

UCSF

UC San Francisco Previously Published Works

Title

Additive value of diffusion-weighted MRI in the I-SPY 2 TRIAL

Permalink

<https://escholarship.org/uc/item/50d8v7dh>

Journal

Journal of Magnetic Resonance Imaging, 50(6)

ISSN

1053-1807

Authors

Li, Wen
Newitt, David C
Wilmes, Lisa J
[et al.](#)

Publication Date

2019-12-01

DOI

10.1002/jmri.26770

Peer reviewed



Published in final edited form as:

J Magn Reson Imaging. 2019 December ; 50(6): 1742–1753. doi:10.1002/jmri.26770.

Additive value of diffusion weighted MRI in the I-SPY 2 TRIAL

Wen Li, PhD¹, David C. Newitt, PhD¹, Lisa J. Wilmes, PhD¹, Ella F. Jones, PhD¹, Vignesh Arasu, MD¹, Jessica Gibbs, BS¹, Bo La Yun, MD^{1,2}, Elizabeth Li, MS^{1,3}, Savannah C. Partridge, PhD⁴, John Kornak, PhD⁵, Laura J. Esserman, MD MBA⁷, Nola M. Hylton, PhD¹ on behalf of the I-SPY 2 Consortium⁶

¹Department of Radiology & Biomedical Imaging, the University of California, San Francisco

²Department of Radiology, Seoul National University Bundang Hospital

³Department of Biomedical Engineering, the University of California, Davis

⁴Department of Radiology, University of Washington

⁵Department of Epidemiology and Biostatistics, the University of California, San Francisco

⁶Quantum Leap Healthcare Collaborative

⁷Department of Surgery, the University of California, San Francisco

Abstract

BACKGROUND—The change in apparent diffusion coefficient (ADC) measured from DWI has shown to be predictive of pathologic complete response (pCR) for patients with locally invasive breast cancer undergoing neoadjuvant chemotherapy.

PURPOSE—To investigate the additive value of tumor ADC in a multi-center clinical trial setting.

STUDY TYPE—Retrospective analysis of multicenter prospective data

POPULATION—415 patients enrolled in the I-SPY 2 TRIAL from 2010 to 2014 were included.

FIELDSTRENGTH/SEQUENCE—1.5T or 3T MRI system using a fat-suppressed single-shot echo planar imaging sequence with b-values of 0 and 800 s/mm^2 for DWI, followed by a T1-weighted sequence for dynamic contrast-enhanced MRI (DCE-MRI) performed at pre-NAC (T0), after 3 weeks of NAC (T1), mid-NAC (T2), and post-NAC (T3).

ASSESSMENT—Functional tumor volume and tumor ADC were measured at each MRI exam, pCR measured at the surgery was assessed as the binary outcome. Breast cancer subtype was defined by hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) status.

STATISTICAL TESTS—A logistic regression model was used to evaluate associations between MRI predictors with pCR. The cross-validated area under the curve (AUC) was calculated to assess the predictive performance of the model with and without ADC.

RESULTS—354 patients (128 HR+/HER2-, 60 HR+/HER2+, 34 HR-/HER2+, 132 HR-/HER2-) were included in the analysis. In the full cohort, adding ADC predictors increased the AUC from 0.76 to 0.78 at mid-NAC and from 0.76 to 0.81 at post-NAC. In HR/HER2 subtypes, the AUC increased from 0.52 to 0.65 at pre-NAC for HR+/HER2-, from 0.67 to 0.73 at mid-NAC and from 0.72 to 0.76 at post-NAC for HR+/HER2+, from 0.71 to 0.81 at post-NAC for triple negatives.

DATA CONCLUSION—The addition of ADC to standard FTV MRI showed improvement in the prediction of treatment response in HR+ and triple negative breast cancer.

Keywords

breast MRI; breast cancer; treatment response; functional tumor volume; apparent diffusion coefficient; pathologic complete response

INTRODUCTION

Neoadjuvant chemotherapy (NAC) is at least as effective as adjuvant chemotherapy for the locally advanced breast cancer^{1,2}. Clinical trials have shown that patients who reached pathologic complete response (pCR) after NAC have better long-term survival rates than patients who do not³⁻⁵. The I-SPY 2 TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response through Imaging and Molecular Analysis 2) is a multi-center, phase 2 trial using response-adaptive randomization within biomarker subtypes to evaluate a series of novel drugs added to standard NAC for women with high-risk stage II/III breast cancer⁶. The primary end point is pCR. A key component of this study is the serial MR imaging, which is used to measure each patient's response to chemotherapy and to predict the likelihood of the patient achieving pCR at the end of the treatment.

In the I-SPY 1 (ACRIN 6657) TRIAL, functional tumor volume (FTV) — an imaging marker computed by applying enhancement thresholds to dynamic-enhanced (DCE) MRI⁷ — showed strong association with pCR⁸ and recurrence-free survival (RFS)⁹. In addition to DCE, the I-SPY 2 TRIAL is testing whether diffusion weighted MRI (DWI), a non-contrast method that characterizes water mobility and cellularity by measuring the apparent diffusion coefficient (ADC), acquired during the same MRI exam as DCE, can provide valuable distinct information about tumor response. The ACRIN (American College of Radiology Imaging Network) 6698 trial, a sub-study of I-SPY 2, evaluated the change in tumor ADC for predicting pCR. The trial found that after 12 weeks of therapy (between drug regimens), the percentage change in tumor ADC predicts pCR¹⁰. Their study also showed that ADC achieved higher predictive performance in hormone receptor (HR) positive and human epidermal growth factor receptor 2 (HER2) negative cancer than other cancer subtypes.

In this study, we propose to investigate the additive value of ADC to FTV alone in predicting pCR in I-SPY 2, in the full cohort and in HR/HER2 breast cancer subtypes. The purpose is to test if there is any additional value ADC can provide to the prediction model that has FTV predictors already in place. Although numerous studies have demonstrated the use of DCE-MRI or DWI in assessing treatment response to NAC, few have tested the approach of combining information from both MR methods¹¹⁻¹³. Furtherly, we propose to test the additive value of ADC in individual HR/HER cancer subtypes based on previous findings

that both FTV and ADC perform differently in predicting pCR in different cancer subtype^{14,15}.

MATERIALS AND METHODS

Patient Population

Women 18 years of age and older diagnosed with stage II or III breast cancer and with tumor size measured ≥ 2.5 cm were eligible to enroll in the I-SPY 2 TRIAL[6]. Patients with evidence of distant metastasis were excluded. Biomarker assessments based on hormone (estrogen and progesterone) receptors (HR+/-) and human epidermal growth factor receptor 2 (HER2+/-) status and a 70-gene assay (MammaPrint, Agendia) were performed at the baseline and used for treatment randomization⁶. In addition to standard immunohistochemical and fluorescence in situ hybridization (FISH) assays, the protocol included a microarray-based assay of HER2 expression (TargetPrint, Agendia) to assign HR and HER2 statuses. Patients with tumors that were designated as HR+/HER2- and low risk according to the 70-gene assay were excluded because the potential benefit of receiving investigated drugs plus chemotherapy for patients with less proliferative tumors are low in the consideration of the risk of drug side effects^{16,17}. All patients provided written informed consent to participate in the study. A second consent was obtained if the patient was randomized to an experimental treatment.

Pathologic Assessment of Response

Figure 1 shows the schema of the I-SPY 2 TRIAL. Pathologic complete response – defined as the absence of residual cancer in the breast or lymph nodes at the time of surgery – is the primary end point of the I-SPY 2 TRIAL. All patients were classified as pCR or non-pCR by a trained pathologist at the time of definitive surgery. Patients who left the study without completing the therapy or patients who did not undergo surgery for any reason were counted as non-pCR.

MRI Acquisition

MRI exams were performed before the initiation of NAC (pre-NAC, T0), after 3 weeks of treatment (early-NAC, T1), after 12 weeks and between drug regimens (mid-NAC, T2), after completion of NAC and prior to surgery (post-NAC, T3). MRI data were acquired on 1.5T or 3T scanners with a dedicated breast radiofrequency coil, across a variety of vendor platforms and institutions. All MRI exams for the same patient were performed using the same magnet configuration (manufacturer, field strength, and breast coil model). The standard image acquisition protocol included T2-weighted, DW-, and DCE-MRI sequences performed bilaterally in the axial orientation (Table S1). DW-MRI were performed using a fat-suppressed single-shot echo planar imaging sequence with the following parameters: TR = 4000 ms, TE = 50–100 ms, FOV = 260–360 mm to achieve full bilateral coverage, acquisition matrix = 128–192 with in-plane resolution ≈ 1.9 mm, slice thickness = 3–5 mm, slice gap = 1 mm, and number of signal averages = 2. Diffusion weighting b-values of 0 and 800 s/mm^2 were specified, with an acquisition time ≈ 5 minutes.

DCE-MRI were performed by acquiring series of three-dimensional fat-suppressed T1-weighted images with the following parameters: TR = 4–10 ms, minimum TE, flip angle = 10–20 degrees, field of view (FOV) = 260–360 mm to achieve full bilateral coverage, acquisition matrix = 384–512 with in-plane resolution 1.4 mm, and slice thickness 2.5 mm, temporal resolution = 80–100s. Gadolinium contrast agent was administered intravenously at a dose of 0.1 mmol/kg body weight, and at a rate of 2 mL/second, followed by a 20 mL saline flush. The same contrast agent brand was used for all MRI exams for the same patient. Pre-contrast and multiple post-contrast images were acquired using identical sequence parameters. Post-contrast imaging continued for at least 8 minutes following contrast agent injection.

Quantitative Image Analysis

The functional tumor volume (FTV) for each imaging visit was calculated from DCE-MRI as previously described¹⁸. Briefly, the segmentation method calculated the volume of all voxels, within a manually-specified 3D region of interest (ROI) encompassing the enhancing lesion, that exceeded a percentage enhancement (PE) threshold of 70% at approximately 2.5 minutes post-contrast. For the consistency of FTV measurements among imaging visits, ROIs for the same patient should be the same size at all visits. If tumor grew larger during the treatment, ROI can be enlarged accordingly but it cannot be shrunk in size only because the tumor shrank. For isolated patients, the 70% PE threshold was adjusted by the imaging core laboratory at UCSF at the T0 visit when needed to provide a satisfactory segmentation of the enhancing lesion. In these cases, the adjusted threshold was used for segmenting all subsequent studies for the patient. The final FTV analysis for each visit was reviewed and approved by a designated breast radiologist at each site and by the imaging core laboratory.

All diffusion images were centrally processed at the core laboratory using in-house software developed in IDL (ITT Visual Information Solutions, Boulder, Colorado). Mono-exponential ADC maps were calculated as previously described¹⁹ based on:

$$ADC = \frac{1}{b} \ln \frac{S_{DWI}}{S_0}$$

where S_{DWI} is the signal intensity at a diffusion weighting of $b = 800 \text{ s/mm}^2$, and S_0 is the signal intensity at $b = 0 \text{ s/mm}^2$. The tumor region of interest (ROI) was manually defined to encompass areas that were hyper-intense on the $b = 800 \text{ s/mm}^2$ images and hypo-intense on the corresponding ADC maps (see Figure 2). Enhanced areas on corresponding DCE-MRI were also used to guide ROI selection. Care was taken to avoid non-enhancing regions with high signal in the T2-weighted ($b = 0 \text{ s/mm}^2$) images arising from cysts, hematomas, or necrosis. Clip artifacts were also excluded. For T2 and T3 studies with no visible residual lesion, the ROI was drawn to include only fibroglandular tissue in the region the tumor was localized to in prior visits, and if possible, the ROI was drawn with the comparable size as in the previous visit when the tumor was visible. All ROIs were drawn by a radiologist certified to evaluate DWI images (B.L.Y.), a graduate student with breast MR background (E.L.), or trained research staff with over 4 years (W.L.) and 10 years (J.G.) of DW-MRI analysis experience. All ROI definitions were reviewed and adjusted if necessary by the first author

(W.L.). Readers were blinded to the pathologic outcome. Tumor ADCs were calculated as the mean of voxels within the ROI for each imaging visit. The quality of the DW-MRI studies for each patient were ranked by W.L. (-1: unacceptable; 0: missing data; 1: acceptable; or 2: good)²⁰. Poor-quality images were excluded because of severe distortion, artifact, fat suppression, or signal-to-noise ratio (SNR) at the tumor area. DW images with acceptable or good qualities were included in this study.

FTV and tumor ADC values were calculated at each treatment time point (T0, T1, T2, T3) and percentage changes from the baseline (T0) value were calculated at each subsequent visit (% FTV0_1, % FTV0_2, and % FTV0_3 for T1 to T3, similarly for ADC). Baseline value and percentage changes of FTV and ADC were analyzed in this study (see Figure 3). Tumor diameters were measured by site breast radiologists on pre-NAC MRI, as the greatest extent of disease.

Statistical Analysis

Statistical analysis was performed to assess the predictive performance of single or multiple MR predictors for pCR versus non-pCR outcomes. All statistical analyses were performed using R version 3.4.1 (R Foundation for Statistical Computing, Vienna Austria).

Based on our observations, FTV and ADC data had a skewed distribution so numeric values were expressed as a median with interquartile ranges in the summary data, except when stated otherwise. In a single-predictor analysis, the Wilcoxon rank sum test was used to test differences in MR metrics in pCR versus non-pCR patients, whereas the Fisher's exact test was used to estimate associations of race, ethnicity, menopausal status, hormone receptor status, HER2 status, node status, with outcomes. The predictive performance of single predictors was estimated by the area under the receiver operating characteristic (ROC) curve (AUC).

The multiple predictor analysis was conducted to study the additive value of ADC to the model with FTV predictors already in place. The analysis was performed separately in the full cohort and in each breast cancer subtype. The AUC with 10-fold cross validation were calculated to assess the predictive performance of a logistic regression model. Specifically, the dataset was randomly split into 10 subsets with equal size. One subset was held as the testing data for validating the model and the remaining 9 subsets were used as training data. The process was then repeated 10 times until each of the 10 subsets had been used exactly once as the testing data. The 10 results can then be averaged to produce a single estimation of AUC. FTV/ADC predictors considered in the model were FTV baseline, ADC baseline and change in FTV/ADC at later treatment time points compared to the baseline. Optimized logistic regression models with FTV predictors only were built at each MR visit, by having the highest AUC among all FTV models upon each MR visit, which included models built with single or all combinations of FTV predictors. ADC predictors were then added to the optimized FTV model and AUCs were calculated. The model achieved highest AUC after all ADC predictors available up to the MR visit were tested was selected as the "FTV+ADC" for the visit. HR/HER2 subtype was included as a categorized variable in models for the full cohort. Interactions between subtype and FTV/ADC predictors were considered in the full cohort analysis. Interactions between baseline and change in the same type of imaging

predictors (FTV or ADC) were also considered in the analysis. The p-value of ADC predictors in the logistic regression model was evaluated by the likelihood ratio test of models with and without ADC predictors. All tests were performed two-sided and at the $\alpha = 0.05$ statistical significance level.

RESULTS

Patient and Tumor Characteristics

Due to the availability of pathologic outcomes, data from 415 patients enrolled in the I-SPY 2 trial between 2010 and 2014 and treated with four experimental drugs were included in this study. Among them, 61 patients (14.7%) were excluded for at least one of the following reasons: 1) only a pre-NAC MRI was performed without subsequent follow-up MR exams (n=9); 2) poor DWI quality for pre-NAC or all subsequent visits (n=51); 3) or missing pCR outcome status (n=1). As a result, we obtained an analysis cohort of 354 patients (see Table 1 for patients characteristics).

In this cohort, 120 (34%) patients achieved pCR and 234 (66%) patients did not (non-pCR). The full cohort (n=354) can be classified into four groups defined by the HR and HER2 positive or negative status. The number of patients and the pCR rates are 128 (16%), 60 (30%), 34 (59%), and 132 (46%) for HR+/HER2-, HR+/HER2+, HR-/HER2+, and HR-/HER2-, respectively. Table 1 shows that tumor diameters measured either by MRI or clinical exam were statistically significantly different between pCR and non-pCR groups. So were pCR rates among the HR/HER2 subgroups.

Single Predictor Analysis

The results of the single predictor analysis for FTV or ADC measures are listed in Table 2. The difference between pCR and non-pCR (Diff. column in Table 2) represents the median differences in these two groups, with a minus sign indicating that predictor values for the pCR group are smaller than the values for the non-pCR group. Table 2 shows that all FTV predictors can predict pCR with estimated AUCs statistically significantly above 0.5 and in the range from 0.63 to 0.70. Similarly, all ADC predictors except the ADC measured at pre-NAC yielded AUCs above 0.5 in statistical significance and in the range from 0.57 to 0.72. AUC values of FTV and ADC increased steadily as treatment progressed, and highest AUCs were observed at post-NAC for both FTV and ADC. Results in stratified subgroups by HR/HER status are listed in Tables S2 and S3.

Multiple Predictor Analysis

The effect of adding more predictors to the logistic regression model to improve AUCs was shown in Table 3. At each treatment time point, the highest AUC of predicting pCR among combinations of FTV predictors available upon each time point was listed under the "Optimized FTV" column. For comparison, the AUC of using the single FTV predictor at the corresponding time point, i.e. FTV0 for pre-NAC, % FTV0_1 for early-NAC, % FTV0_2 for mid-NAC, etc. was listed under the "Single FTV predictor" column. Please note, subtype was added to the "Single FTV predictor" model for the analysis in the full cohort. The highest AUC found by adding any ADC predictors to the "Optimized FTV"

model was shown in column “FTV + ADC”. The “n”s list in Table 3 were for number of patients who had both FTV and ADC available up to each treatment time point so they are different from the ones shown in Table 2 where “n” was for number of patients with single FTV predictor available. Cases where AUCs of “Optimized FTV” increased after adding ADC predictors (at least one among all available predictors upon the corresponding time point) were bolded under the “FTV + ADC” column in Table 3. The table also shows cases where AUCs of “FTV + ADC” yielded lower AUCs than “Optimized FTV” from the same row. In those cases, ADC predictor(s) were “forced” to be added to the “Optimized FTV” model without improving actual predictive value. Figure 4 shows plots of AUCs for the cohorts and time points where ADC did (Figure 4a) and did not (Figure 4b) contribute to the increase of AUC.

In the full cohort, adding ADC increased AUCs from 0.76 to 0.78 at mid-NAC and from 0.76 to 0.81 at post-NAC. The ROC curves from which these two pairs of AUCs were calculated are shown in Figure 5(a) and Figure 5(b). At mid-NAC, the ADC predictors added to the FTV only model were ADC0, % ADC0_2, and the interaction between subtype and ADC0 with p-values of 0.13, 0.00013, and 0.075, respectively. At post-NAC, the ADC predictors added to the FTV only model were ADC0, % ADC0_3, and interactions between subtype and ADC0, and between subtype and % ADC0_3 with p-values of 0.025, <0.0001 for both ADC0 and % ADC0_3, and 0.12 and 0.17 for the interactions. In the HR+/HER2- subtype, AUC increased from 0.52 to 0.65 at pre-NAC when ADC0 was added to the model. The ROC curves of models with FTV0 only and with FTV0 + ADC0 are shown in Figure 5(c). The p-value of ADC0 in the combined model was estimated to be 0.95.

In the HR+/HER2+ subtype, adding ADC increased the AUCs from 0.67 to 0.73 at mid-NAC and from 0.72 to 0.76 at post-NAC. These ROC curves are shown in Figure 5(d) and Figure 5(e). Although adding ADC achieved higher AUC at post-NAC (Figure 5(e)), the additive value only showed in the area when the sensitivity 0.7 and the specificity 0.6. The ADC predictor added to the optimized FTV model at mid-NAC was % ADC0_1 with a p-value of 0.60. ADC predictors added to the optimized FTV model at post-NAC were % ADC0_1, % ADC0_2, and % ADC0_3 with p-values of 0.77, 0.80, 0.51, respectively.

In the HR-/HER2- subtype (triple negative), the AUC increased from 0.71 to 0.81 at T3. The ROC curves shown in Figure 5(f) demonstrate that the improvement of sensitivity occurred when the specificity was < 0.9. ADC predictors added to the optimized FTV only model at T3 were ADC0 and % ADC0_3, with corresponding p-values of 0.011 and <0.0001. Example images of a patient with triple negative breast cancer are shown in Figure 6 and Figure 7. No additive value of ADC was observed in HR-/HER2+ subtype.

DISCUSSION

This study demonstrates the value of adding tumor ADC measured from DWI to the prediction model of using FTV measured from DCE-MRI, which increased the AUC values at mid-NAC (between regimen) and post-NAC (before pre- surgery) in the full cohort.

Furthermore, the additive value was also observed in HR/HER2 subtypes (i.e. HR+/HER2-, HR+/HER2+, and HR-/HER2-).

MRI provides both structural and functional information for tumor tissues. Functional tumor volume measured by DCE-MRI has been shown to be predictive of pCR in previous clinical trials, when chemotherapy reduces the tumor vascularity and thus decreases the contrast-enhanced volume in the tumor^{8,9}. However, the apparent diffusion coefficient can characterize tumor biology by measuring water diffusion (Brownian motion). Cancer tissue has higher cellularity so it should have lower ADC value (more restricted water motion) than benign tumor and normal tissue. This study demonstrated that the tumor mean ADC value increases during the course of NAC, consistent with findings of the ACRIN 6698 clinic trial and other clinical trials^{14,20–22}.

Recently, with the increased image quality and standardization of DWI in clinical applications^{20,23,24}, researchers have started to integrate DWI with DCE-MRI to better predict response in NAC. Li et al. combined pharmacokinetic parameters from DCE-MRI with ADC as a multi-parametric imaging biomarker and showed that the multi biomarker was superior to single-parametric measurements using DCE-MRI or DWI alone after one cycle of NAC²⁵. However, it is a single institute study with 3T DCE- and DW-MRI data with a small patient cohort (n=33). Another study published by Pinker, et al. tested the diagnostic accuracy of multiparametric MRI using DCE-MRI, DWI, and 3-dimensional proton magnetic resonance spectroscopic imaging on 113 lesions²⁶. Their results showed that multiparametric MRI with 3 MRI parameters yielded significantly higher AUC (0.936) in comparison with DCE-MRI alone (0.814). However, the combination of DCE-MRI and DWI did not yield a higher AUC (0.808) than DCE-MRI alone.

One of the advantages of treating cancer patients with NAC is that we can monitor the tumor response using serial MRIs at multiple treatment time points. Several studies investigated single time points when MRI biomarkers are most predictive for pCR^{8,12,27–29}. However, they only tested the prediction of variables measured at specific time point and very few studies investigated the combined model with predictors measured at current and previous timepoints if applicable. Since achieving a better prediction of pCR was the goal for this study, AUC was used as a numeric estimation to compare different models. All of our optimized models were built by achieving the highest AUCs and 10-fold cross validations were used in the estimation of AUC for multiple predictor models to avoid overfitting and improve the validity of our conclusions. Thus, optimized models in this study included only predictors that contribute to the increase of AUC while minimizing bias in our results. However, even with cross-validation, the optimized model selected may not always be replicated in a different set of patients.

AUCs can only provide general estimations for prediction models. To fully appreciate the improvement in prediction, the ROC curve needs to be plotted to evaluate the trade-off between the sensitivity and specificity when cut-point varies. For example, ROC curves of the combined model and the “optimized FTV” model crossed at post-NAC in HR+/HER2+ and triple negative subtypes. In these cases, the additive value of ADC is partial depending on the range of sensitivity/specificity in clinical interests. If the interest is to predict pCR

with modest to high specificities (> 0.5) in HR+/HER2+, adding ADC can help to improve sensitivities at post-NAC. Similarly, ADC can improve sensitivity in triple negative cancer at post-NAC only if specificity > 0.9 . Our results also showed that there were many cases where AUCs did not increase or even decrease by adding ADC predictors to FTV only models. In these models, ADC did not add any predictive value but noise.

In previously published results of the ACRIN 6698¹⁰, the authors explored the combination of ADC, FTV, and HR/HER2 subtype at mid-NAC. A model combining percentage change ADC, percentage change FTV, and cancer subtype resulted in an AUC of 0.71. They also found that the predictive value of ADC may be comparable to or higher than that of FTV between drug regimens, particularly in HR+/HER2- cancer patients. This study went further in the width and depth to study the additional predictive value of ADC to FTV at all treatment time points and in individual cancer subtypes. In our study, we found the model combining ADC, FTV, and subtype achieved higher AUCs than FTV plus subtype between drug regimens and at post-NAC. The highest AUC=0.81 was observed at the post-NAC. By subtype, ADC was able to increase the AUCs at pre-NAC for HR+/HER2-, at mid-NAC and post-NAC for HR+/HER2+, and at post-NAC for HR-/HER2-.

Although a subset of our study cohort (n=95) were also included in the ACRIN 6698 study¹⁰, these two studies are different: 1) the primary aim of ACRIN 6698 was to evaluate the prediction of tumor ADC to pCR while this study focused on the additive value of ADC; 2) ADC maps in ACRIN 6698 were generated from 4-b DWI but in this study ADC maps were generated from 2-b DWI; 3) as an imaging trial, ACRIN 6698 applied carefully designed quality control and management while DWI collected in I-SPY 2 had no quality control. That is why there were a substantial number of patients (n=51) excluded from the analysis of this study due to the poor image quality. This may be the disadvantage of using the I-SPY 2 data instead of ACRIN 6698. However, the much larger patient population of I-SPY 2 (>2,000 enrollments) provide a larger sample size (n=354 for this study versus n=242 for ACRIN 6698 study) and more available for future analysis.

All HR/HER2 subtypes had increased AUC after ADC was added to extend the model with FTV only, except HR-/HER2+. This may in part be due to the sample size being low (n=24–33, depending on the visit). In the HR+/HER2- subtype, AUC increased from 0.52 to 0.65 at pre-NAC and ROC curves demonstrated that the predictive performance of the FTV + ADC model was overall better than the optimized FTV model. The AUC of the combined model achieved statistical significance even though neither FTV nor ADC alone had AUCs above the statistical significance level (Table S1). HR+/HER2- breast cancer has showed limited benefit from NAC^{30,31}. If imaging predictors can identify patients who will and will not benefit before NAC starts, it will help doctors plan treatment more effectively and timely.

The I-SPY framework represents a prospective trial with careful quality control of patient inclusion criteria, MRI acquisition and measurement, and clearly defined pathologic outcomes that is advantageous for answering this research question. However, this study has limitations. First, the clinical trial used a low temporal resolution (90s) in DCE-MRI, which may preclude the use of pharmacokinetic modeling even though it meets the current American College of Radiology guidelines. Second, although the DCE-MRI and FTV

measurements were made under careful quality control management by the imaging core lab of I-SPY 2, the DW-MR images had limited quality which could affect the ADC measurement. Third, this was a multi-center clinical trial so MRI scanners with different magnet strengths (1.5T, 3T) from different vendors were used to acquire data. The phantom study by Keenan et al. reported that ADC values vary among vendors and magnet strengths³². Newitt et al. recently investigated multisite concordance of ADC measurements across the National Cancer Institute's quantitative imaging network and found discrepancies among different platforms[24]. Fourth, the patient cohort of this study was taken from the experimental arms of four different completed drugs in I-SPY 2. Drug agents in I-SPY 2 target different breast cancer subtypes, so adjusting the combined model by HR/HER2 status may have been confounded by the different treatment effects. Lastly, the optimized models were built from the data in this study only. Even we tried to avoid the problem of over-fitting or selection bias by using 10-fold cross-validation, we should treat final forms of optimize models with caution. Due to the nature of being a clinical trial for targeted therapy, subtype cancer cohorts in this study did not share the same distribution as in the general population of breast cancer.

In conclusion, the results of this study showed that tumor ADC measured during the treatment of neoadjuvant chemotherapy may provide additive value to the functional tumor volume in predicting pathologic complete response, especially in HR+ and triple negative breast cancer patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The I-SPY2 Trial is supported by Quantum Leap Healthcare Collaborative (2013 to present), the Foundation for the National Institutes of Health (2010 to 2012) and the National Cancer Institute Center for Biomedical Informatics and Information Technology (2010–2012). The authors would like to thank the patients who participated in I-SPY 2.

Grant Support: NIH R01 CA132870 and NIH U01 CA225427

REFERENCES

1. Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B. Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr.* 2001;96:102. [PubMed: 11773300]
2. Deo SVS, Bhutani M, Shukla NK, Raina V, Rath GK, Purkayasth J. Randomized trial comparing neo-adjuvant versus adjuvant chemotherapy in operable locally advanced breast cancer (T4b N0–2 M0). *J Surg Oncol.* 2003;84:192–197. [PubMed: 14756429]
3. Thompson AM, Moulder-Thompson SL. Neoadjuvant treatment of breast cancer. *Ann Oncol.* 2012;23 Suppl 1(suppl_10):x231–6.
4. Esserman LJ, Berry DA, Cheang MCU, et al. Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat.* 2012;132:1049–1062. [PubMed: 22198468]

5. Kong X, Moran MS, Zhang N, Haffty B, Yang Q. Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. *Eur J Cancer*. 2011;47:2084–2090. [PubMed: 21737257]
6. 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. [PubMed: 19440188]
7. Newitt DC, Aliu SO, Witcomb N, 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*. 2014;7:94–100. [PubMed: 24772212]
8. Hylton NM, Blume JD, Bernreuter WK, et al. 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. [PubMed: 22623692]
9. Hylton NM, Gatsonis CA, Rosen MA, 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*. 2016;279:44–55. [PubMed: 26624971]
10. Partridge SC, Zhang Z, Newitt DC, et al. Diffusion-weighted MRI Findings Predict Pathologic Response in Neoadjuvant Treatment of Breast Cancer: The ACRIN 6698 Multicenter Trial. *Radiology*. 2018;289:618–627. [PubMed: 30179110]
11. Marino MA, Helbich T, Baltzer P, Pinker-Domenig K. Multiparametric MRI of the breast: A review. *J Magn Reson Imaging*. 2018;47:301–315. [PubMed: 28639300]
12. Minarikova L, Bogner W, Pinker K, et al. Investigating the prediction value of multiparametric magnetic resonance imaging at 3 T in response to neoadjuvant chemotherapy in breast cancer. *Eur Radiol*. 2017;27:1901–1911. [PubMed: 27651141]
13. Li X, Arlinghaus LR, Ayers GD, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: pilot study findings. *Magn Reson Med*. 2014;71:1592–1602. [PubMed: 23661583]
14. Partridge SC, Zhang Z, Newitt DC, et al. ACRIN 6698 trial: Quantitative diffusion-weighted MRI to predict pathologic response in neoadjuvant chemotherapy treatment of breast cancer. In: *Journal of Clinical Oncology*. ; 2017:11520–11520.
15. Li W, Arasu V, Newitt DC, et al. Effect of MR Imaging Contrast Thresholds on Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Subtypes: A Subgroup Analysis of the ACRIN 6657/I-SPY 1 TRIAL HHS Public Access. *Tomography*. 2016;2:378–387. [PubMed: 28066808]
16. Rugo HS, Olopade OI, DeMichele A, et al. Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. *N Engl J Med*. 2016;375:23–34. [PubMed: 27406347]
17. Park JW, Liu MC, Yee D, et al. Adaptive Randomization of Neratinib in Early Breast Cancer. *N Engl J Med*. 2016;375:11–22. [PubMed: 27406346]
18. Hylton NM. Vascularity assessment of breast lesions with gadolinium-enhanced MR imaging. *Magn Reson Imaging Clin N Am*. 1999;7:411–20, x. [PubMed: 10382170]
19. Partridge SC, McDonald ES. Diffusion weighted magnetic resonance imaging of the breast: protocol optimization, interpretation, and clinical applications. *Magn Reson Imaging Clin N Am*. 2013;21:601–624. [PubMed: 23928248]
20. Newitt DC, Zhang Z, Gibbs JE, et al. Test-retest repeatability and reproducibility of ADC measures by breast DWI: Results from the ACRIN 6698 trial. *J Magn Reson Imaging*. 10 2018.
21. Galbán CJ, Ma B, Malyarenko D, et al. Multi-Site Clinical Evaluation of DW-MRI as a Treatment Response Metric for Breast Cancer Patients Undergoing Neoadjuvant Chemotherapy. Schwarz AJ, ed. *PLoS One*. 2015;10:e0122151. [PubMed: 25816249]
22. Sharma U, Danishad KKA, Seenu V, Jagannathan NR. 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*. 2009;22:104–113. [PubMed: 18384182]
23. Partridge SC, Nissan N, Rahbar H, Kitsch AE, Sigmund EE. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. *J Magn Reson Imaging*. 2017;45:337–355. [PubMed: 27690173]

24. Newitt DC, Chenevert TL, Quarles CC, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. *J Med Imaging*. 2018; 5: 011003.
25. Li X, Abramson RG, Arlinghaus LR, et al. Multiparametric Magnetic Resonance Imaging for Predicting Pathological Response After the First Cycle of Neoadjuvant Chemotherapy in Breast Cancer. *Invest Radiol*. 2015; 50: 195–204. [PubMed: 25360603]
26. Pinker K, Bogner W, Baltzer P, 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*. 2014;49: 421–430. [PubMed: 24566292]
27. Fangberget A, Nilsen LB, Hole KH, 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*. 2011;21:1188–1199. [PubMed: 21127880]
28. Partridge SC, Gibbs JE, Lu Y, et al. MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. *AJR Am J Roentgenol*. 2005;184:1774–1781. [PubMed: 15908529]
29. Marinovich ML, Sardanelli F, Ciatto S, et al. Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI. *Breast*. 2012;21:669–677. [PubMed: 22863284]
30. Houssami N, Macaskill P, von Minckwitz G, Marinovich ML, Mamounas E. Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. *Eur J Cancer*. 2012;48:3342–3354. [PubMed: 22766518]
31. Colleoni M, Montagna E. Neoadjuvant therapy for ER-positive breast cancers. *Ann Oncol*. 2012;23(suppl 10):x243–x248. [PubMed: 22987970]
32. Keenan KE, Peskin AP, Wilmes LJ, et al. Variability and bias assessment in breast ADC measurement across multiple systems. *J Magn Reson Imaging*. 2016;44:846–855. [PubMed: 27008431]

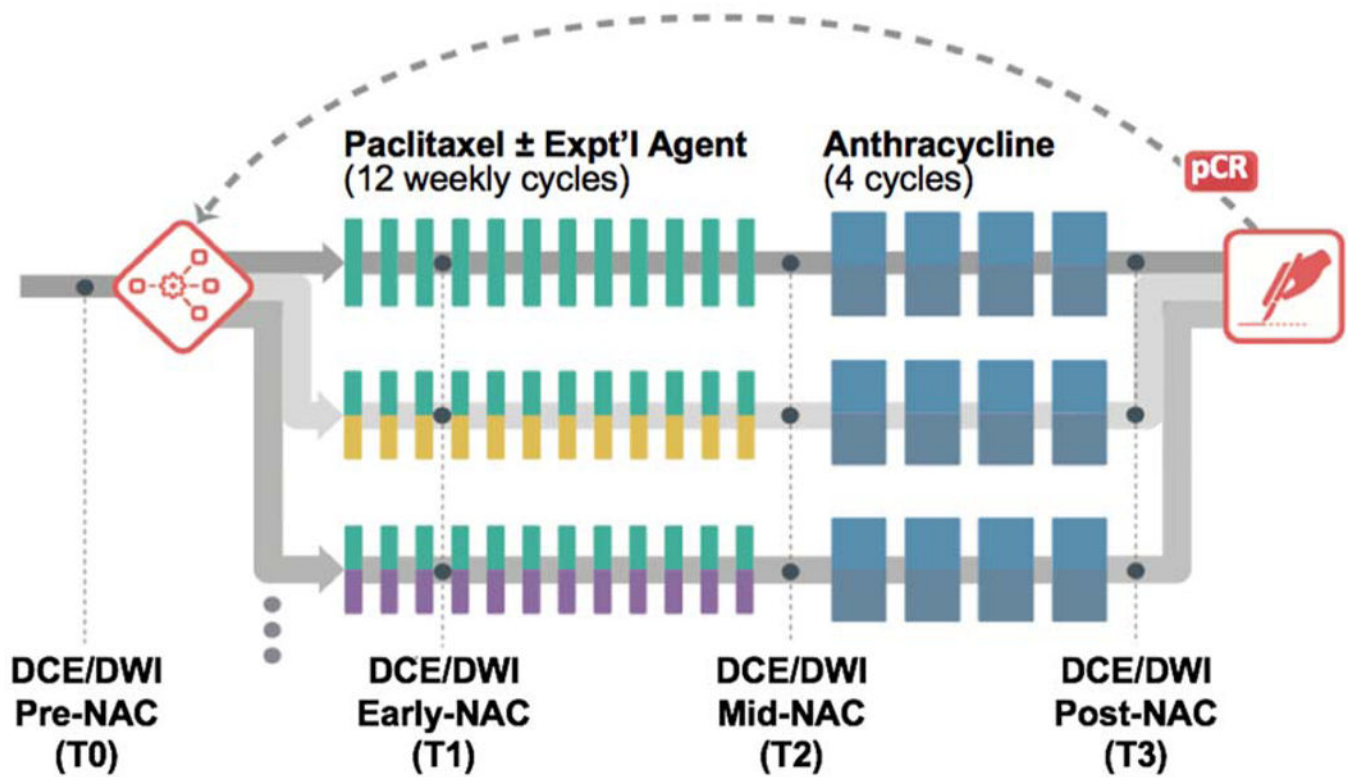


FIGURE 1.

I-SPY 2 study schema and adaptive randomization. Patients were randomized to the control (paclitaxel for HER2- or paclitaxel + trastuzumab for HER2+) or one of the experimental drug arms. Participants received a weekly dose of paclitaxel alone (control) or in combination with an experimental agent for 12 weekly cycles followed by four (every 2–3 weeks) cycles of anthracycline-cyclophosphamide (AC) prior to surgery.

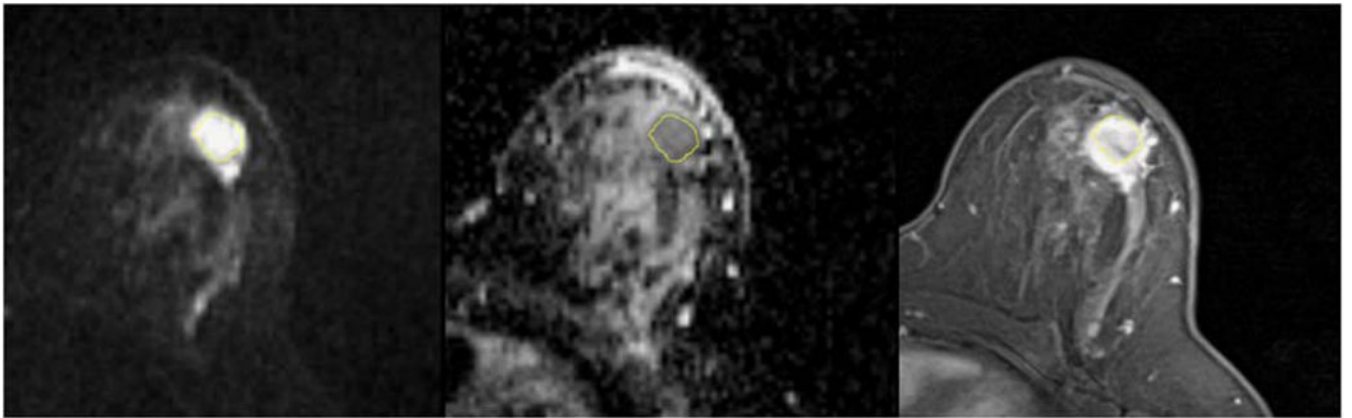


FIGURE 2.

ROI delineation in diffusion weighted MRI. Representative images were chosen from the same slice location in the axial view. The ROI was delineated on the ADC map (in the middle) to enclose the area that is hyper-intense in the $b = 800 \text{ s/mm}^2$ DW-MRI (on the left) and hypo-intense in the ADC map. The DCE-MRI is shown on the right to guide the location of the tumor

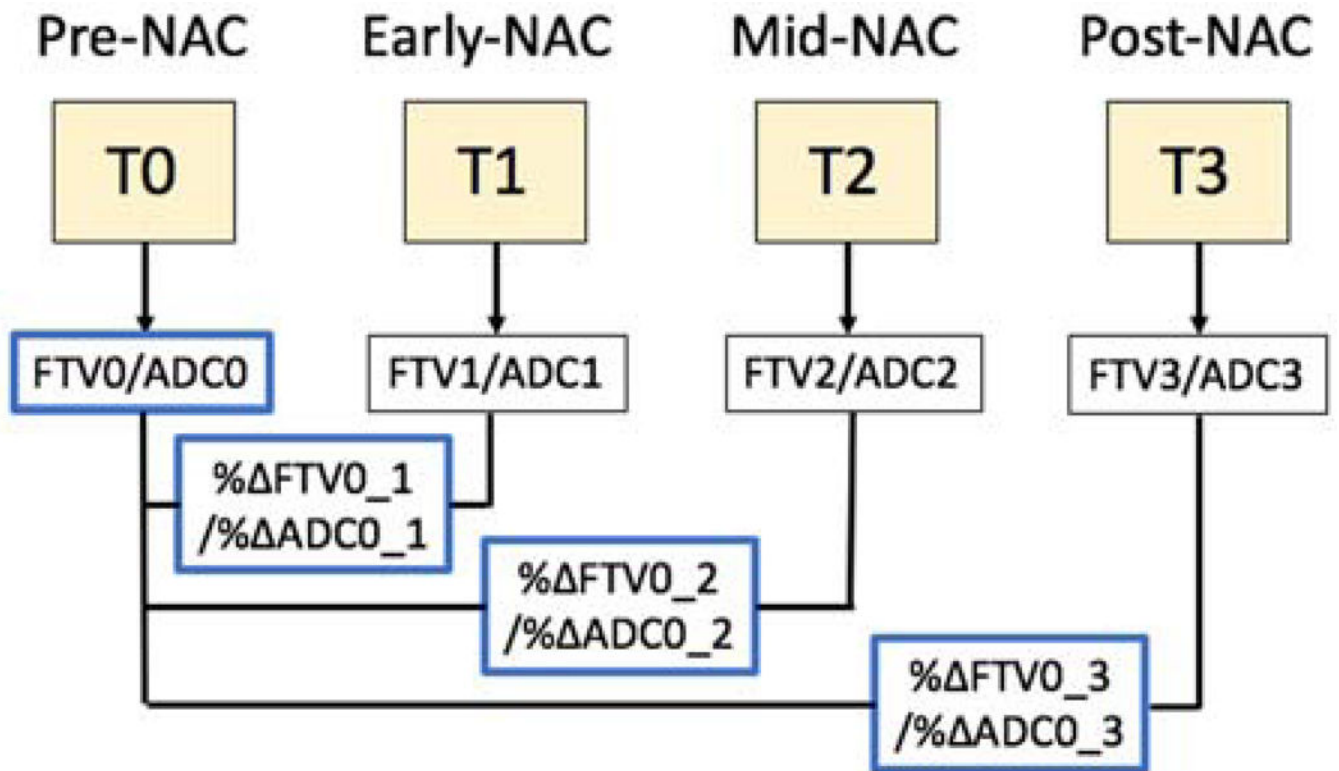
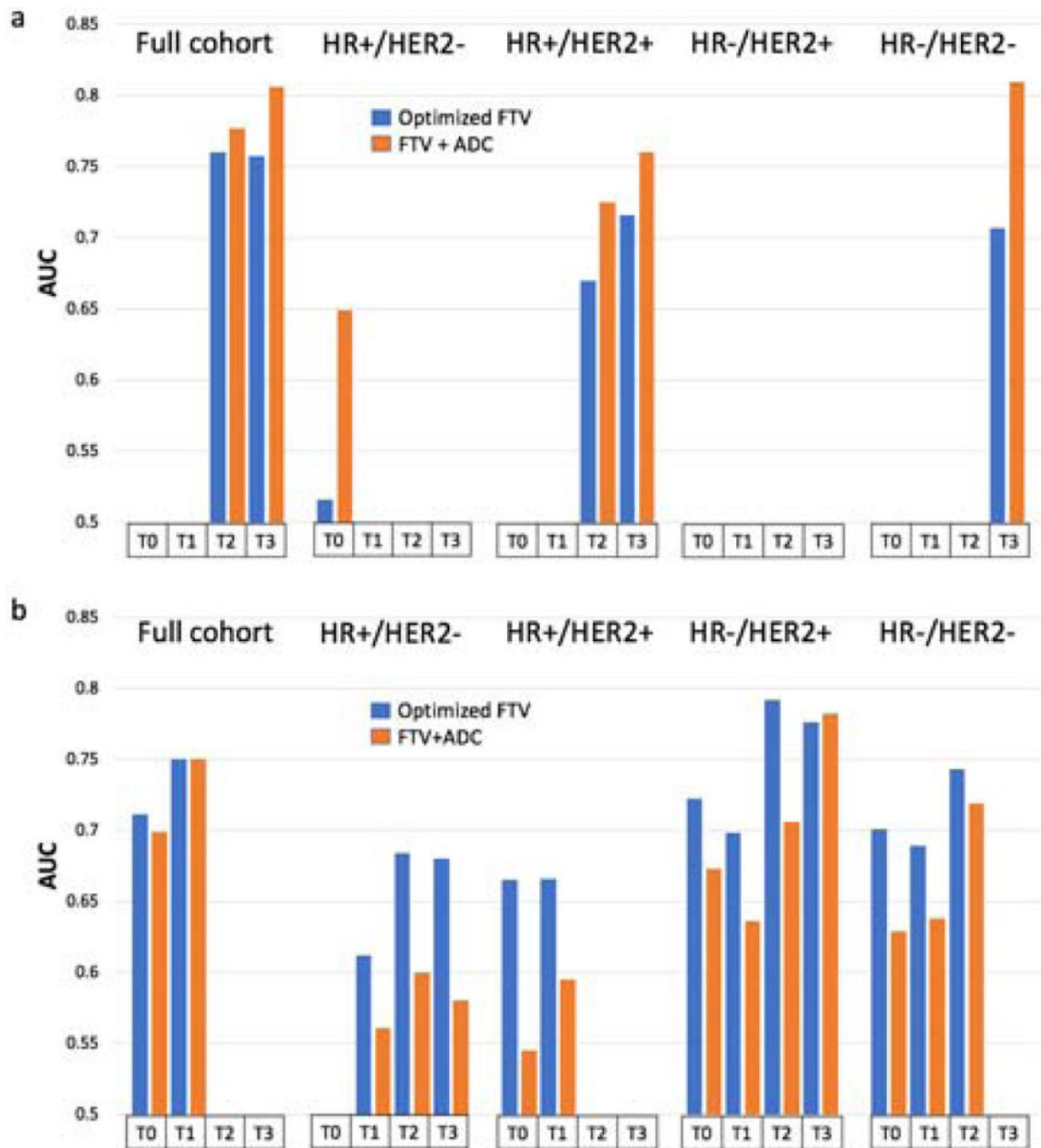


FIGURE 3. MR predictors calculated at multiple treatment time points. Predictors in bold frame were included in the analysis.

**FIGURE 4.**

AUCs for the optimized models with FTV predictors only and the same FTV predictors plus ADC predictors. The plots were generated using the full cohort and by HR/HER2 subtype labeled at the top of each subfigure. Within each cohort, a pair of FTV only and FTV + ADC are plotted at each treatment time point: T0 (pre-NAC), T1 (early-NAC), T2 (mid-NAC), and T3 (post-NAC). Subfigures show time points and cohorts where: (a) AUC increased for FTV + ADC compared to FTV only and (b) AUC did not increase or decreased.

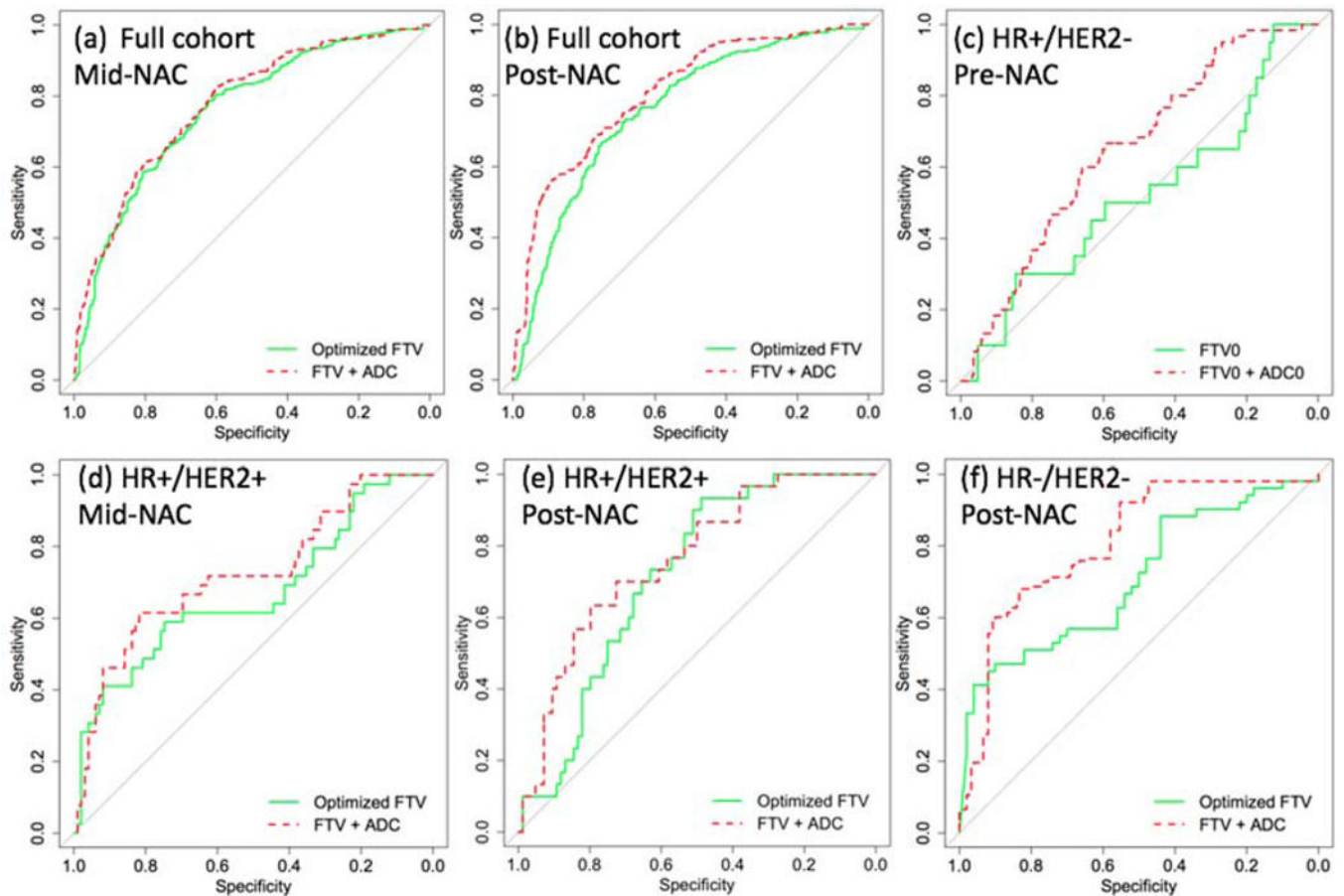


FIGURE 5.

The comparison of ROC curves for logistic regression models with vs. without ADC. In each subfigure, two pairs of ROC curves were plotted. “Optimized FTV” refers to the ROC curve generated for the model with FTV predictors only (FTV plus subtype for the full cohort) that had the highest AUC among models with all combinations of FTV predictors available upon specified treatment time point. “FTV + ADC” refers to the ROC curve for the extended model with ADC predictors added to the “Optimized FTV” model. (a) Full cohort at mid-NAC; (b) Full cohort at post-NAC; (c) HR+/HER2- at pre-NAC. Since there is no optimized model, ROC curves were marked as “FTV0” and “FTV0 + ADC0”; (d) HR+/HER2+ at mid-NAC; (e) HR+/HER2+ at post-NAC; (f) HR-/HER2- at post-NAC.

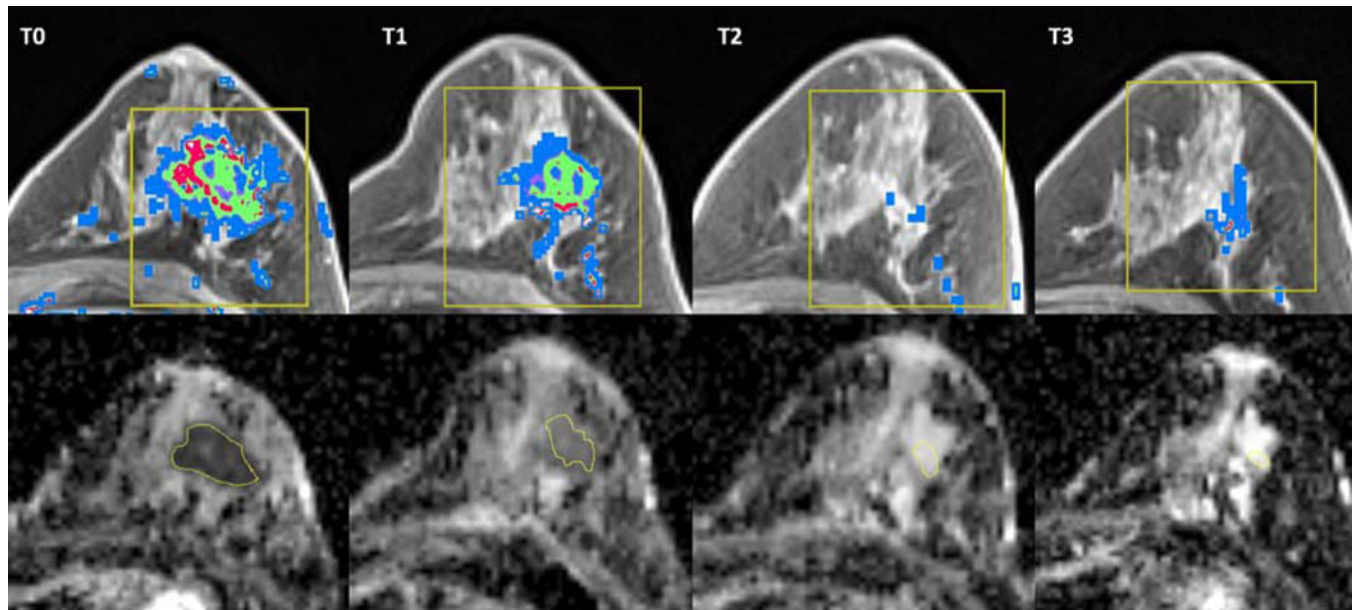


FIGURE 6.

Example MR images of an I-SPY 2 patient who achieved pCR after the neoadjuvant chemotherapy. The patient was 33 years old when diagnosed with a triple negative (HR-/HER2-) breast cancer. The top row shows representative slices of her DCE-MRIs at all 4 treatment time points. At each time point, the displayed slice is chosen from the volume acquired at early enhancement (137s after contrast injection), superimposed by the tumor voxels (in blue, green and red) identified by PE threshold. Pre-NAC FTV was 39cc and % FTV was -57.1% at T1, -93.1% at T2, and -92.5% at T3. The bottom row shows representative slices of her ADC maps at matching treatment time points, superimposed by the manually traced tumor ROI. Pre-NAC mean tumor ADC was $0.804 \times 10^{-3} \text{ mm}^2/\text{sec}$ and % ADC was 60.8% at T1, 180.9% at T2, and 172.1% at T3. The optimized FTV only model to predict pCR at T3 for triple negative cancer is: $y = -2.18 + (-0.246) * \% \text{ FTV0_3}$, based on which the probability of this patient to achieve pCR after NAC was 52.4%. The optimized model after adding ADC to the FTV only model is: $y = -9.17 + (-0.099) * \% \text{ FTV0_3} + 4.92 * \text{ADC0} + 0.47 * \% \text{ ADC0_3}$. Based on the new model, her probability to reach pCR increased to 87.5%.

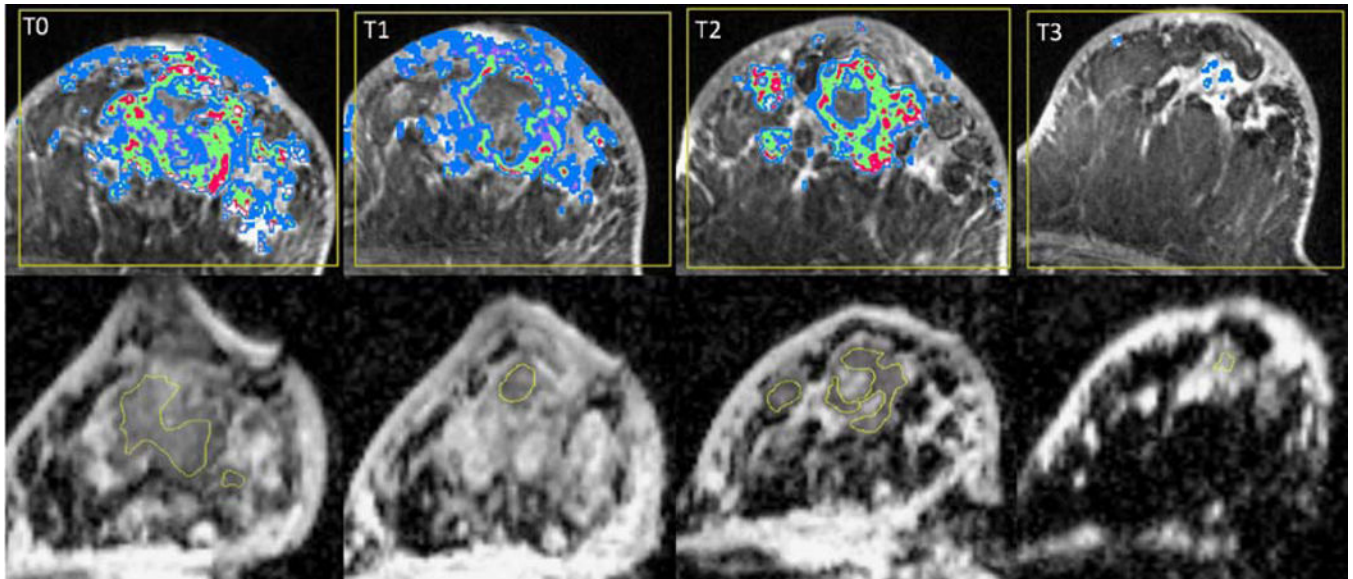


FIGURE 7.

Example MR images of an I-SPY 2 patient who did not achieve pCR after the neoadjuvant chemotherapy. The patient was 35 years old when diagnosed with a triple negative (HR-/HER2-) breast cancer. The top row shows representative slices of her DCE-MRIs at all 4 treatment time points. At each time point, the displayed slice is chosen from the volume acquired at early enhancement (134s after contrast injection), superimposed by the tumor voxels (in blue, green and red) identified by PE threshold. Pre-NAC FTV was 151.9cc and % FTV was -37.1% at T1, -57.7% at T2, and -98.8% at T3. The bottom row shows representative slices of her ADC maps at matching treatment time points, superimposed by the manually traced tumor ROI. Pre-NAC mean tumor ADC was $1.28 \times 10^{-3} \text{ mm}^2/\text{sec}$ and % ADC was -1.8% at T1, -10.6% at T2, and -6.0% at T3. The optimized FTV only model to predict pCR at T3 for triple negative cancer is: $y = -2.18 + (-0.246) * \% \text{ FTV0_3}$, based on which the probability of this patient to not achieve pCR after NAC was 43.8%. The optimized model after adding ADC to the FTV only model is: $y = -9.17 + (-0.099) * \% \text{ FTV0_3} + 4.92 * \text{ADC0} + 0.47 * \% \text{ ADC0_3}$. Based on the new model, her probability to non-pCR increased to 90.0%.

Table 1

Patient characteristics (n=354)

Characteristics	pCR (n=120)	Non-pCR (n=234)	<i>P</i> *
Age, median (range) – yr	49 (28–70)	50 (25–71)	0.46
Median tumor diameter by MRI (IQR) – cm	3.3 (2.5–4.4)	4.0 (3.0–5.6)	0.0018
Median tumor diameter by clinical exam (IQR) – cm	4.0 (3.0–5.5)	5.0 (3.5–6.0)	0.0012
Race – no. (%)			0.84
Asian	9 (7.5)	14 (6.0)	
Black or African American	15 (12.5)	30 (12.8)	
Native Hawaiian or Pacific Islander	2 (1.7)	2 (0.9)	
White	93 (77.5)	187 (79.9)	
Mix race	1 (0.8)	1 (0.4)	
Ethnicity – no. (%)			0.59
Hispanic or Latino	14 (11.7)	23 (9.8)	
Not Hispanic or Latino	106 (88.3)	211 (90.2)	
Menopausal status – no. (%)			1.00
Premenopausal	62 (51.7)	121 (51.7)	
Perimenopausal	2 (1.7)	5 (2.1)	
Postmenopausal	41 (34.2)	78 (33.3)	
Not applicable	14 (11.7)	27 (11.5)	
Unknown	1 (0.8)	3 (1.3)	
HR/HER2 subtype – no. (%)			<0.0001
HR+/HER2-	21 (17.5)	107 (45.7)	
HR+/HER2+	18 (15.0)	42 (17.9)	
HR-/HER2+	20 (16.7)	14 (0.6)	
HR-/HER2- (Triple negative)	61 (50.8)	71 (30.3)	
Node status – no. (%)			0.48
Palpable	47 (39.2)	104 (44.4)	
Nonpalpable	68 (56.7)	117 (50.0)	
Unknown	5 (4.2)	13 (5.6)	

* Wilcoxon *p* value was used for continuous variables and Fisher's exact test was used for categorical variables

Numbers in parentheses are range for age, interquartile range for tumor diameters, and percentage in pCR or non-pCR groups

Table 2.

Median of predictor values and their differences in pCR vs. non-pCR in the full cohort (n=354, pCR rate: 34%)

Predictor	n	pCR*	Non-pCR*	Diff. (95% CI**)	AUC (95% CI)	P**
FTV0 (cc)	352	9.4 (5.3, 22.7)	19.5 (8.7, 38.1)	-5.7 (-9.2, -2.8)	0.63 (0.57, 0.69)	<0.0001
% FTV0_1 (%)	347	-59.6 (-81.4, -32.8)	-41.3 (-68.5, -13.1)	-16.5 (-24.2, -8.6)	0.63 (0.57, 0.70)	<0.0001
% FTV0_2 (%)	328	-95.7 (-98.7, -85.7)	-87.5 (-95.4, -62.7)	-5.9 (-9.2, -3.3)	0.67 (0.61, 0.74)	<0.0001
% FTV0_3 (%)	329	-98.3 (-99.9, -92.5)	-92.3 (-98.2, -81.7)	-3.5 (-5.6, -1.8)	0.70 (0.64, 0.76)	<0.0001
ADC0($\times 10^{-3}$ mm ² /sec)	348	1.04 (0.94, 1.12)	1.05 (0.94, 1.15)	-0.005 (-0.04, 0.03)	0.51 (0.45, 0.57)	0.79
% ADC0_1 (%)	328	19.3 (6.7, 34.7)	13.5 (5.2, 25.1)	4.5 (0.5, 8.7)	0.57 (0.51, 0.64)	0.03
% ADC0_2 (%)	302	57.2 (28.9, 87.7)	30.8 (12.1, 60.3)	23.3 (13.7, 32.7)	0.67 (0.60, 0.73)	<0.0001
% ADC0_3 (%)	301	84.3 (58.1, 107.3)	46.7 (19.1, 80.8)	35.1 (25.4, 44.7)	0.72 (0.66, 0.78)	<0.0001

* Values were given as median (IQR)

** 95% CI and P were calculated by Wilcoxon Rank Sum test

Table 3.

Comparison of AUCs of optimized models with FTV predictors only and with additional ADC predictors

Patient cohort	Visit	n	pCR rate (%)	AUC (95% CI)		
				Single FTV predictor	Optimized FTV	FTV + ADC
Full cohort	T0	346	34	0.71 (0.68, 0.75)	0.71 (0.68, 0.75)	0.70 (0.66, 0.73)
	T1	323	33	0.71 (0.68, 0.75)	0.75 (0.72, 0.78)	0.75 (0.72, 0.78)
	T2	282	33	0.73 (0.69, 0.76)	0.76 (0.73, 0.79)	0.78 (0.74, 0.81)
	T3	257	34	0.72 (0.68, 0.76)	0.76 (0.72, 0.79)	0.81 (0.77, 0.84)
HR+/HER2-	T0	124	16	0.52 (0.37, 0.66)	0.52 (0.37, 0.66)	0.65 (0.51, 0.69)
	T1	116	14	0.61 (0.46, 0.76)	0.61 (0.46, 0.76)	0.56 (0.48, 0.65)
	T2	101	13	0.68 (0.51, 0.86)	0.68 (0.51, 0.86)	0.60 (0.51, 0.69)
	T3	94	14	0.68 (0.51, 0.85)	0.68 (0.51, 0.85)	0.58 (0.49, 0.67)
HR+/HER2+	T0	58	31	0.67 (0.50, 0.83)	0.67 (0.50, 0.83)	0.55 (0.44, 0.65)
	T1	52	33	0.65 (0.47, 0.82)	0.67 (0.49, 0.84)	0.60 (0.49, 0.70)
	T2	46	28	0.58 (0.38, 0.77)	0.67 (0.56, 0.78)	0.73 (0.63, 0.83)
	T3	38	26	0.58 (0.36, 0.79)	0.72 (0.62, 0.81)	0.76 (0.66, 0.86)
HR-/HER2+	T0	33	58	0.72 (0.54, 0.90)	0.72 (0.54, 0.90)	0.67 (0.57, 0.78)
	T1	28	57	0.61 (0.39, 0.84)	0.70 (0.50, 0.90)	0.64 (0.52, 0.76)
	T2	25	56	0.79 (0.60, 0.98)	0.79 (0.60, 0.98)	0.71 (0.58, 0.83)
	T3	24	54	0.72 (0.50, 0.94)	0.78 (0.58, 0.98)	0.78 (0.67, 0.90)
HR-/HER2-	T0	131	47	0.70 (0.61, 0.79)	0.70 (0.61, 0.79)	0.63 (0.57, 0.69)
	T1	127	46	0.63 (0.53, 0.73)	0.69 (0.60, 0.78)	0.64 (0.58, 0.69)
	T2	110	47	0.70 (0.60, 0.80)	0.74 (0.69, 0.80)	0.72 (0.66, 0.77)
	T3	101	50	0.71 (0.61, 0.81)	0.71 (0.61, 0.81)	0.81 (0.76, 0.86)