchymal enhancement as a biomarker for response to neoadjuvanttherapy. ISMRM 27th Annual Meeting and Exhibition, Montreol, Canada; 2019 May 13.
13. Hylton NM. Vascularity assessment of breast lesions with gadolinium-enhanced MR imaging. Magn Reson Imaging Clin N Am. 1999;7:411–420.
14. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
15. Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions. 2019.
16. Davison AC, Hinkley DV. Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge: Cambridge University Press; 1997.
17. Cortazar P, Geyer CE. Pathological complete response in neoadjuvant treatment of breast cancer. Ann Surg Oncol. 2015;22:1441–1446.
18. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P, Swain SM, Prowell T, Loibl S, Wickerham DL, Bogaerts J, Baselga J, Perou C, Blumenthal G, Blohmer J, Mamounas EP, Bergh J, Semiglazov V, Justice R, Eidtmann H, Paik S, Piccart M, Sridhara R, Fasching PA, Slaets L, Tang S, Gerber B, Geyer CE, Pazdur R, Ditsch N, Rastogi P, Eiermann W, von Minckwitz G. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384:164–172.
19. Croshaw R, Shapiro-Wright H, Svensson E, Erb K, Julian T. Accuracy of clinical examination, digital mammogram, ultrasound, and MRI in determining postneoadjuvant pathologic tumor response in operable breast cancer patients. Ann Surg Oncol. 2011;18:3160–3163.
20. Shin HJ, Kim HH, Ahn JH, Kim SB, Jung KH, Gong G, Son BH, Ahn SH. Comparison of mammography, sonography, MRI and clinical examination in patients with locally

advanced or inflammatory breast cancer who underwent neoadjuvant chemotherapy. Br J Radiol. 2011;84:612–620.

21. Akazawa K, Tamaki Y, Taguchi T, Tanji Y, Miyoshi Y, Kim SJ, Ueda S, Yanagisawa T, Sato Y, Tamura S, Noguchi S. Preoperative evaluation of residual tumor extent by three-dimensional magnetic resonance imaging in breast cancer patients treated with neoadjuvant chemotherapy. Breast J. 2006;12:130–137.

22. Yeh E, Slanetz P, Kopans DB, Rafferty E, Georgian-Smith D, Moy L, Halpern E, Moore R, Kuter I, Taghian A. Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. AJR Am J Roentgenol. 2005;184:868–877.

23. Londero V, Bazzocchi M, Del Frate C, Puglisi F, Di Loreto C, Francescutti G, Zuiani C. Locally advanced breast cancer: comparison of mammography, sonography and MR imaging in evaluation of residual disease in women receiving neoadjuvant chemotherapy. Eur Radiol. 2004;14:1371–1379.

24. Rosen EL, Blackwell KL, Baker JA, Soo MS, Bentley RC, Yu D, Samulski TV, Dewhirst MW. Accuracy of MRI in the detection of residual breast cancer after neoadjuvant chemotherapy. AJR Am J Roentgenol. 2003;181:1275–1282.

25. Montemurro F, Martincich L, De Rosa G, Cirillo S, Marra V, Biglia N, Gatti M, Sismondi P, Aglietta M, Regge D. Dynamic contrast-enhanced MRI and sonography in patients receiving primary chemotherapy for breast cancer. Eur Radiol. 2005;15:1224–1233.

26. Harvey JA, Hendrick RE, Coll JM, Nicholson BT, Burkholder BT, Cohen MA. Breast MR imaging artifacts: how to recognize and fix them. Radiographics. 2007;27: S131–45.