diagnostics

MDPI

Article

MRI-Based Models for Personalizing Neoadjuvant Treatment in Breast Cancer: A Validation Study

Wen Li [1,*], Natsuko Onishi [1], John Kornak [2], Teffany Joy Bareng [1], Nu N. Le [1], Pouya Metanat [1], Matthew Gibbons [1], Lisa J. Wilmes [1], Elissa Price [1], Bonnie N. Joe [1,*], Savannah C. Partridge [3], Christina Yau [4], Denise M. Wolf [5], Mark J. Magbanua [5], Sara Venters [5], W. Fraser Symmans [6], Jane Perlmutter [7], Barbara LeStage [7], on behalf of the I-SPY 2 Imaging Working Group, on behalf of the I-SPY 2 Investigators, Douglas Yee [8], Hope S. Rugo [9], Laura J. van't Veer [5], Angela M. DeMichele [10], Laura J. Esserman [4] and Nola M. Hylton [1]

1 Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA 94158, USA; natsuko.onishi@ucsf.edu (N.O.); teffanyjoy.bareng@ucsf.edu (T.J.B.); nu.le@ucsf.edu (N.N.L.); pouya.metanat@ucsf.edu (P.M.); matthew.gibbons@ucsf.edu (M.G.); lisa.wilmes@ucsf.edu (L.J.W.); elissa.price@ucsf.edu (E.P.); nola.hylton@ucsf.edu (N.M.H.)
2 Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94158, USA; john.kornak@ucsf.edu
3 Department of Radiology, University of Washington, Seattle, WA98195, USA; scp3@uw.edu
4 Department of Surgery, University of California, San Francisco, CA 94158, USA; hoisze.yau@ucsf.edu (C.Y.); laura.esserman@ucsf.edu (L.J.E.)
5 Department of Laboratory Medicine, University of California, San Francisco, CA 94158, USA; denise.wolf@ucsf.edu (D.M.W.); mark.magbanua@ucsf.edu (M.J.M.); sara.venters@ucsf.edu (S.V.); laura.vantveer@ucsf.edu (L.J.v.V.)
6 Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; fsymmans@mdanderson.org
7 I-SPY 2 Advocacy Group, San Francisco, CA 94158, USA; janep@gemini-grp.com (J.P.); blestage51@gmail.com (B.L.)
8 Department of Medicine, University of Minnesota, Minneapolis, MN 55455, USA; yeexx006@umn.edu
9 Division of Breast Medical Oncology, City of Hope National Medical Center, Duarte, CA 91010, USA; hrugo@coh.org
10 Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; angela.demichele@pennmedicine.upenn.edu
* Correspondence: wen.li@ucsf.edu (W.L.); bonnie.joe@ucsf.edu (B.N.J.); Tel.: +1-415-885-3764 (W.L.); +1-415-885-7464 (B.N.J.)

Abstract

Background/Objectives: Robust, validated, subtype-specific MRI-based predictive models are needed to predict breast cancer treatment response accurately and in a timely manner. This study aims to validate such models using data collected from a large, multicenter I-SPY 2 trial to enable personalized neoadjuvant therapy and optimize individual patient outcomes. **Methods**: This study included 2117 patients enrolled between May 2010 and June 2022 in the I-SPY 2 trial. The full cohort was randomly divided into training and test sets in a 70/30 ratio. Subtype-specific MRI prediction models were developed using functional tumor volume (FTV) derived from the dynamic contrast-enhanced MRI acquired at baseline (T0), early treatment (T1), and inter-regimen (T2). The endpoint for prediction was pathologic complete response (pCR), assessed at surgery. The area under the receiver operating characteristic curve (AUC) was used to evaluate the overall performance of the model, and sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were evaluated for the optimal cut point selected by the Youden Index. Our analysis was subtype-specific, with breast cancer subtypes defined by hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) statuses. **Results**: FTV variables included in the optimal models differed by subtype: FTV percent change at T1 was selected for HR−/HER2+ and HR+/HER2+ cancers, and FTV

Check for updates

Academic Editor: Takouhie Catherine Maldjian

Received: 15 October 2026
Revised: 6 November 2026
Accepted: 11 November 2026
Published: 24 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

percent change at T2 for HR+/HER2− and HR−/HER2− (triple-negative) cancers. Performance was similar in the training and test cohorts. AUCs estimated in the test set were HR+/HER2−: 0.68 [95% confidence interval: 0.59, 0.77], HR+/HER2+: 0.71 [0.59, 0.82], HR−/HER2+: 0.65 [0.46, 0.84], HR−/HER2−: 0.74 [0.67, 0.81]. Optimal cut points selected for individual subtypes also resulted in similar sensitivities, specificities, PPVs, and NPVs. **Conclusions**: We developed and validated breast cancer subtype-specific MRI models to predict pCR using longitudinal FTV measurements before and during neoadjuvant therapy. Our findings indicate that the optimal MRI time point for prediction varies by cancer subtype, with early-treatment FTV change showing the greatest predictive value for HER2+ breast cancer, whereas inter-regimen FTV change shows the greatest predictive value for HR+/HER2− and TNBC.

Keywords: magnetic resonance imaging; dynamic-contrast enhanced; functional tumor volume; breast cancer; neoadjuvant therapy; pathologic complete response

1. Introduction

Breast cancer is one of the most extensively studied malignancies, leading to the development of numerous prediction models aimed at assessing treatment outcomes. In patients with early-stage breast cancer, neoadjuvant chemotherapy (NAC) is increasingly tailored according to tumor biology, including hormone receptor (HR) status, human epidermal growth factor receptor 2 (HER2) status, and other molecular features [1–3]. The goal of personalizing NAC based on tumor biology is to maximize treatment response and to identify patients who may benefit from therapy escalation or de-escalation early in treatment.

The increasing adoption of personalized NAC has highlighted the need for accurate and robust prediction of treatment response in breast cancer. The ongoing I-SPY 2.2 clinical trial has introduced the ability to adjust the amount of therapy based on response, highlighting the need to accurately and robustly predict response during the course of therapy. Magnetic resonance imaging (MRI) is considered the most accurate imaging method for evaluating tumor response during NAC [4]. Dynamic contrast-enhanced (DCE) MRI is the most widely used and clinically adopted MRI modality for monitoring response to NAC because it provides high-resolution anatomical and functional information of breast tumors. Functional tumor volume (FTV), derived from DCE-MRI, quantifies the volume of enhancing tumor tissue based on predefined enhancement thresholds [5]. The I-SPY 1/American College of Radiology Imaging Network (ACRIN) 6657 trial found that changes in FTV during NAC, especially at the early treatment time point, are strongly associated with pathologic complete response (pCR) and long-term outcomes [6,7]. Building on these findings, I-SPY 2 has collected high-quality DCE-MRI exams at multiple time points and longitudinal FTV measurements for each patient [8–10].

Numerous MRI-based imaging biomarkers and predictive models have been developed to estimate treatment response [11,12], particularly to predict pCR and, in recent years, to predict response for specific breast cancer subtypes [13–15]. However, many published MRI models are retrospective, single-institution, or based on relatively small cohorts. In contrast, subtype-specific MRI models, developed and validated across all subtypes in a large-scale, multicenter clinical trial, are rare. Previously, subtype-specific MRI-based predictive models were developed to predict pathologic complete response (pCR). The models and the previously determined probability thresholds have been part of the therapy de-escalation algorithm in I-SPY 2 (completed in June 2022) and I-SPY 2.2 (ongoing).

The primary objective of this study was to validate subtype-specific models based on longitudinal FTV measurements for predicting pCR and to determine the treatment time point at which FTV provides the greatest predictive information for each breast cancer subtype. The study evaluated FTV as an imaging biomarker rather than developing a comprehensive prediction model incorporating clinical, pathologic, molecular, and imaging variables. The validation consisted of two analyses. The primary analysis was to develop the models using the prespecified training cohort by following the method of our previous work [10] and subsequently validate them without refitting in a held-out test cohort. The secondary analysis was to re-develop the models using the entire cohort.

2. Materials and Methods

2.1. Study Population and Eligibility

This retrospective study consisted of the complete cohort of the I-SPY 2 trial (www.ClinicalTrials.gov (accessed on 10 November 2026); registration number: NCT01042379), an adaptive phase II neoadjuvant breast cancer study that uses molecular signatures to test novel agents in combination with standard chemotherapy. This cohort included 2117 patients enrolled between May 2010 and June 2022 across 35 participating sites. Eligibility for enrollment generally included adult women ≥ 18 years old with histologically confirmed invasive breast cancer, clinically or radiologically measurable disease in the breast (typically ≥ 2.5 cm), clinical stage II or III (or equivalent criteria) with no distant metastases, and molecularly high-risk disease if hormone receptor (HR)-positive [16,17]. HR and HER2 (human epidermal growth factor receptor 2) status (negative versus positive) were assessed prior to treatment initiation. HR-positive (HR+) was defined as 5% or more staining for either estrogen or progesterone receptor. HER2-positive (HER2+) was defined as 3+ by immunohistochemistry or FISH + (fluorescence in situ hybridization) per the College of American Pathologists (CAP) guideline. All patients received a 12-week course of paclitaxel-based chemotherapy (80 mg/m^2 intravenous paclitaxel weekly), with or without an experimental agent (Supplemental Table S1). Patients with HER2+ breast cancer also received intravenous trastuzumab (4 mg/kg load in week 1, followed by 2 mg/kg weekly × 11 weeks) with the weekly paclitaxel. All patients then received four cycles of doxorubicin/cyclophosphamide (AC) except for the last 171 patients where there was an option to proceed to surgery early without AC if a complete response was predicted. See Figure 1 for the study schema [18,19]. The primary endpoint is pCR, defined as no invasive cancer in the breast and lymph nodes at the time of surgery [17,20]. All patients signed written informed consent, and all participating sites received approval from their local human study institutional review boards (IRBs) before they started recruiting patients.

2.2. MRI Acquisition

DCE-MRI was performed by acquiring a series of three-dimensional fat-suppressed T1-weighted images bilaterally. See Supplemental Table S2 for the detailed I-SPY 2 image acquisition protocol, which was standardized based on clinical breast MR imaging, our experience in the I-SPY 1 and 2 trials, and in compliance with the American College of Radiology (ACR) guidance [6,7]. To summarize, the imaging parameters were TR = 4–10 msecs, minimum TE, flip angle = 10–20°, in-plane resolution ≤ 1.4 mm, slice thickness ≤ 2.5 mm, temporal resolution = 80–100 s, field of view (FOV) = 260–360 mm to achieve full bilateral coverage, acquisition matrix = 384–512, and axial orientation. A gadolinium-based contrast agent was administered intravenously at a dose of 0.1 mmol/kg using an injection rate of 2 mL/s, followed by a 20 mL saline flush. The pre-contrast and multiple post-contrast acquisitions were repeated using identical sequence parameters, and post-contrast imaging continued for at least 8 min after the contrast agent injection.

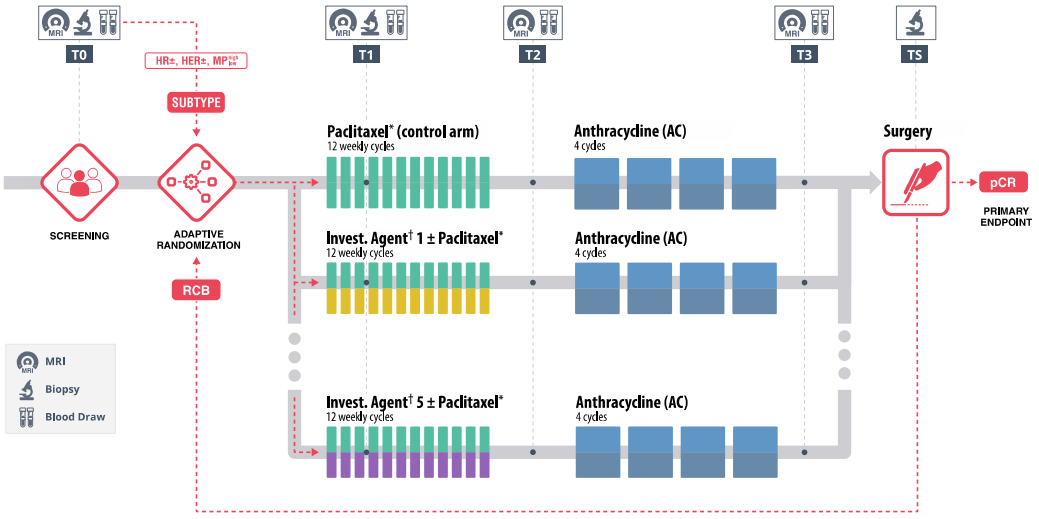


Figure 1. I-SPY 2 study schema. I-SPY 2 is a phase II adaptive platform trial that uses Bayesian adaptive randomization to assign patients to receive standard neoadjuvant chemotherapy alone or in combination with one of several investigational agents. After the first regimen, every patient received anthracycline (AC). MRI scans were performed at T0: baseline and before the treatment started, T1: early treatment, approximately 3 weeks after the treatment started, T2: inter-regimen, approximately 12 weeks after the treatment started, and T3: pre-surgery, after the completion of AC and prior to the surgery. This study validated MRI-based models to predict pathologic complete response at T2. Solid gray arrows indicate progression through treatment to surgery. Red dashed arrows depict the biomarker-based treatment allocation decision point. Green bars represent paclitaxel, yellow and purple bars distinguish investigational agents, and blue blocks represent anthracycline-based chemotherapy. * HER2-positive patients may additionally receive trastuzumab (Herceptin). † One or more investigational agents may replace some or all standard therapy.

MRI exams were performed at four time points: baseline (T0), early treatment (3 weeks, T1), inter-regimen (12 weeks, T2), and pre-surgery (T3) (see Figure 1). The same image acquisition protocol, magnet configuration (vendor, field strength, breast coil), and contrast agent brand were required for the acquisition of all exams for the same patient.

2.3. FTV Variables

FTV is a standardized imaging biomarker in the I-SPY trials [6,7,9] and was obtained from DCE-MRI using semiautomated in-house software developed in IDL (Interactive Data Language, version 8.4; NV5 Geospatial Software, Inc., Broomfield, CO, USA) at all four time points for all patients. FTV was calculated as the sum of voxel volumes within the 3-dimensional (3D) volume of interest (VOI). The VOI was delineated manually on the subtracted image between the pre-contrast and early enhancement phases (acquired approximately 2.5 min post-contrast) by a site radiologist or a trained imaging coordinator. Voxels within the VOI that exceeded a predefined percentage enhancement (PE) threshold and a signal enhancement ratio (SER) threshold were included in the FTV calculation, where $PE = \frac{S_1 - S_0}{S_0} \times 100\%$ and $SER = \frac{S_1 - S_0}{S_2 - S_0}$, S0, S1, and S2 were signal intensities at pre-contrast, early enhancement, and late enhancement (at approximately 7.5 min post-contrast), respectively. MRI-based models were built at T2 to predict pathological outcome using FTV variables available at T2: FTV at T0 (baseline FTV: FTV0), the early percent change in FTV from T0 to T1 ($\Delta FTV0_1 = \frac{FTV1 - FTV0}{FTV0} \times 100\%$), and the inter-regimen percent change in FTV from T0 to T2 ($\Delta FTV0_2 = \frac{FTV2 - FTV0}{FTV0} \times 100\%$).

2.4. Statistical Analysis

For patient characteristics, age was reported using mean ± standard deviation (SD), and other categorical characteristics, such as subtype (HR and HER2), treatment (experimental versus control), pCR status, and residual cancer burden (RCB), were reported using the number of patients and percentage within the cohort. The p-value for the age difference between the training and test sets was acquired using a two-sided Student's t-test, while those for differences in categorical characteristics were generated using Fisher's exact test for binary variables, and Fisher-Freeman-Halton test was used for variables with more than two categories. Fisher's exact test was also used to compare the difference between two confusion matrices when the same cut point was applied to the training and test sub-cohorts.

We split the entire cohort randomly into the training (70%) and test (30%) sets. Subtype-specific MRI-based models for predicting pCR were developed using single- and multivariable logistic regression within the training set, as previously described [10]. To obtain robust estimates of model performance and reduce overfitting, repeated 10-fold cross-validation was performed within the training set, with the 10-fold cross-validation procedure repeated 10 times using different random partitions of the training data. The overall cross-validated (CV-) AUC was summarized across the repeated cross-validation runs. The 95% confidence interval (CI) for CV-AUC was estimated by bootstrapping with 1000 replicates. Candidate models included all possible single- and multivariable combinations of FTV0, $\Delta FTV0_1$, and $\Delta FTV0_2$. Because simple single-predictor models were preferred for interpretability and potential clinical implementation, when the model with the highest CV-AUC contained multiple FTV predictors, we additionally compared the individual FTV predictors within the training set and selected the single-predictor model with the highest CV-AUC for validation. All model and predictor selection was performed using the training set without consideration of performance in the test set. The selected subtype-specific models were then fit using the complete training set and applied without further feature selection or refitting to the held-out test set to generate predicted probabilities. In the test set, AUCs were calculated using the probabilities and 95% CIs were calculated using the DeLong method [21]. We assessed differences between the two AUCs using a two-sided bootstrap test.

Youden Index–based cut points were determined from the ROC curves of the models fit in the training set and subsequently applied unchanged to the test set. With pCR considered the positive outcome and non-pCR the negative outcome, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were estimated in both cohorts. The purpose of this analysis was to evaluate whether classification performance at a fixed, training-derived threshold was maintained in the held-out test set, as a complementary assessment of model generalizability. The Youden Index was used as a standardized statistical criterion for this validation analysis and was not intended to define a clinically optimal threshold for treatment escalation or de-escalation. Selection of clinically actionable thresholds would require consideration of the intended clinical application and the relative consequences of false-positive and false-negative predictions.

Following completion of the training/test validation analysis, we conducted a separate secondary analysis in which the model-development procedure was repeated using the full cohort (training and test sets combined). The purpose of this analysis was to derive subtype-specific models using all available data for future prospective evaluation. This full-cohort analysis was performed after the training/test validation and was not used to select, modify, or evaluate the models in the held-out test set; consequently, no test-set performance estimates are reported for models developed using the full cohort.

All statistical tests were two-sided, with a significance level of 0.05. Analyses were performed using R version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria).

3. Results

3.1. Patient Population

From May 2010 to June 2022, 2117 patients were enrolled and randomized to treatment in the I-SPY 2 trial across 35 sites. Patients with missing HR or HER2 status ($n = 10$), missing pCR status ($n = 88$), and missing FTV measurements ($n = 248$) at T0 ($n = 0$), T1 ($n = 88$), or T2 ($n = 160$) were excluded. The final full cohort included 1771 patients: 1227 in the training set and 544 in the test set. See Figure 2 for the data inclusion and exclusion flowchart. Table 1 shows the patient characteristics in the full cohort and in the training/test sets separately. *p*-values did not flag any statistically significant differences between the training and test sets. Supplemental Table S3 shows the number of patients accrued in I-SPY 2, the number of patients included in this analysis, the field strength, and the vendor for MRI scanners by site.

Table 1. Patient characteristics.

	Full $n = 1771$	**Training $n = 1227$**	**Test $n = 544$**	***p* (Training vs. Test)**
Age (y/o, mean ± SD)	48.5 ± 11.3	48.8 ± 11.5	48.1 ± 10.9	0.23
HR/HER2 subtype (n [%])				0.82
HR+/HER2−	748 (42%)	527 (43%)	221 (41%)	
HR+/HER2+	275 (16%)	189 (15%)	86 (16%)	
HR−/HER2+	133 (8%)	90 (7%)	43 (8%)	
HR−/HER2−	615 (35%)	421 (34%)	194 (36%)	
Clinical T Stage (n [%])				0.20
T0	1 (0%)	1 (0%)	0 (0%)	
T1	50 (3%)	39 (3%)	11 (2%)	
T2	1165 (66%)	802 (65%)	363 (67%)	
T3	466 (26%)	327 (27%)	139 (26%)	
T4	58 (3%)	42 (3%)	16 (3%)	
Unknown	31 (2%)	16 (1%)	15 (3%)	
Clinical N Stage (n [%])				0.43
N0	789 (45%)	556 (45%)	233 (43%)	
N1	755 (43%)	514 (42%)	241 (44%)	
N2	83 (5%)	58 (5%)	25 (5%)	
N3	94 (5%)	69 (6%)	25 (5%)	
Unknown	50 (3%)	30 (2%)	20 (4%)	
Treatment (n [%])				0.19
Experimental	1268 (72%)	890 (73%)	378 (69%)	
Control	503 (28%)	337 (27%)	166 (31%)	
Pathologic complete response (pCR)				0.83
pCR	660 (37%)	455 (37%)	205 (38%)	
Non-pCR	1111 (63%)	772 (63%)	339 (62%)	
Residual cancer burden (RCB) class				0.50
RCB-0 (pCR)	660 (37%)	455 (37%)	205 (38%)	
RCB-I	253 (14%)	182 (15%)	71 (13%)	
RCB-II	606 (34%)	424 (35%)	182 (33%)	
RCB-III	252 (14%)	166 (14%)	86 (16%)	

SD: standard deviation. HR: hormone receptor. HER2: HER2: human epidermal growth factor receptor 2. pCR: pathologic complete response. RCB: residual cancer burden.

3.2. Validation of FTV-Based Models

3.2.1. Primary Training/Test Validation

MRI-based models were developed using individual HR/HER2 sub-cohorts in the training set. Based on achieving the highest AUC, the selected model for HR+/HER2+ included $\Delta FTV0_1$ only, the HR−/HER2+ model included $\Delta FTV0_1$ only, and the triple negative model included $\Delta FTV0_2$ only (Supplemental Table S4). For the HR+/HER2− subtype, the model with the highest cross-validated AUC included FTV0, $\Delta FTV0_1$, and $\Delta FTV0_2$, whereas the models selected for the other three subtypes each contained a single FTV predictor. Because we prioritized simple and clinically interpretable single-predictor

models, we additionally compared the three individual FTV predictors for HR+/HER2− using the training data. Supplemental Table S5 and Figure S1 show comparisons of the optimal model and single-variable models for HR+/HER2− in the training set. Among the single-predictor models, $\Delta FTV0_2$ achieved the highest cross-validated AUC and was therefore selected as the HR+/HER2− model for validation in the held-out test set. This model-selection decision was made without consideration of test-set performance.

Table 2 compares model performance in the training and test sets. All models achieved consistent AUCs in the test set, compared to the model performance observed in the training set. See Figure 3a–d for ROC curves estimated in the training and test sets for HR+/HER2−, HR+/HER2+, HR−/HER2+, and TNBC, respectively. For the HR+/HER2− subtype, the AUC was 0.68 (95% CI: 0.59, 0.77). For HR+/HER2+, the AUC was 0.71 (95% CI: 0.59, 0.82). For HR−/HER2+, the AUC was 0.65 (95% CI: 0.46, 0.84). For TNBC, the AUC was 0.74 (95% CI: 0.67, 0.81).

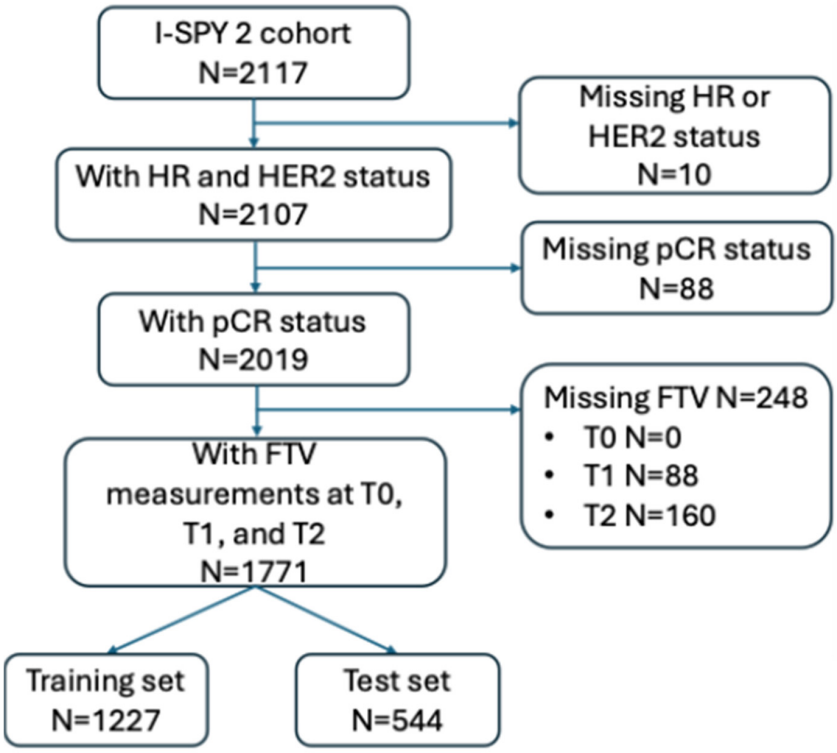


Figure 2. Data inclusion and exclusion flowchart. HR: hormone receptor. HER2: human epidermal growth factor receptor 2. pCR: pathologic complete response. FTV: functional tumor volume. T0: baseline. T1: early-treatment time point. T2: inter-regimen time point. Arrows indicate patient flow through the inclusion and exclusion steps and the division of the final cohort into training and test sets.

Table 2. Comparison of model performances in the training/test sets.

Subtype	Training Set		Test Set		*p*
	N (pCR Rate, %)	CV-AUC (95% CI)	*N* (pCR Rate, %)	AUC (95% CI)	
HR+/HER2−	527 (22%, 117/527)	0.70 (0.64, 0.76)	221 (23%, 50/221)	0.68 (0.59, 0.77)	0.64
HR+/HER2+	198 (44%, 88/198)	0.68 (0.57, 0.75)	90 (40%, 36/90)	0.71 (0.59, 0.82)	0.60
HR−/HER2+	92 (71%, 65/92)	0.73 (0.60, 0.85)	43 (74%, 32/43)	0.65 (0.46, 0.84)	0.39
HR−/HER2−	437 (46%, 200/437)	0.76 (0.72, 0.80)	201 (46%, 92/201)	0.74 (0.67, 0.81)	0.53

HR: hormone receptor, HER2: human epidermal growth factor receptor 2, pCR: pathologic complete response, AUC: area under the receiver operating characteristic curve, CI: confidence interval, CV-AUC: cross-validated AUC.

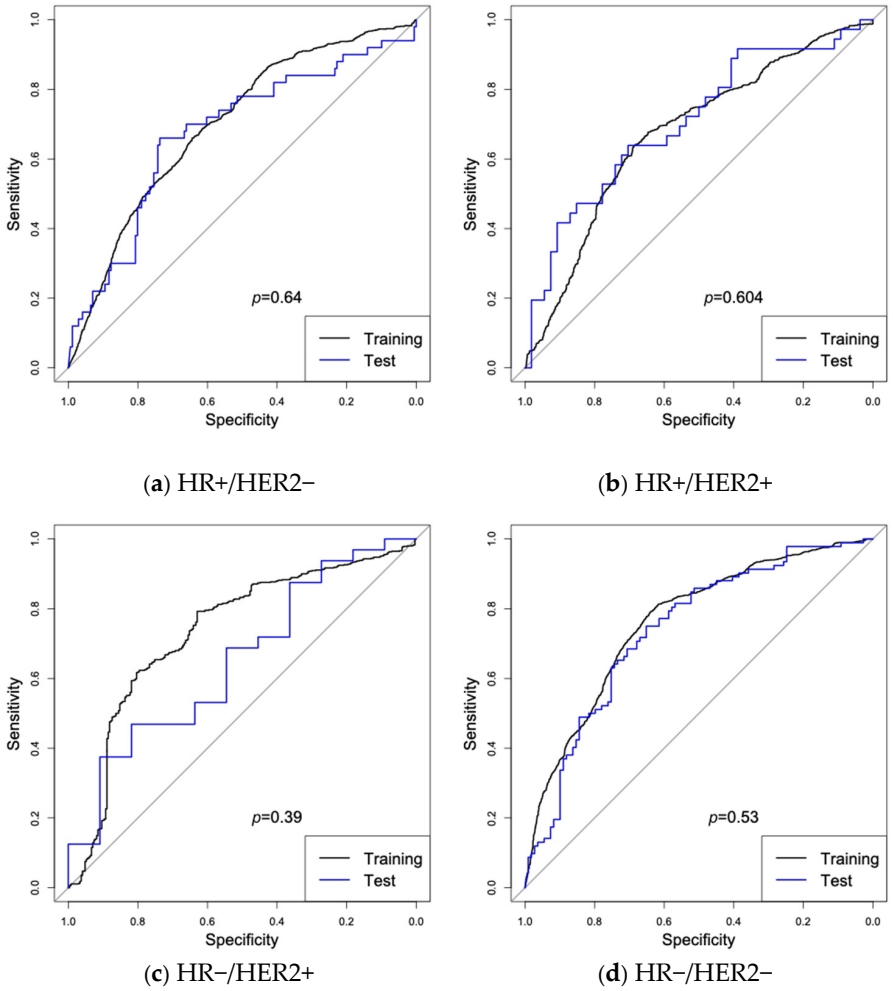


Figure 3. Receiver operating characteristic (ROC) curves of the functional tumor volume (FTV)-based predictive models estimated in the training and test sets stratified by breast cancer subtype. ROC curves are shown for the training (black) and test (blue) sets for (**a**) HR+/HER2−, (**b**) HR+/HER2+, (**c**) HR−/HER2+, and (**d**) HR−/HER2− (triple-negative) breast cancer. *p* values (estimated using DeLong's test) indicate the comparison of the areas under the ROC curves (AUCs) between the two curves. ROCs for the training sets were generated with predictions in the hold-out folds of the 10-fold cross-validation.

3.2.2. Secondary Model Development Using the Full Cohort

When MRI-based models were developed in the full cohort (the combination of the training and test sets, $n = 1771$) using the same methodology as the model development in the training set, the HR+/HER2− specific model included Δ*FTV*0_2 only (different from the model developed in the training set), the HR+/HER2+ included Δ*FTV*0_1 only, the HR−/HER2+ included Δ*FTV*0_1 only, and the triple negative included Δ*FTV*0_2 only (Supplemental Table S6).

3.3. Validation of the Youden Index-Based Cut Points

Youden Index–based cut points were determined in the training set and applied unchanged to the held-out test set to evaluate the generalizability of classification performance at fixed thresholds. These cut points were used as standardized statistical thresholds for model validation and were not intended to represent clinically optimized thresholds for treatment decision-making. The FTV variable included in the optimal model, the Youden Index-based cut point (in the FTV values and predicted probabilities), sensitivity, specificity, PPV, and NPV are listed in Table 3. Supplemental Figure S2 shows plots of estimated values

with 95% CIs. Figure 4 shows heatmaps of the confusion matrices when the Youden Index-based cut points were applied to the test set, compared to the confusion matrices generated by the same cut points in the training set. The same cut points resulted in similar sensitivity, specificity, PPV, and NPV in the test set compared to the training set in all subtypes except HR−/HER2+, in which specificities were substantially lower in the test set.

Table 3. Validation of Youden Index-based cut points.

Subtype	FTV Variable Included in the Optimal Model	Cut Point * ΔFTV (pCR Probability)	Cohort	Sensitivity (Fraction, [95% CI])	Specificity (Fraction, [95% CI])	PPV (Fraction, [95% CI])	NPV (Fraction, [95% CI])
HR+/HER2−	ΔFTV0_2	−89% (0.26)	Training	70% (82/117, [61, 78] %)	61% (252/410, [57, 66] %)	34% (82/240, [28, 41] %)	88% (252/287, [83, 91] %)
			Test	74% (37/50, [60, 85] %)	56% (95/171, [48, 63] %)	33% (37/113, [24, 42] %)	88% (95/108, [80, 93] %)
HR+/HER2+	ΔFTV0_1	−69% (0.47)	Training	68% (60/88, [57, 78] %)	67% (74/110, [58, 76] %)	63% (60/96, [52, 72] %)	73% (74/102, [63, 81] %)
			Test	67% (24/36, [49, 81] %)	59% (32/54, [45, 72] %)	52% (24/46, [37, 67] %)	73% (32/44, [57, 85] %)
HR−/HER2+	ΔFTV0_1	−72% (0.75)	Training	65% (42/65, [52, 76] %)	81% (22/27, [62, 94] %)	89% (42/47, [77, 96] %)	49% (22/45, [34, 64] %)
			Test	72% (23/32, [53, 86] %)	45% (5/11, [17, 77] %)	79% (23/29, [60, 92] %)	36% (5/14, [13, 65] %)
HR−/HER2−	ΔFTV0_2	−85% (0.52)	Training	79% (158/200, [73, 84] %)	65% (154/237, [59, 71] %)	66% (158/241, [59, 72] %)	79% (154/196, [72, 84] %)
			Test	71% (65/92, [60, 80] %)	67% (73/109, [57, 76] %)	64% (65/101, [54, 74] %)	73% (73/100, [63, 81] %)

* The same cut points determined in the training set were applied to the test set. Pathologic complete response (pCR) is considered a positive outcome, while non-pCR is considered a negative outcome. Patients with ΔFTV (functional tumor volume) lower than the cut point were considered test positive. Cut-point values are selected for the corresponding functional tumor volume (FTV) variables shown in the column on the left. ΔFTV0_1: early fractional change in FTV. ΔFTV0_2: inter-regimen fractional change in FTV. HR: hormone receptor. HER2: human epidermal growth factor receptor 2. PPV: positive predictive value. NPV: negative predictive value.

3.4. Representative Images

Representative MRI cases for three breast cancer subtypes (HR+/HER2−, HR−/HER2+, and triple-negative breast cancer) are shown in Figure 5. For each subtype, representative images from a patient who achieved pCR and a patient with residual disease (non-pCR) after NAC are shown. In general, patients achieving pCR demonstrated larger FTV reduction compared to patients with residual disease. These images show that patients with HR−/HER2+ breast cancer who have greater FTV reduction at an early time point are more likely to achieve pCR (Figure 5a–d). For HR+/HER2− (Figure 5e–h) and triple-negative (Figure 5i–l) patients, FTV reductions at inter-regimen are more predictive. Representative images were selected to illustrate patterns of FTV change associated with treatment response within each subtype, but these images do not represent the complete spectrum of imaging findings.

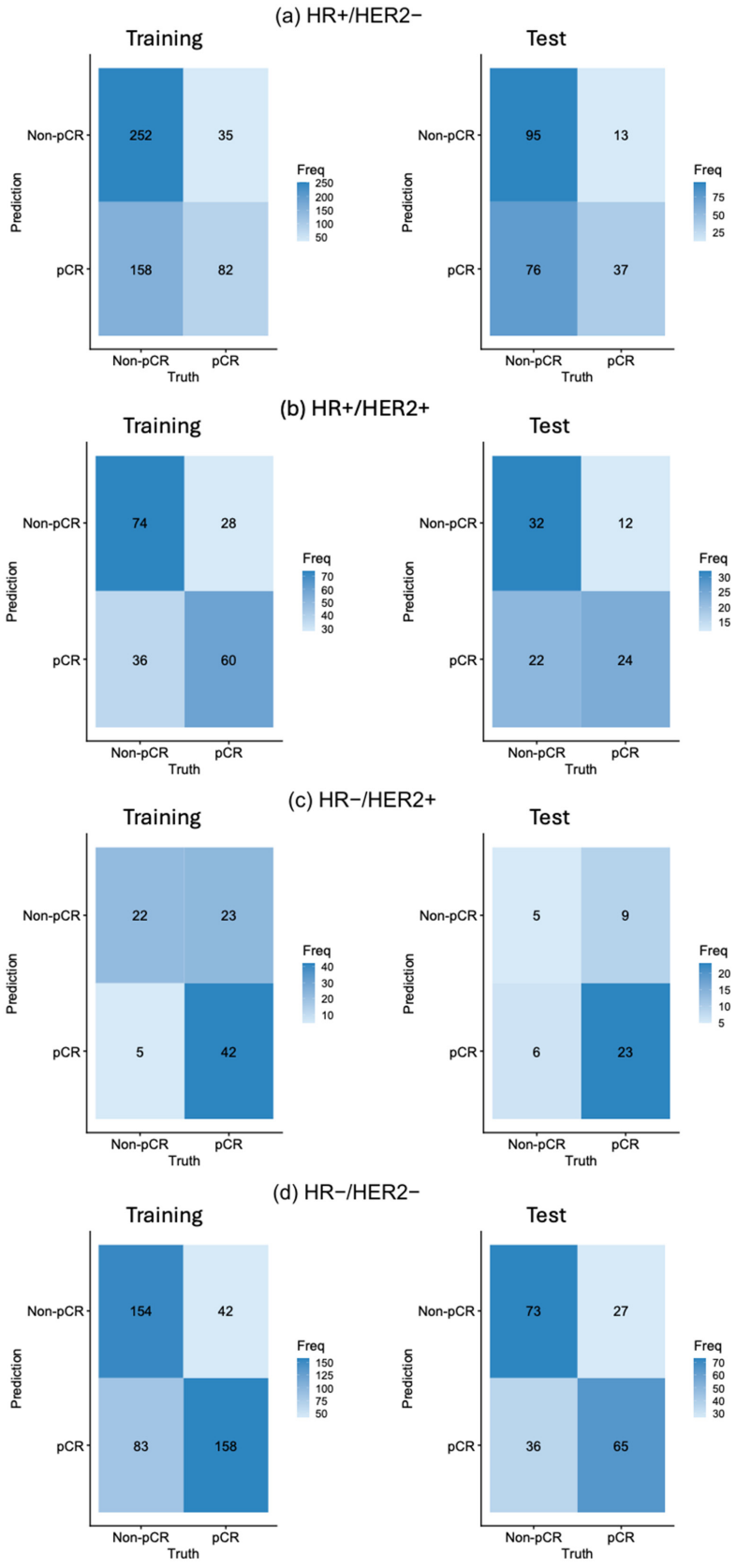
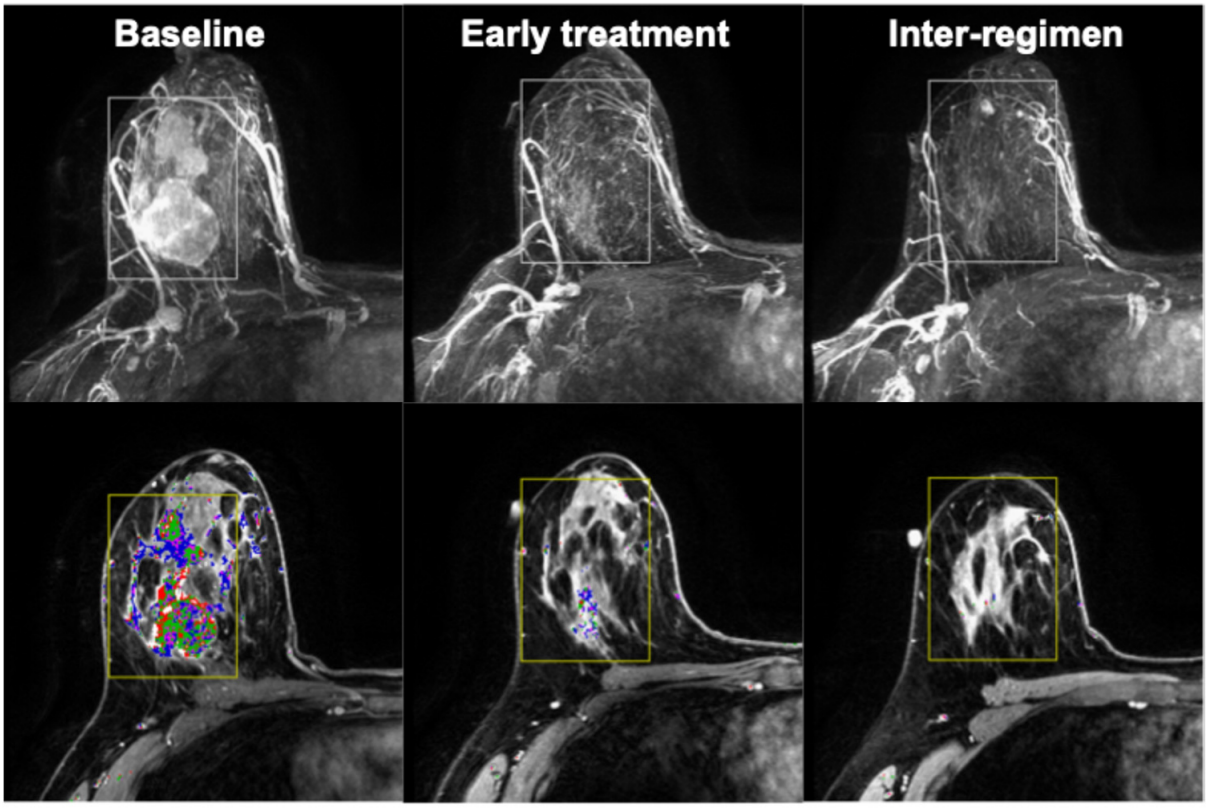
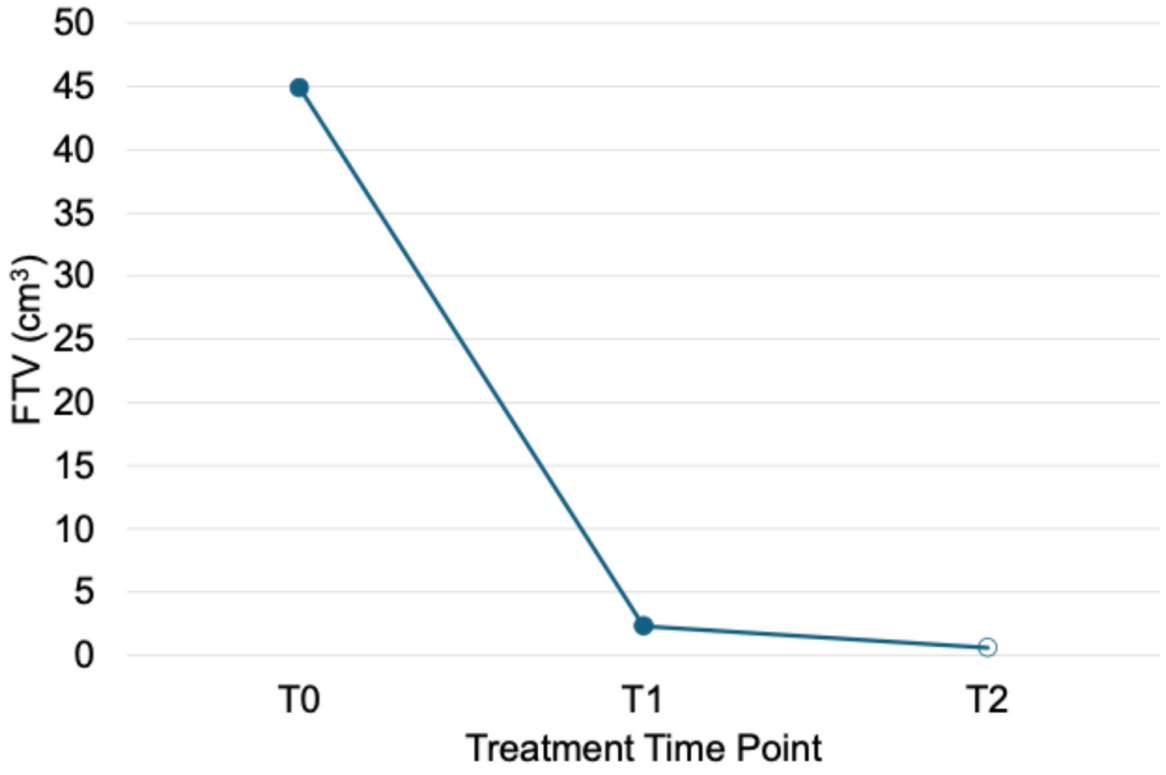


Figure 4. Heatmaps of the confusion matrices generated using the Youden Index-based cut points for each breast cancer subtype: (**a**) HR+/HER2−; (**b**) HR+/HER2+; (**c**) HR−/HER2+; and (**d**) HR−/HER2−. In each panel, the training set is shown on the left and the test set on the right. For each subtype, the cut point was determined in the training set and applied unchanged to the test set. Patient counts are shown within each cell. pCR, pathologic complete response.

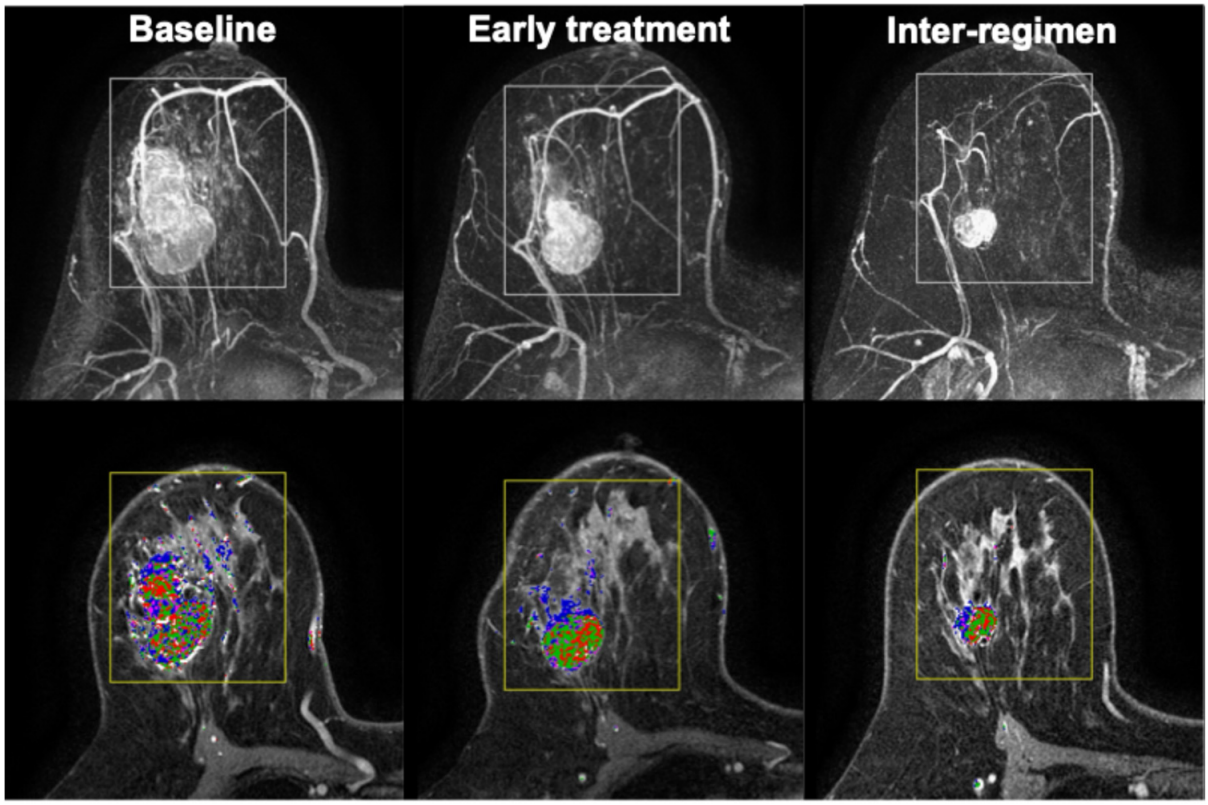


(**a**)

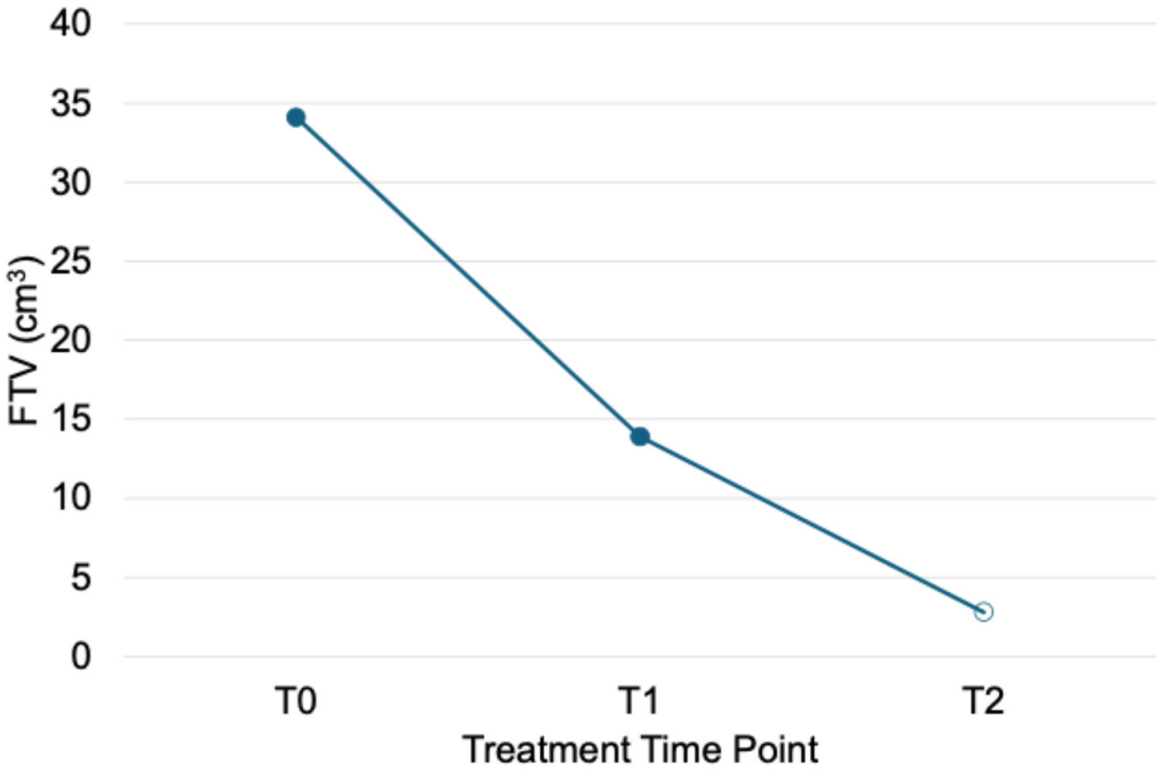


(**b**)

Figure 5. *Cont.*

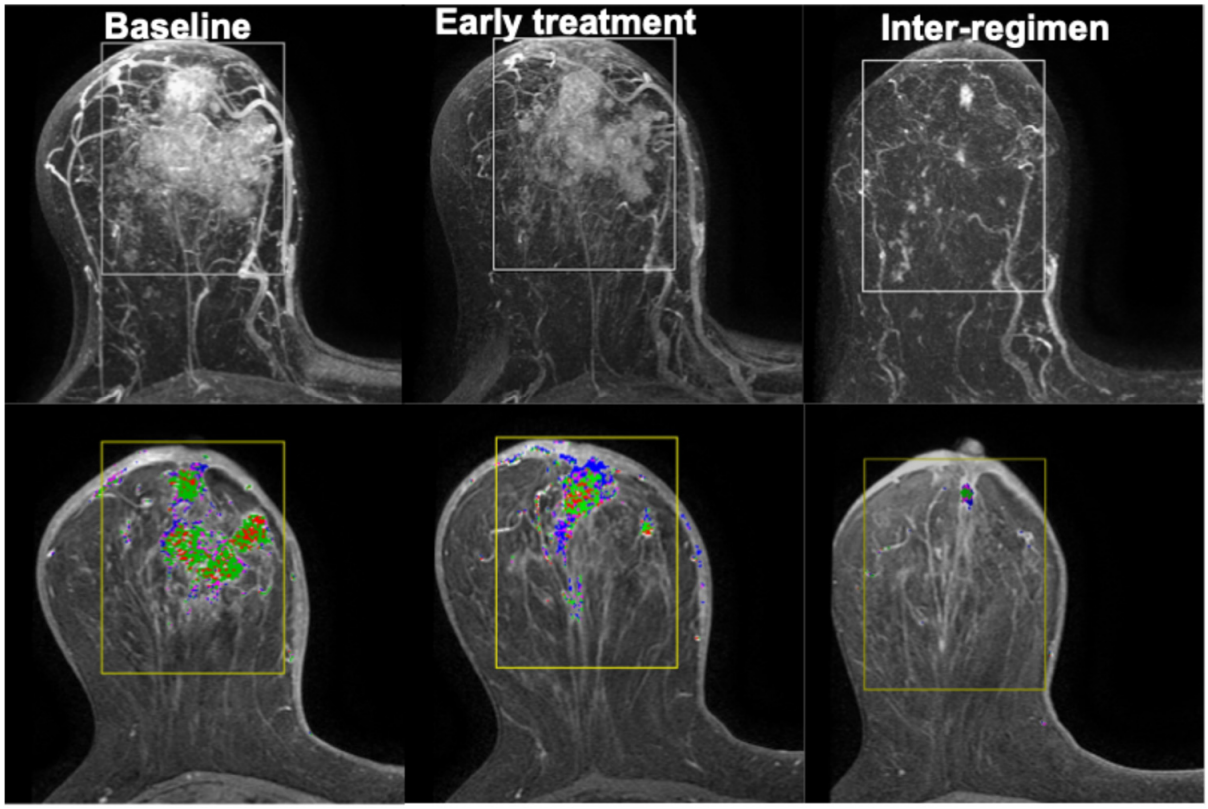


(**c**)

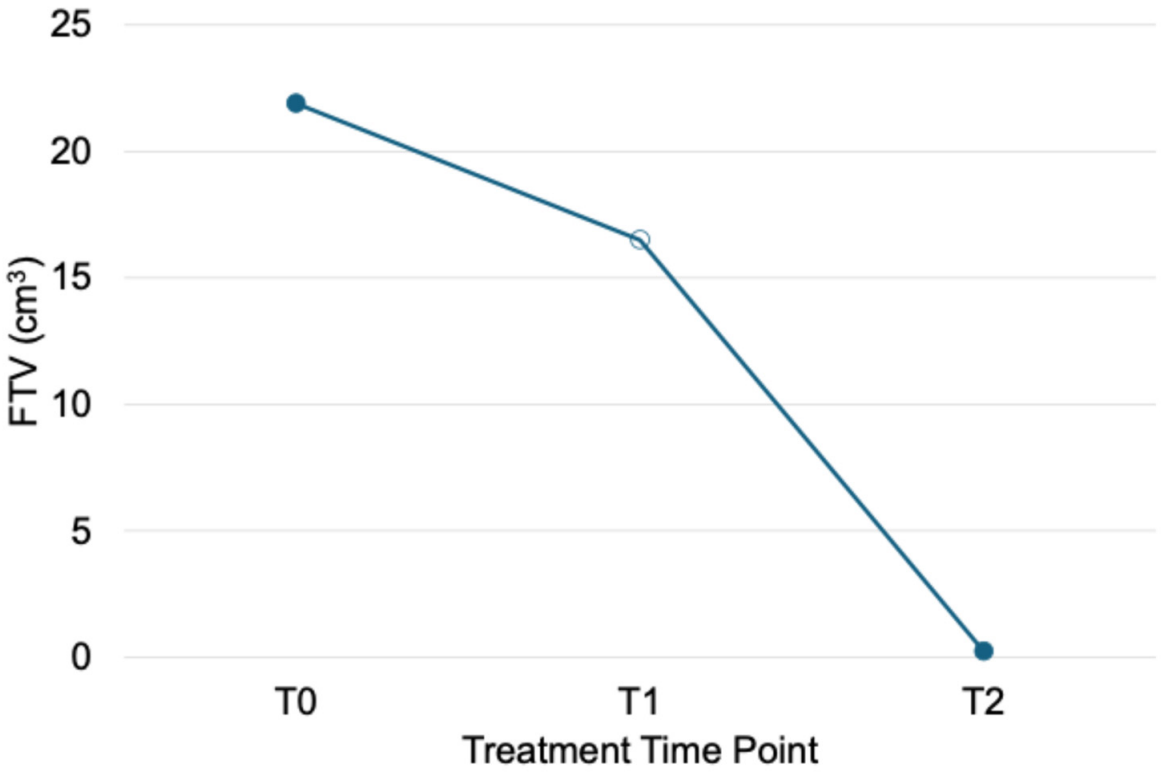


(**d**)

Figure 5. *Cont.*

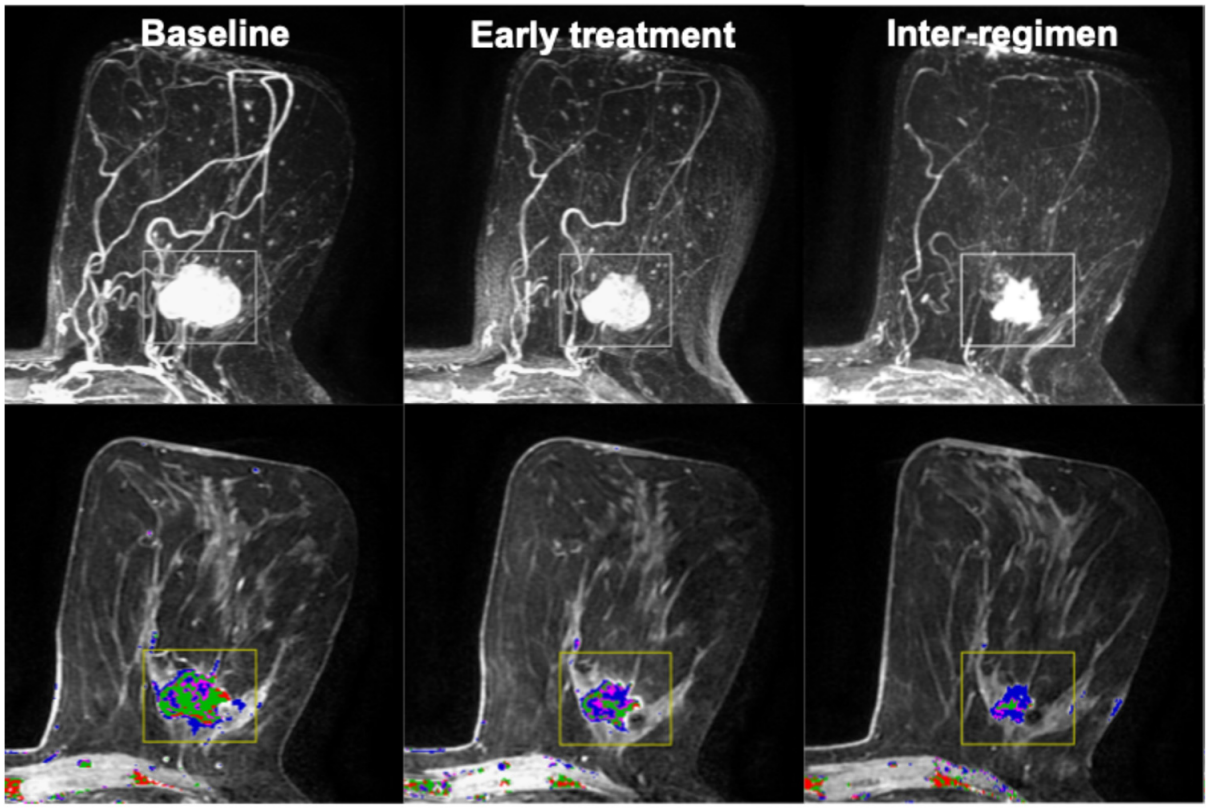


(**e**)

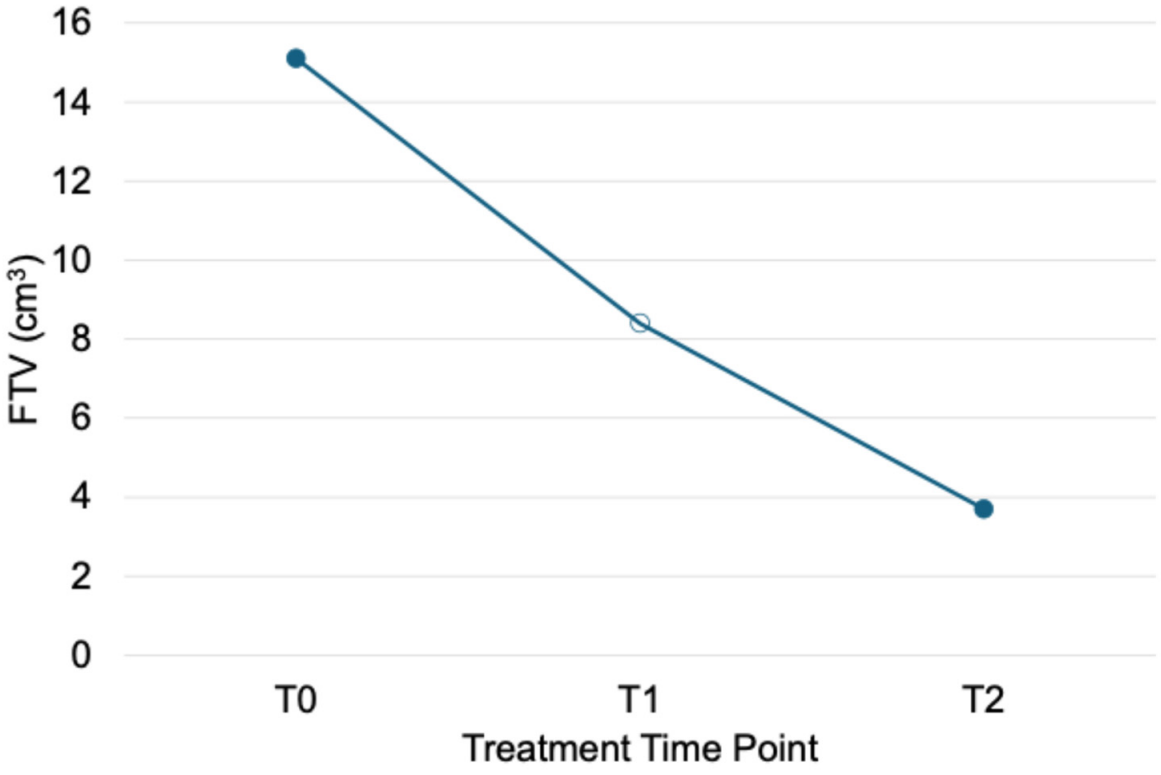


(**f**)

Figure 5. *Cont.*

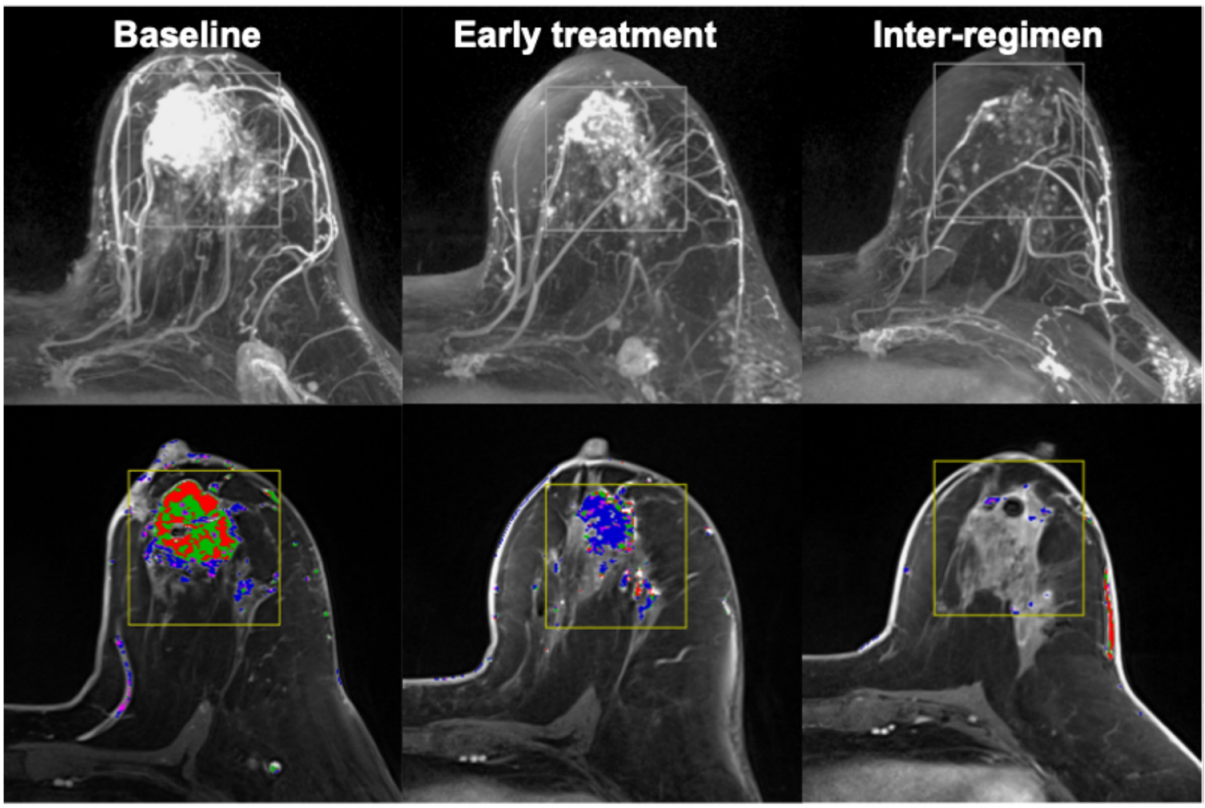


(**g**)

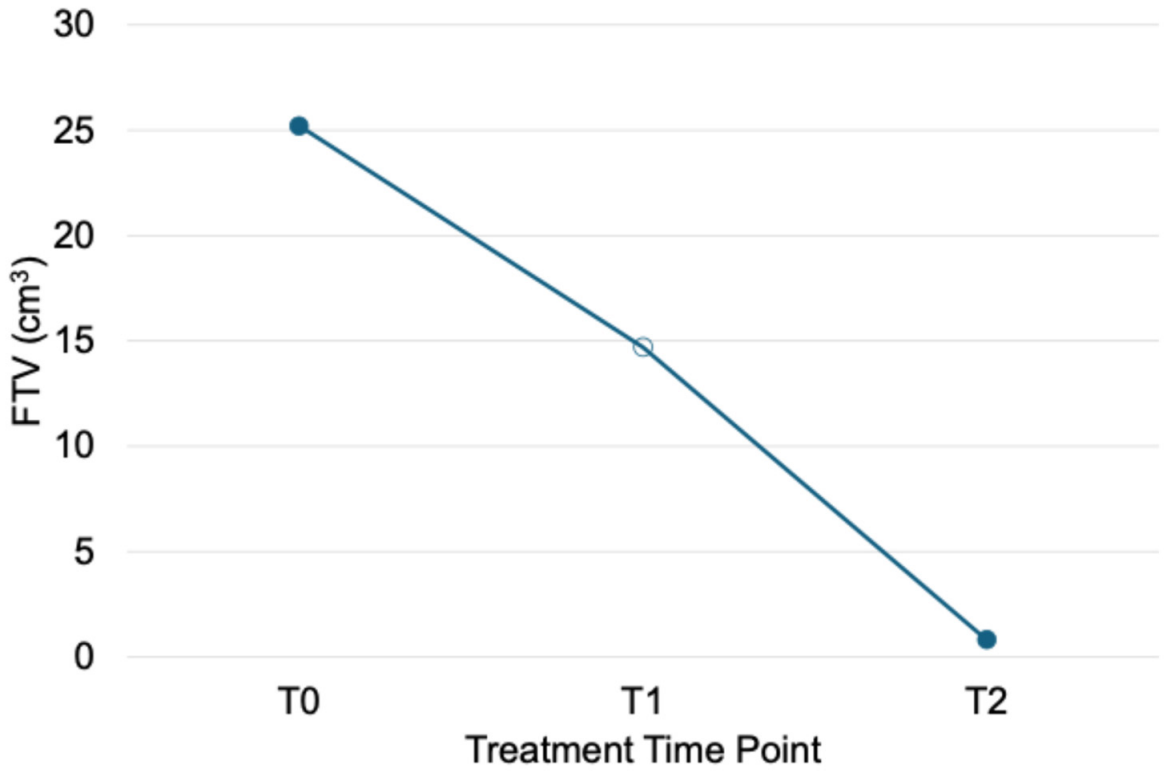


(**h**)

Figure 5. *Cont.*

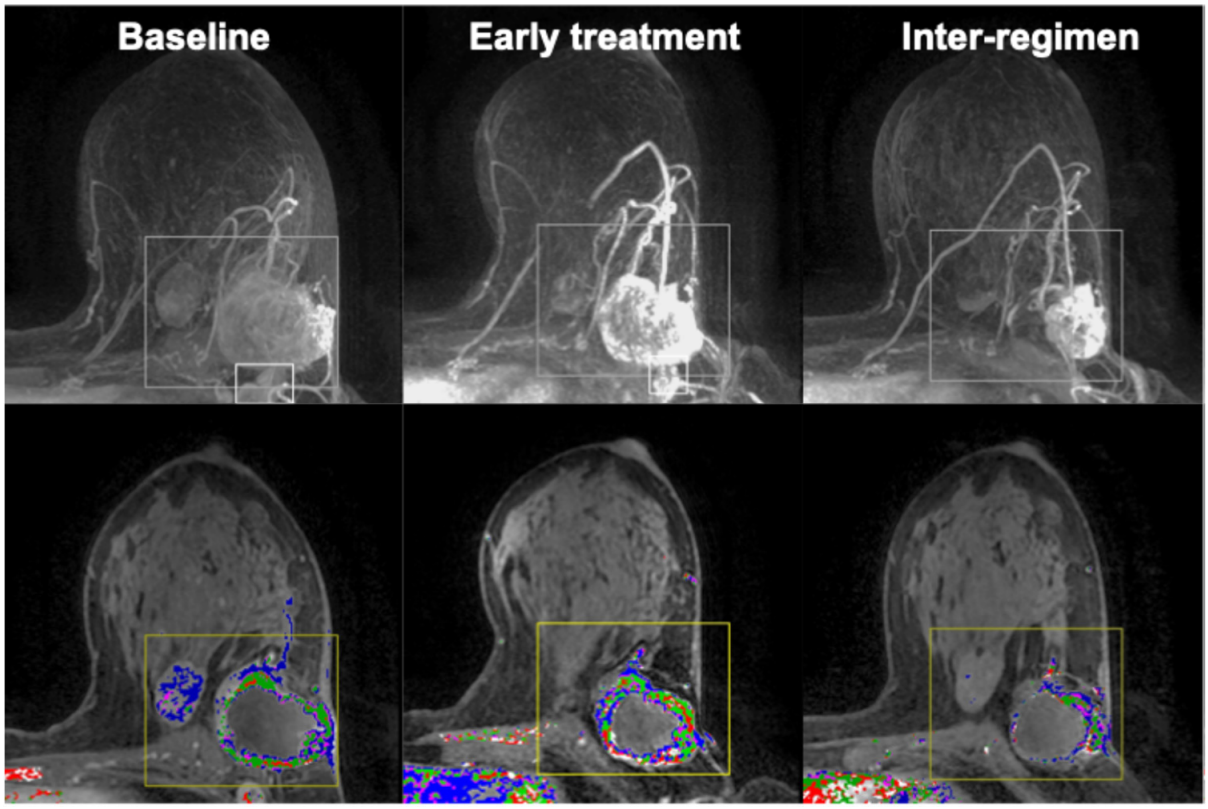


(**i**)

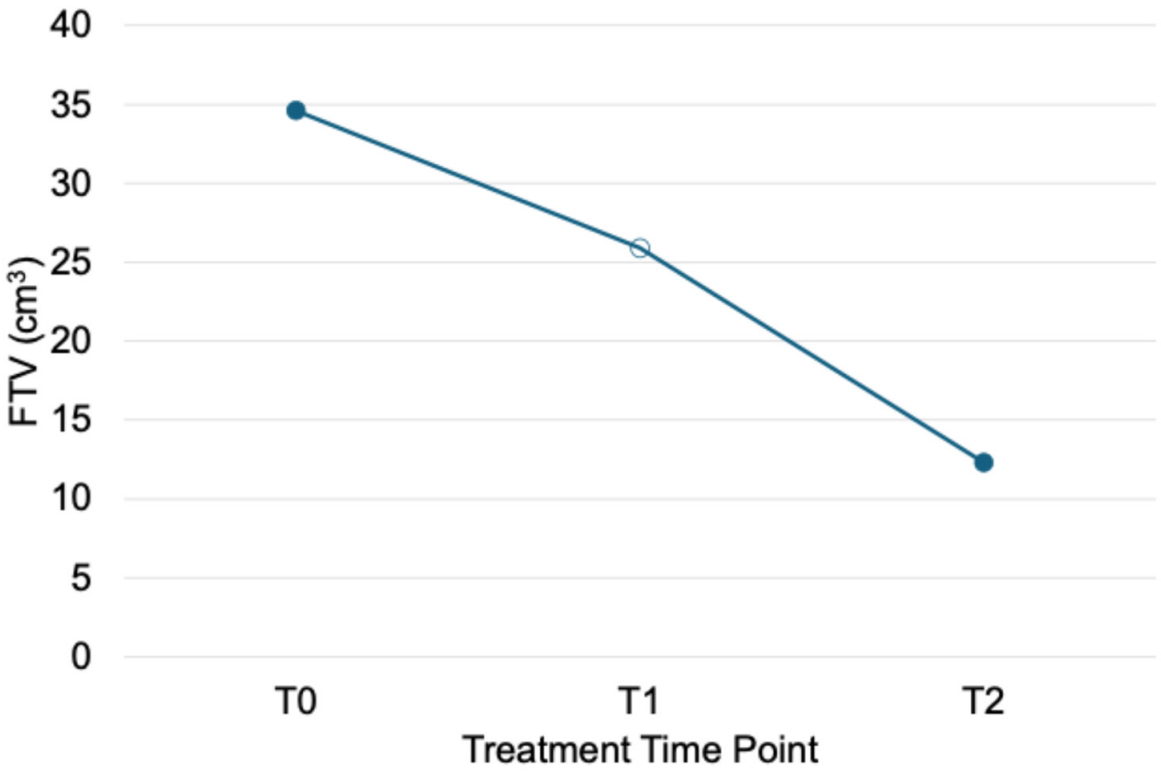


(**j**)

Figure 5. *Cont.*



(**k**)



(**l**)

Figure 5. Representative serial breast MRI examples and functional tumor volume (FTV) trajectories by breast cancer subtype and pathologic response. Representative cases are shown for HR−/HER2+ (**a**–**d**), HR+/HER2− (**e**–**h**), and HR−/HER2− (**i**–**l**) breast cancer, showing pathologic complete response (pCR) examples (**a**,**b**,**e**,**f**,**i**,**j**) first followed by non-pCR examples (**c**,**d**,**g**,**h**,**k**,**l**). Representative slides

of dynamic contrast-enhanced MRI acquired at baseline (T0), early treatment (T1), and inter-regimen (T2) are shown together with the corresponding FTV measurements and longitudinal FTV trajectories over the course of treatment. The rectangular box illustrates a 3-dimensional volume of interest (VOI) in which contrast enhancement thresholds are applied to select voxels, shown as colored pixels within the box, for FTV calculation. Each MRI example is shown in two rows. The top row shows the maximum intensity projection (MIP) of the subtraction between early enhancement (approximately 2 min and 30 s post-contrast injection) and pre-contrast, and the bottom row shows the signal enhancement ratio (SER) map overlaid on an axial MR image. Colors in the SER maps indicate different SER ranges: blue, 0–0.9; purple, 0.9–1.0; green, 1.0–1.3; red, 1.3–1.75; and white, ≥1.75. The yellow rectangle outlines the volume of interest.

4. Discussion

In this study, we validated the MRI-based subtype-specific predictive models built with FTV variables using a large patient population from a multicenter clinical trial. Our results demonstrated that the predictive performance in the test set was consistent with that in the training set for all breast cancer subtypes. However, the consistency varied slightly for HR+/HER2− and HR−/HER2+ subtypes, which had the lowest and highest pCR rates.

FTV is a standardized semiautomated MRI measurement that quantifies tumor volume changes during NACT in breast cancer, and the multicenter ACRIN 6657 trial demonstrated that FTV change can predict pCR early in the treatment [6]. Our results showed that early FTV changes best predicted pCR in HR−/HER2+ and HR+/HER2+ breast cancer subtypes, whereas late FTV changes best predicted pCR in HR+/HER2− and triple-negative breast cancer. These findings are consistent with our previously published results for HR+/HER2+ breast cancers from the I-SPY 2 study, which included 990 patients [10,22]. Other studies also reported that MRI effectively predicted pCR early in HER2+ tumors undergoing NAC [23,24]. In a retrospective single-center study by Yao et al., 86 HER2+ patients were included (42 with HR+ and 44 with HR−), and 59.3% achieved pCR at surgery after NAC [24]. The study found that the tumor volume change rate (the same calculation as FTV change in our study) was superior to long-axis diameter change rate at inter-regimen (equivalent to T2 in our study) in pCR prediction. Their study did not evaluate any MRI measurements earlier than T2 and tumor volume was calculated by manual segmentation of the tumor on MRI. The subtype-specific multicenter ADAPT (Adjuvant Dynamic Marker-Adjusted Personalized Therapy Trial Optimizing Risk Assessment and Therapy Response Prediction in Early Breast Cancer, NCT01779206) subtrials reported that MRI was more accurate in predicting pCR in the HR−/HER2+ subtype than in HR+/HER2+ and triple-negative subtypes after one cycle of NACT (Week 3) [23]. This is likely because HER2−targeted treatments are most effective in the HR−/HER2+ subtype, and rapid reduction in volume is highly predictive of obtaining a pCR and subsequent high rates of distant recurrence-free survival. It is possible that as we develop more targeted and effective treatments, we will see a shift to higher MRI FTV accuracy at the T1 (3-week) time point.

Importantly, our validation of the MRI-based subtype-specific predictive models suggests that clinical decision-making regarding therapy escalation or de-escalation should be considered at different treatment time points for different breast cancer subtypes. In addition, we performed a complementary analysis of Youden Index-based thresholds to evaluate the generalizability of classification performance at fixed, training-derived cut points rather than to establish thresholds for clinical treatment decisions. The Youden Index gives equal weight to sensitivity and specificity and does not account for the asymmetric clinical consequences of misclassification. This distinction is particularly important for treatment de-escalation, for which incorrectly classifying a patient with residual disease as likely to achieve pCR may have greater clinical consequences than failing to identify a patient who ultimately achieves pCR. Therefore, the thresholds reported here should not be used directly to guide

treatment escalation or de-escalation. Clinical implementation will require thresholds tailored to specific treatment decisions, with predefined acceptable risks of misclassification and prospective validation. The present results instead establish the performance and generalizability of longitudinal FTV as a response-prediction biomarker and provide a foundation for future studies evaluating clinically optimized decision thresholds.

Hormone receptor-positive (HR+/HER2−) breast cancer is the most common subtype and is generally less sensitive to cytotoxic chemotherapy than other subtypes [25]. In addition, HR+/HER2− breast cancers have a lower pCR rate than HER2+ or triple-negative breast cancers. Altogether, it is challenging to predict pCR accurately, especially for identifying excellent responses (positive prediction). Our model validation showed that the best FTV variable for predicting pCR in this subtype is the FTV change at inter-regimen, rather than the early treatment time point. The optimal cut point selected by the Youden Index achieved a relatively high NPV (88% in the test set) and a moderate specificity (56% in the test set), suggesting inter-regimen MRI may be best used for identifying patients who are not responding well to NAC and can be escalated to alternative treatment methods such as endocrine therapy [26,27].

TNBC is known to be an aggressive cancer subtype. Unlike the HR+/HER2− subtype, triple-negative tumors are generally sensitive to chemotherapy [28] and several studies have shown that DCE-MRI can accurately predict pCR early after just a few cycles of NAC [29–31]. However, our study found that FTV change detected in the inter-regimen MRI, rather than the early-treatment MRI, was the optimal FTV variable to predict pCR for this subtype. This finding suggests that while early-treatment FTV change may be a good predictor of pCR, inter-regimen FTV change may better reflect cumulative treatment sensitivity and residual viable tumor burden, thus providing the greatest predictive information for distinguishing pCR and non-pCR. This may also suggest the heterogeneous and dynamic nature of breast cancer response to NAC, especially for triple-negative breast cancer. When the optimal cut point was selected by the Youden Index, the FTV model demonstrated balanced sensitivity and specificity, with a higher NPV than PPV. These findings suggest that the model may be somewhat more effective at identifying patients unlikely to achieve pCR than at confirming those who will achieve pCR, although the clinically meaningful cut point depends on the intended application.

The predictive performance observed in our study was modest, with test-set AUCs of 0.68, 0.71, 0.65, and 0.74 for predicting pCR in HR+/HER2−, HR+/HER2+, HR−/HER2+, and TNBC, respectively. These results should be interpreted in the context of the considerable variability in performance reported for MRI-based pCR prediction models. In a multicenter study of 586 patients from four institutions, Liu et al. reported an AUC of 0.79 for a pretreatment multiparametric MRI radiomic signature and an AUC of 0.86 after incorporating clinicopathologic factors [32]. In another multicenter study of 1262 patients from four centers, pretreatment MRI models achieved mean external-validation AUCs of approximately 0.78 across HR+/HER2−, HER2+, and TNBC subtypes; performance increased substantially when post-treatment and longitudinal changes in MRI features were incorporated, with AUCs of 0.84–0.93 across external validation cohorts [33]. More recently, a multicenter deep-learning study of 1339 patients reported external-validation AUCs of 0.85–0.86 using longitudinal MRI, whereas a single-scan model showed substantially lower external-validation performance (AUCs, 0.61–0.64) [34]. Collectively, these findings suggest that MRI-based pCR prediction performance depends strongly on the imaging information available, modeling approach, and validation setting. Although our models' prediction performance is lower than that of some multiparametric or longitudinal MRI models, we obtained our results in a large cohort of 2117 patients across 35 participating sites, providing an assessment of model performance across substantially greater institutional and imaging heterogeneity.

An important potential application of longitudinal FTV is the early identification of patients with an inadequate or sufficient response to neoadjuvant therapy. Because FTV provides a quantitative measure of tumor response during treatment, it may complement clinical, pathologic, and molecular information in future response-adaptive treatment strategies. However, the models evaluated in this study are not intended to serve as stand-alone tools for treatment selection or modification, and the observed discriminatory performance does not support changing therapy solely based on an individual FTV-based prediction. Rather, the present study establishes the performance of longitudinal FTV across a large, heterogeneous multicenter population, an important prerequisite for evaluating its use as a robust imaging biomarker in prospective response-adapted treatment strategies. Further studies are needed to determine whether FTV-guided treatment modification, in conjunction with other clinical and biologic information, can improve patient outcomes.

We have developed response predictive subtypes in I SPY 2, molecular tools that predict which patients will be most responsive to chemotherapy [35]. These may help us to further optimize FTV prediction. As new therapies emerge that are highly effective, for example, in the immune-positive response subtypes, we may see that the timing for optimal predictability may shift to the earlier time point (3 weeks or T1). Furthermore, we are refining the early endpoint and have found that minimal residual disease (RCB1) plus pCR (RCB0) have the same excellent outcomes, and may be better than pCR alone, for all of the subtypes except TNBC [20].

This study has limitations. First, this study conducted an internal validation. Although the MRI model was evaluated on a separate test set, both the training and test cohorts were derived from the same clinical trial. External validation in independent populations is needed to further establish the generalizability of the model. Second, all data in this study were from I-SPY 2, a highly regulated and standardized trial, in which patients were enrolled according to trial eligibility criteria, treatment regimens were mixed between standard and experimental drugs (not representative of routine clinical practice), and imaging protocols were standardized (which may not be true in an actual clinical setting). Therefore, the model performance in this study may differ in routine clinical practice with more heterogeneous patient populations and imaging protocols. Third, we presented subtype-specific models developed and validated within each subgroup. This stratification analysis may have limited statistical power, resulting in increased uncertainty in performance estimates, as we observed in HR+/HER2− and HR−/HER2+ subtypes. Lastly, the present study evaluated the predictive performance of longitudinal FTV as an imaging biomarker and did not assess its incremental value beyond routinely available clinical and pathologic predictors. Future studies should evaluate FTV in combination with factors such as age, tumor grade, clinical tumor size, and other molecular or imaging biomarkers to determine whether longitudinal FTV provides additional predictive value beyond established clinical information and whether such integrated models improve clinically relevant prediction of treatment response.

In summary, we developed and validated breast cancer subtype-specific MRI models to predict pCR using longitudinal FTV measurements before and during NAC. There is substantial clinical impact to our findings, indicating that the optimal MRI time point for prediction varies by cancer subtype, with early-treatment FTV change showing the greatest predictive value for HER2+ breast cancer, whereas inter-regimen FTV change shows the greatest predictive value for HR+/HER2− and TNBC. We are prospectively testing these models in the ongoing I-SPY 2.2 trial. External validation is also warranted to determine whether personalized imaging strategies can be generalized to improve treatment outcomes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/diagnostics16233947/s1, Supplemental Figure S1: ROCs of single variable models compared to the optimal model for HR+/HER2− subtype in the training set; Supplemental Figure S2: Estimates and 95% confidence intervals of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) in the training and test sets using the same optimal cut points; Supplemental Table S1: List of experimental treatments or controls in the included cohort; Supplemental Table S2: Standardized I-SPY 2 DCE-MRI acquisition parameters; Supplemental Table S3: Site, vendor, field strength in accrual and included patient cohorts; Supplemental Table S4: FTV-based subtype-specific model developed in the training set ($n = 1227$); Supplemental Table S5: Comparison of the optimal model with single-variable models for HR+/HER2− in the training ($n = 527$, pCR rate = 22%); Supplemental Table S6: FTV-based subtype-specific model developed in the full cohort.

Author Contributions: Conceptualization, W.L., C.Y., D.M.W., L.J.E. and N.M.H.; methodology, W.L., J.K., D.M.W. and N.M.H.; software, W.L. and J.K.; validation, W.L., N.O., N.N.L., L.J.W., C.Y. and D.M.W.; formal analysis, W.L.; investigation, W.L., N.O., P.M., M.G. and S.V.; resources, J.P., B.L., L.J.E. and N.M.H.; data curation, T.J.B., N.N.L., L.J.W., E.P., B.N.J., M.G., on behalf of the I-SPY 2 Imaging Working Group and on behalf of the I-SPY 2 Investigators; writing—original draft preparation, W.L.; writing—review and editing, J.K., E.P., S.C.P., M.J.M., B.L. and L.J.E.; visualization, W.L., J.K. and L.J.E.; supervision, W.F.S., D.Y., H.S.R., L.J.v.V., A.M.D., L.J.E. and N.M.H.; project administration, T.J.B.; funding acquisition, L.J.v.V., M.J.M., L.J.E. and N.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NCI of the National Institutes of Health (NIH) under award numbers: R01CA132870, U01CA225427, P01CA210961, R01CA255442.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University of California San Francisco Institutional Review Board (protocol code 10-01565 and date of approval: 12 April 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Part of the imaging and clinical data used in this study are available on the Cancer Imaging Archive (TCIA): $n = 719$ are available at DOI: 10.7937/TCIA.D8Z0-9T85, and $n = 385$ are available at DOI: 10.7937/tcia.kk02-6d95. We are currently working on releasing the remaining data on TCIA in the near future. In the meantime, please contact the corresponding author, W.L., with the data queries. All data queries are subject to the regulations of the I-SPY Data Access & Publications Committee (DAPC).

Acknowledgments: The authors wish to acknowledge the generous support of the 501(c)3 study sponsors, Quantum Leap Healthcare Collaborative (QLHC, 2013 to present) and the Foundation for the NIH (2010 to 2012; to L.E.). The authors sincerely appreciate the ongoing support for the I-SPY program from the Safeway Foundation, the William K. Bowes, Jr. Foundation, Give Breast Cancer the Boot, QLHC and the Breast Cancer Research Foundation (to L.E.). Sincere thanks to all the patients who volunteered to participate in the I-SPY trials.

Conflicts of Interest: WFS reports shares of IONIS Pharmaceuticals; received consulting fees for pathology review from AstraZeneca; is a cofounder with equity in Delphi Diagnostics; issued patents for: (1) a method to calculate residual cancer burden, and (2) a genomic signature to measure sensitivity to endocrine therapy. BNJ reports royalties from UpToDate and consulting fees from Medicolegal. SCP reports research support from GE Healthcare and grant funding from Philips Healthcare and Guerbet, LLC. CY reports an institutional research grant from NCI/NIH; salary support and travel reimbursement from Quantum Leap Healthcare Collaborative; US patent title, "Breast cancer response prediction subtypes" (No. 18/174,491); and University of California Inventor Share. DY reports personal fees from Quantum Leap Healthcare Collaborative. HSR reports institutional research support to her prior institution (UCSF) from AstraZeneca, Gilead Sciences, Inc., Lilly, Merck & Co., Daiichi Sankyo, Inc., Novartis Pharmaceuticals, Pfizer, F. Hoffman-La Roche/Genentech, StemlineTherapeutics, and Ambryx; and institutional research support from Bicycle Therapeutics, F. Hoffman-La Roche/Genentech, Merck & Co., and Stemline Therapeutics; advisory and consulting

roles with Napo, Bristol Myers Squibb, Helsinn, and BioNTech. LvtV is an advisor and shareholder of Exai Bio; part-timeemployee and owns stock in Agendia. ADM reports institutional research funding from Novartis, Pfizer, Genentech and Neogenomics; Program Chair, Scientific Advisory Committee, ASCO. LJE reports personal fees from UpToDate and is an unpaid board member of Quantum Leap Healthcare. NMH reports institutional research funding from NIH. All other authors declare no competing interests.

Abbreviations

The following abbreviations are used in this manuscript:

MRI	Magnetic Resonance Imaging
I-SPY Trial	Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis
pCR	Pathologic Complete Response
AUC	Area Under the Curve (AUC)
ROC	Receiver Operating Characteristic
PPV	Positive Predictive Value
NPV	Negative Predictive Value
HR	Hormone Receptor
HER2	Human Epidermal growth factor Receptor 2
FTV	Functional Tumor Volume
NAC	Neoadjuvant Chemotherapy
DCE	Dynamic Contrast-Enhanced
ACRIN	American College of Radiology Imaging Network
AC	Anthracycline
IRB	Institutional Review Board
VOI	Volume of Interest
PE	Percentage Enhancement
SER	Signal Enhancement Ratio
SD	Standard Deviation
RCB	Residual Cancer Burden
CV	Cross-Validated
CI	Confidence Interval
N	Number

References

1. Cardoso, F.; Paluch-Shimon, S.; Senkus, E.; Curigliano, G.; Aapro, M.S.; André, F.; Barrios, C.H.; Bergh, J.; Bhattacharyya, G.S.; Biganzoli, L.; et al. 5th ESO-ESMO internatonal consensus guidelines for advanced breast cancer (ABC 5). *Ann. Oncol.* **2020**, *31*, 1623–1649. [CrossRef] [PubMed]
2. Cardoso, F.; Kyriakides, S.; Ohno, S.; Penault-Llorca, F.; Poortmans, P.; Rubio, I.T.; Zackrisson, S.; Senkus, E.; ESMO Guidelines Committee. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **2019**, *30*, 1194–1220. [CrossRef] [PubMed]
3. Gradishar, W.J.; Moran, M.S.; Abraham, J.; Aft, R.; Agnese, D.; Allison, K.H.; Anderson, B.; Burstein, H.J.; Chew, H.; Dang, C.; et al. Breast Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2022**, *20*, 691–722. [CrossRef] [PubMed]
4. Reig, B.; Heacock, L.; Lewin, A.; Cho, N.; Moy, L. Role of MRI to Assess Response to Neoadjuvant Therapy for Breast Cancer. *J. Magn. Reson. Imaging* **2020**, *52*, 1587–1606. [CrossRef] [PubMed]
5. Newitt, D.C.; Aliu, S.O.; Witcomb, N.; Sela, G.; Kornak, J.; Esserman, L.; Hylton, N.M. 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. [CrossRef] [PubMed]
6. Hylton, N.M.; Blume, J.D.; Bernreuter, W.K.; Pisano, E.D.; Rosen, M.A.; Morris, E.A.; Weatherall, P.T.; Lehman, C.D.; Newstead, G.M.; Polin, S.; 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. [CrossRef] [PubMed]
7. Hylton, N.M.; Gatsonis, C.A.; Rosen, M.A.; Lehman, C.D.; Newitt, D.C.; Partridge, S.C.; Bernreuter, W.K.; Pisano, E.D.; Morris, E.A.; Weatherall, P.T.; 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. [CrossRef] [PubMed]

8. Onishi, N.; Li, W.; Gibbs, J.; Wilmes, L.J.; Nguyen, A.; Jones, E.F.; Arasu, V.; Kornak, J.; Joe, B.N.; Esserman, L.J.; et al. Impact of MRI Protocol Adherence on Prediction of Pathological Complete Response in the I-SPY 2 Neoadjuvant Breast Cancer Trial. *Tomography* **2020**, *6*, 77–85. [CrossRef] [PubMed]
9. Onishi, N.; Bareng, T.J.; Gibbs, J.; Li, W.; Price, E.R.; Joe, B.N.; Kornak, J.; Esserman, L.J.; Newitt, D.C.; Hylton, N.M.; et al. Effect of Longitudinal Variation in Tumor Volume Estimation for MRI-guided Personalization of Breast Cancer Neoadjuvant Treatment. *Radiol. Imaging Cancer* **2023**, *5*, e220126. [CrossRef] [PubMed]
10. Li, W.; Onishi, N.; Gibbs, J.E.; Wilmes, L.J.; Le, N.N.; Metanat, P.; Price, E.R.; Joe, B.N.; Kornak, J.; Yau, C.; et al. MRI-Based Model for Personalizing Neoadjuvant Treatment in Breast Cancer. *Tomography* **2025**, *11*, 26. [CrossRef] [PubMed]
11. Janssen, L.M.; Dekker, B.M.D.; Gilhuijs, K.G.A.; van Diest, P.J.; van der Wall, E.; Elias, S.G. MRI to assess response after neoadjuvant chemotherapy in breast cancer subtypes: A systematic review and meta-analysis. *npj Breast Cancer* **2022**, *8*, 107. [CrossRef] [PubMed]
12. Pesapane, F.; Agazzi, G.M.; Rotili, A.; Ferrari, F.; Cardillo, A.; Penco, S.; Dominelli, V.; D'ECclesiis, O.; Vignati, S.; Raimondi, S.; et al. Prediction of the Pathological Response to Neoadjuvant Chemotherapy in Breast Cancer Patients with MRI-Radiomics: A Systematic Review and Meta-analysis. *Curr. Probl. Cancer* **2022**, *46*, 100883. [CrossRef] [PubMed]
13. Zhang, J.; Wu, Q.; Lei, P.; Zhu, X.; Li, B. MRI-based radiomics models for early predicting pathological response to neoadjuvant chemotherapy in triple-negative breast cancer: A systematic review and meta-analysis. *J. Appl. Clin. Med. Phys.* **2025**, *26*, e70296. [CrossRef] [PubMed]
14. Wang, L.; Yang, J.; Yang, L.; Zhu, Y.; Tang, X.; Cao, X.; Kang, W.; Sun, H.; Xie, Z. Developing a predictive model for neoadjuvant therapy in HER2 overexpression breast cancer using multi-parameter MRI radiomics: Two-center retrospective study. *Front. Oncol.* **2025**, *15*, 1544058. [CrossRef] [PubMed]
15. Hajri, R.; Aboudaram, C.; Lassau, N.; Assi, T.; Antoun, L.; Ribeiro, J.M.; Lacroix-Triki, M.; Ammari, S.; Balleyguier, C. Prediction of Breast Cancer Response to Neoadjuvant Therapy with Machine Learning: A Clinical, MRI-Qualitative, and Radiomics Approach. *Life* **2025**, *15*, 1165. [CrossRef] [PubMed]
16. Symmans, W.F.; Yau, C.; Chen, Y.Y.; Balassanian, R.; Klein, M.E.; Pusztai, L.; Nanda, R.; Parker, B.A.; Datnow, B.; Krings, G.; et al. Assessment of Residual Cancer Burden and Event-Free Survival in Neoadjuvant Treatment for High-risk Breast Cancer: An Analysis of Data From the I-SPY2 Randomized Clinical Trial. *JAMA Oncol.* **2021**, *7*, 1654–1663. [PubMed]
17. SPY2 Trial Consortium. Association of Event-Free and Distant Recurrence-Free Survival With Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer: Three-Year Follow-up Analysis for the I-SPY2 Adaptively Randomized Clinical Trial. *JAMA Oncol.* **2020**, *6*, 1355–1362. [PubMed]
18. Clark, A.S.; Yau, C.; Wolf, D.M.; Petricoin, E.F.; Veer, L.J.v.'.; Yee, D.; Moulder, S.L.; Wallace, A.M.; Chien, A.J.; Isaacs, C.; et al. Neoadjuvant T-DM1/pertuzumab and paclitaxel/trastuzumab/pertuzumab for HER2(+) breast cancer in the adaptively randomized I-SPY2 trial. *Nat. Commun.* **2021**, *12*, 6428. [CrossRef] [PubMed]
19. Pusztai, L.; Yau, C.; Wolf, D.M.; Han, H.S.; Du, L.; Wallace, A.M.; String-Reasor, E.; Boughey, J.C.; Chien, A.J.; Elias, A.D.; et al. Durvalumab with olaparib and paclitaxel for high-risk HER2-negative stage II/III breast cancer: Results from the adaptively randomized I-SPY2 trial. *Cancer Cell* **2021**, *39*, 989–998.e5. [CrossRef] [PubMed]
20. Yau, C.; Osdoit, M.; van der Noordaa, M.; Shad, S.; Wei, J.; de Croze, D.; Hamy, A.-S.; Laé, M.; Reyal, F.; Sonke, G.S.; et al. Residual cancer burden after neoadjuvant chemotherapy and long-term survival outcomes in breast cancer: A multicentre pooled analysis of 5161 patients. *Lancet Oncol.* **2022**, *23*, 149–160. [CrossRef] [PubMed]
21. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **1988**, *44*, 837–845. [CrossRef] [PubMed]
22. Li, W.; Le, N.N.; Nadkarni, R.; Onishi, N.; Wilmes, L.J.; Gibbs, J.E.; Price, E.R.; Joe, B.N.; Mukhtar, R.A.; Gennatas, E.D.; et al. Tumor Morphology for Prediction of Poor Responses Early in Neoadjuvant Chemotherapy for Breast Cancer: A Multicenter Retrospective Study. *Tomography* **2024**, *10*, 1832–1845. [CrossRef] [PubMed]
23. Graeser, M.; Schrading, S.; Gluz, O.; Strobel, K.; Würstlein, R.; Kümmel, S.; Schumacher, C.; Grischke, E.; Forstbauer, H.; Braun, M.; et al. Early response by MR imaging and ultrasound as predictor of pathologic complete response to 12-week neoadjuvant therapy for different early breast cancer subtypes: Combined analysis from the WSG ADAPT subtrials. *Int. J. Cancer* **2021**, *148*, 2614–2627. [CrossRef] [PubMed]
24. Yao, H.; Qian, X.; Zheng, R.; Sheng, X.; Ding, J.; Wang, M.; Zha, X.; Wang, S.; Wang, J. Early MRI-Derived Volumetric Thresholds Predict Response and Guide Personalization in HER2-Positive Breast Cancer: A Retrospective Study. *Biomedicines* **2025**, *13*, 2906. [CrossRef] [PubMed]
25. Torrisi, R.; Marrazzo, E.; Agostinetto, E.; De Sanctis, R.; Losurdo, A.; Masci, G.; Tinterri, C.; Santoro, A. Neoadjuvant chemotherapy in hormone receptor-positive/HER2-negative early breast cancer: When, why and what? *Crit. Rev. Oncol. Hematol.* **2021**, *160*, 103280. [CrossRef] [PubMed]

26. Soliman, S.; Zitter, R.; Walker, S.; Chmura, S.; Eniu, A.; Graff, S.L. Evolving Treatment Paradigms in Hormone Receptor–Positive, Human Epidermal Growth Factor 2–Negative Metastatic Breast Cancer. *Am. Soc. Clin. Oncol. Educ. Book* **2026**, *46*, e520518. [CrossRef] [PubMed]
27. Burciu, O.M.; Merce, A.-G.; Cerbu, S.; Iancu, A.; Popoiu, T.-A.; Cobec, I.M.; Sas, I.; Dimofte, G.M. Current Endocrine Therapy in Hormone-Receptor-Positive Breast Cancer: From Tumor Biology to the Rationale for Therapeutic Tunning. *Medicina* **2025**, *61*, 1280. [CrossRef] [PubMed]
28. Obidiro, O.; Battogtokh, G.; Akala, E.O. Triple Negative Breast Cancer Treatment Options and Limitations: Future Outlook. *Pharmaceutics* **2023**, *15*, 1796. [CrossRef] [PubMed]
29. Guirguis, M.S.; Adrada, B.E.; Yam, C.; Tripathy, D.; Candelaria, R.; Yang, W.; Patel, M.; Moseley, T.; Perez, F.; Whitman, G.J.; et al. Early tumor volume reduction by breast DCE MRI predicts pathologic complete response to neoadjuvant therapy in triple negative breast cancer. *npj Breast Cancer* **2025**, *11*, 129. [CrossRef] [PubMed]
30. Panthi, B.; Adrada, B.E.; Candelaria, R.P.; Guirguis, M.S.; Yam, C.; Boge, M.; Chen, H.; Hunt, K.K.; Huo, L.; Hwang, K.-P.; et al. Assessment of Response to Neoadjuvant Systemic Treatment in Triple-Negative Breast Cancer Using Functional Tumor Volumes from Longitudinal Dynamic Contrast-Enhanced MRI. *Cancers* **2023**, *15*, 1025. [CrossRef] [PubMed]
31. Liu, B.; Wu, L.; Liu, C.; Long, X.; Hu, S.; Zhang, L.; Liu, Z.; Liang, C. Baseline and Early Treatment MRI Model for Predicting Complete Pathologic Response to Neoadjuvant Chemoimmunotherapy in Patients With Triple-Negative Breast Cancer. *Am. J. Roentgenol.* **2025**, *225*, e2533178. [CrossRef] [PubMed]
32. Liu, Z.; Li, Z.; Qu, J.; Zhang, R.; Zhou, X.; Li, L.; Sun, K.; Tang, Z.; Jiang, H.; Li, H.; et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin. Cancer Res.* **2019**, *25*, 3538–3547. [CrossRef] [PubMed]
33. Huang, Y.; Zhu, T.; Zhang, X.; Li, W.; Zheng, X.; Cheng, M.; Ji, F.; Zhang, L.; Yang, C.; Wu, Z.; et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: A multicenter, retrospective study. *eClinicalMedicine* **2023**, *58*, 101899. [CrossRef] [PubMed]
34. Huang, X.; Xu, Z.; Zhao, Y.; Wang, Y.; Liu, Y.; Hu, W.; Zhao, K.; Yao, L.; He, J.; Yu, Y.; et al. Longitudinal MRI-based deep learning model for predicting pathological complete response in breast cancer: A multicenter, retrospective cohort study. *npj Precis. Oncol.* **2026**, *10*, 57. [CrossRef] [PubMed]
35. Wolf, D.M.; Yau, C.; Wulfkuhle, J.; Brown-Swigart, L.; Gallagher, R.I.; Lee, P.R.E.; Zhu, Z.; Magbanua, M.J.; Sayaman, R.; O'gRady, N.; et al. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell* **2022**, *40*, 609–623.e6. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.