HHS Public Access

Author manuscript

J Magn Reson Imaging. Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

J Magn Reson Imaging. 2017 July ; 46(1): 290–302. doi:10.1002/jmri.25560.

Magnetic Resonance Spectroscopy of Breast Cancer for Assessing Early Treatment Response: Results from the ACRIN 6657 MRS Trial

Patrick J. Bolan, PhD,
Department of Radiology, University of Minnesota, Minneapolis, MN

Eunhee Kim, PhD,
National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD. American College of Radiology Imaging Network (ACRIN), Philadelphia, PA

Benjamin A. Herman, SM,
American College of Radiology Imaging Network (ACRIN), Philadelphia, PA. Center for Statistical Sciences, Brown University, Providence, RI

Gillian M. Newstead, MD,
Department of Radiology, University of Chicago, Chicago, IL

Mark A. Rosen, MD, PhD,
Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA

Mitchell D. Schnall, MD, PhD,
American College of Radiology Imaging Network (ACRIN), Philadelphia, PA. Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA

Etta D. Pisano, MD,
Department of Radiology, Medical College of South Carolina, Charleston, SC

Paul T. Weatherall, MD,
Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX

Elizabeth A. Morris, MD,
Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY

Constance D. Lehman, MD, PhD,
Department of Radiology, Massachusetts General Hospital, Boston, MA

Michael Garwood, PhD,
Department of Radiology, University of Minnesota, Minneapolis, MN

Michael T. Nelson, MD,
Department of Radiology, University of Minnesota, Minneapolis, MN

Douglas Yee, MD,
Masonic Cancer Center and Department of Medicine, University of Minnesota, Minneapolis, MN

Corresponding Author: Patrick J. Bolan, Center for Magnetic Resonance Research, 2021 6th Street SE, Minneapolis MN 55455, 612/625-6526, bola0035@umn.edu.

Sandra M. Polin, MD,
Washington Radiology Associates, P.C., Fairfax, VA

Laura J. Esserman, MD, MBA,
Department of Surgery, University of California, San Francisco, CA

Constantine A. Gatsonis, PhD,
American College of Radiology Imaging Network (ACRIN), Philadelphia, PA. Center for Statistical Sciences, Brown University, Providence, RI

Gregory J. Metzger, PhD,
Department of Radiology, University of Minnesota, Minneapolis, MN

David C. Newitt, PhD,
Department of Radiology, University of California, San Francisco, CA

Savannah C. Partridge, PhD, and
Department of Radiology, University of Washington, Seattle, WA

Nola M. Hylton, PhD
Department of Radiology, University of California, San Francisco, CA

for the ACRIN 6657 Trial team and ISPY-1 Investigators

Abstract

Purpose—To estimate the accuracy of predicting response to neoadjuvant chemotherapy (NACT) in patients with locally advanced breast cancer using magnetic resonance spectroscopy (MRS) measurements made very early in treatment.

Materials and Methods—This prospective HIPAA-compliant protocol was approved by the American College of Radiology and local-site institutional review boards. 119 women with invasive breast cancer of ≥3 cm undergoing NACT were enrolled between September 2007 and April 2010. MRS measurements of the concentration of choline-containing compounds ([tCho]) were performed prior to the first chemotherapy regimen (time point 1, TP1) and 20–96 hours after the first cycle of treatment (TP2). The change in [tCho] was assessed for its ability to predict pathologic complete response (pCR) and radiologic response using the area under the receiver operating characteristic curve (AUC) and logistic regression models.

Results—Of the 119 subjects enrolled, only 29 cases (24%) with 8 pCRs provided usable data for the primary analysis. Technical challenges in acquiring quantitative MRS data in a multi-site trial setting limited the capture of usable data. In this limited data set, the decrease in tCho from TP1 to TP2 had poor ability to predict either pCR (AUC = 0.53, 95% CI: [0.27, 0.79]) or radiologic response (AUC = 0.51, 95% CI: [0.27, 0.75]).

Conclusions—The technical difficulty of acquiring quantitative MRS data in a multi-site clinical trial setting led to a low yield of analyzable data, which was insufficient to accurately measure the ability of early MRS measurements to predict response to NACT.

Keywords

magnetic resonance spectroscopy; breast cancer; choline; treatment response

Introduction

Systemic chemotherapy in combination with surgery is currently the standard treatment for women with locally advanced breast cancer (stages IIb and III). Neoadjuvant chemotherapy (NACT) (systemic therapy prior to definitive surgery) has the advantage of increasing the rate of breast conservation surgery (1, 2). NACT also provides an opportunity to monitor an individual patient's response and tailor her therapeutic regimen. Such adaptive therapy requires a minimally-invasive means to distinguish responders from non-responders early in the course of treatment. Several studies have shown that contrast-enhanced MRI can distinguish responders from non-responders after the first cycle of NACT (3–5).

Previous work has suggested that changes in the total choline (tCho) signal, as measured by magnetic resonance spectroscopy (MRS), may provide an earlier indicator of treatment response than size changes. Breast cancers commonly exhibit an elevated level of tCho due to changes in phospholipid metabolism (6), which have been attributed to frequent mutations in choline kinase and other enzymes (7). An early clinical study by Kvistad et al. (8) observed that the in vivo tCho signal disappeared in response to successful chemotherapy. A number of subsequent studies found that a decrease in tCho signal during chemotherapy was associated with favorable pathologic (9–12) and radiologic responses (13–15).

While the above studies looked at tCho changes measured tCho one or more weeks after starting chemotherapy, in vitro studies have shown that taxane treatment leads to cell death on the 24–48 hour timescale (16), well before changes in tumor size would be measureable. Meisamy et al. (17) sought to measure these early metabolic changes in vivo by quantifying the tCho concentration ([tCho]) one day after starting chemotherapy. In a small study of 13 patients, they found that [tCho] changes could distinguish radiologic responders from non-responders in all cases, with decreasing [tCho] indicating response and increasing [tCho] indicating stable or progressive disease.

Motivated by the Meisamy et al. results (17), this study was designed to assess MRS measurements of [tCho] as an early indicator of NACT response in a multi-site, multi-vendor trial setting. The primary hypothesis was that a decrease in [tCho] measured acutely (1–4 days) following the first chemotherapy treatment would predict pathologic complete response (pCR). This study was performed by extending the American College of Radiology Imaging Network (ACRIN) 6657 trial (*Contrast-Enhanced Breast MRI for Evaluation of Patients Undergoing Neoadjuvant Treatment for Locally Advanced Breast Cancer*), which assessed contrast-enhanced MRI metrics for response prediction, with added MRS measurements and modified time points. This prospective trial extension, herein called ACRIN 6657-ext, was a companion study to Cancer and Leukemia Group B (CALGB) 150007 and the Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular Analysis (I-SPY TRIAL), which studied imaging- and tissue-based biomarkers for predicting response and survival. The purpose of this study was to estimate the accuracy of predicting response to neoadjuvant chemotherapy (NACT) in patients with locally advanced breast cancer using magnetic resonance spectroscopy (MRS) measurements made very early in treatment.

Materials and Methods

Participant Eligibility and Enrollment

Female patients enrolling in CALGB 150007 who had tumors that measured at least 3 cm in diameter by clinical examination or imaging, and who were receiving NACT with a weekly paclitaxel-based regimen for 12 weeks, in combination with optional trastuzumab for HER+ patients, followed by four cycles of an anthracycline-cyclophosphamide regimen, were eligible for this prospective study. Pregnant patients and those with ferromagnetic prostheses, intracranial aneurysm clips and other absolute contraindications to MRI were excluded from the study. Following amendments to both the CALGB 150007 and ACRIN 6657 protocols, both studies were re-opened to enrollment from September 2007 to April 2010. Patients signed a single consent form for both studies. The Health Insurance Portability and Accountability Act (HIPAA)–compliant protocol and informed consent process were approved by the American College of Radiology Institutional Review Board and local-site institutional review boards. Patients were screened for eligibility, consented, enrolled through CALGB 150007, and then registered to ACRIN 6657-ext.

MRS Quality Control

As MR spectroscopy of the breast is not routinely used in clinical practice, a training and quality control program was used to help control MRS data quality (18). Prior to enrolling patients, each site was required to qualify by submitting MRS measurements with acceptable accuracy and spectral quality from a trial-specific spectroscopy phantom (detailed in appendix). Two sequence variations were allowed to account for variations between vendors: a measurement with fixed echo time of 125 ms, 64 or more averages, and optional fat suppression; or a TE-averaged (19) acquisition using an array of echo times from 50–200 ms in 64 or more increments. The trial team provided consultation and training materials to help standardize MRS data acquisition, file transfer, and in vivo voxel placement. Throughout study accrual the trial team monitored data quality from both in vivo and scheduled phantom scans, and provided sites with feedback and consulting as needed.

Imaging and Spectroscopy Data Acquisition Procedures

MR imaging and spectroscopic examinations were performed within 4 weeks prior to the start of the first chemotherapy regimen (time point 1, TP1), 1–4 days after the first cycle of treatment (TP2), between the first and second chemotherapy regimens (TP3), and after the second regimen and prior to surgery (TP4). The 1–4 day timing of the TP2 scan was further divided into 1-day (20–28 hour) and 2–4 day (48–96 hour) windows, based on scheduling preference. Examinations TP2–4 were performed on the same MR system or similar system (same vendor, breast coil, and field strength). The MRI data from TP3 and the MRS data from TP3 and TP4 were not considered for the analyses in this manuscript. The MRI and MRS acquisition protocols were the same for all time points.

MR examinations were performed with either 1.5T or 3T MR scanners (determined by availability) using a dedicated breast radiofrequency coil and prone positioning. The acquisition protocol included a localizer, a T2-weighted sequence, a contrast-enhanced image series, and finally MRS. The contrast-enhanced series consisted of a high-spatial-

resolution (in-plane spatial resolution ≤ 1 mm) three-dimensional fat-suppressed T1-weighted imaging volume acquired either bilaterally (with axial orientation) or unilaterally (in sagittal orientation) using a gradient-echo sequence with a repetition time of 24 ms or less, minimum echo time with water and fat in phase, a flip angle of 45° or less, a field of view 16–18 cm (sagittal) or 32–40 cm (axial), a minimum matrix of 256 × 192 (sagittal) or 512 × 384 (axial), and 64 or more sections with a thickness of 2.5 mm or less. Imaging time for the T1-weighted volume was required to be a maximum of 3 minutes. The first volume in the series was acquired before injection of a gadolinium-based contrast agent at a dose of 0.1 mmol/kg over 15 seconds, followed by a 10 mL saline flush over 15 seconds. The second volume acquisition was started simultaneously with the contrast agent injection, and additional volumes were acquired sequentially for a minimum of 8 minutes post-injection.

MR spectroscopy was performed after the completion of the contrast-enhanced series. For each subject's pre-treatment scan (TP1), a 3D cubic voxel was prescribed to cover the enhancing lesion as much as possible while excluding adipose tissue, normal-appearing fibroglandular tissue, necrotic regions, and artifacts due to biopsy tissue markers ("clips"), using the post-contrast images for guidance. The voxel size was variable with a minimum of 10 mm per side. For subsequent scans (TP2–4), operators were instructed to adjust the voxel size and position as needed to best imitate the voxel placement of the TP1 scan (e.g., if the lesion shrank the voxel should as well). Operators also recorded their confidence (on a 1–5 scale) that the voxel was well-placed based on the above criteria. After static field shim adjustments a water reference scan and a "metabolite" scan were acquired using a PRESS (20) localization sequence. The water reference scan was acquired with 5 echo times (TE=50, 75, 100, 125, 150 ms), repetition time 6 s, 1 or 2 averages, and resonance frequency set on water at 4.7 ppm. The metabolite scan was acquired with water suppression enabled, repetition time 3 s, and the resonance frequency set at 3.2 ppm.

MR Imaging Analysis

Image data were de-identified and centrally archived at the American College of Radiology Imaging Core Laboratory. Staff members of the breast MR imaging laboratory at the University of California at San Francisco performed the quantitative analysis of all MR image data. The primary quantitative measurement was an estimate of functional tumor volume (FTV) (21), computed as the sum of voxels with a percentage enhancement (PE) greater than 70%, where PE was defined as [(S1 - S0)/S0] *100%, where S0 and S1 represent the signal intensities on the precontrast, and early postcontrast images, respectively. Site-specific adjustments to the 70% threshold were necessary to account for variability in MR imaging systems and imaging parameters. Radiologic response to treatment was defined as the change in functional tumor volume over the complete therapy, ($\%\Delta FTV_{1\rightarrow4} = ((FTV_4 - FTV_1)/FTV_1)*100$).

MR Spectroscopy Analysis

De-identified MRS data were transferred from the American College of Radiology Imaging Core Laboratory to University of Minnesota for processing. MRS analysis was based on the methods previously described (22). Preprocessing for all spectra included frequency and phase adjustments, noise normalization, Lorentzian line broadening, and zero-filling to

produce comparable spectral discretization for both field strengths. The water reference spectra were processed by fitting water and 1.3 ppm lipid resonances with a Voigt lineshape. The peak amplitudes for each echo time were fit with a monoexponential model to estimate the T_2 relaxation rate and T_2-corrected amplitude for both water and lipid, as well as the lipid fraction (defined as lipid/(water+lipid)). The individual traces from the metabolite scans were individually phase and frequency-corrected (23) prior to averaging. The residual water and lipids were fit with Voigt lineshapes and removed to reduce baseline distortions prior to fitting a total choline (tCho) resonance at 3.2 ppm. A tCho peak was considered detectable if four criteria were met: peak frequency-domain signal-to-noise ≥ 2, ratio of tCho to water linewidth ≥ 0.33 (to avoid fitting noise), Cramer-Rao minimum variance bound ≤ 200%, and tCho center frequency between 3.1 and 3.3 ppm. After fitting the datasets were manually reviewed and those with poor quality (e.g., poor shim, artifacts, baseline, etc.) were removed from the analysis set. The tCho and water amplitudes were then converted to a concentration in units of mmol tCho per kilogram water as described previously (22), with the exception that water T_2 was measured and corrected in each dataset rather than using an assumed water T_2. The T_2 for tCho was taken from the literature using 263 ms at 1.5 T (24–26) and 204 ms at 3T (27). If a tCho peak was not detected at TP1, the concentration was considered unmeasurable. A representative case showing results of imaging and spectroscopic analysis is shown in Figure 1.

Histopathologic Analysis

Histopathologic analysis of surgical specimens was performed locally by each institution's pathologist. Pathologic complete response (pCR) was reported when no residual invasive disease was present in the breast and axillary lymph nodes (ypT0/is, ypN0). Hormone receptor (HR) positivity (ER-positive and/or PR-positive) and HER2 receptor expression were determined from pretreatment core biopsy by immunohistochemistry and Allred score. For subtype analysis, patient groups were defined as those with HR+/HER2-, HER2+ and HR-/HER2- (triple negative, TN) tumors.

Statistical Analysis

The primary analysis was to determine if a change in tCho concentration from pre-treatment to the early treatment time point, defined as $\Delta[tCho]_{1\rightarrow2} = [tCho]_2 - [tCho]_1$, could discriminate patients with pCR from non-responders. This was assessed using three analyses. First, the nonparametric area under the ROC curve (AUC) of $\Delta[tCho]_{1\rightarrow2}$ was calculated for its ability to predict pCR (28, 29). Second, the association between dichotomized $\Delta[tCho]_{1\rightarrow2}$ and pCR status was determined by Fisher's exact test, where $\Delta[tCho]_{1\rightarrow2}$ was dichotomized at the optimal cutpoint by maximizing the Youden's index (sensitivity+specificity−1). Third, univariate logistic regression models were used to assess the association between continuous $\Delta[tCho]_{1\rightarrow2}$ and pCR, and the odds ratios (OR) and 95% confidence intervals (CI) for $\Delta[tCho]_{1\rightarrow2}$ were calculated. Secondary analyses were performed using the total tCho change in mmol/kg units ($\Delta[tCho]_{1\rightarrow2}$) and the percent change relative to baseline ($\%\Delta[tCho]_{1\rightarrow2} = (\Delta[tCho]_{1\rightarrow2} / [tCho]_1)*100$) as explanatory variables, and with radiologic response ($\%\Delta FTV_{1\rightarrow4}$), dichotomized at the median value, as the outcome. Additionally, Spearman's rho was used to determine if $\Delta[tCho]_{1\rightarrow2}$ was correlated with the continuous measure of $\%\Delta FTV_{1\rightarrow4}$. Two exploratory analyses were also

performed to determine if the early changes in either spectroscopically-measured water T_2 ($\Delta T_2w_{1\rightarrow2} = T_2w_2\text{-}T_2w_1$) or MRI-measured FTV ($\%\Delta FTV_{1\rightarrow2}$) were associated with pCR.

To understand which factors lead to unmeasurable tCho at TP1, preventing the calculation of $\Delta[tCho]_{1\rightarrow2}$ and $\%\Delta[tCho]_{1\rightarrow2}$, a set of 15 potentially explanatory factors that reflect the MRS measurement quality (voxel size, SNR_{water}, water line width, lipid fraction, field strength, imaging system, voxel placement confidence) and lesion biology (age, race, pCR status, radiologic response status, receptor status (HR+/HER2+/TN), breast density, lesion type (mass vs non-mass), and morphologic sub-classification) were identified and independently evaluated using logistic regression models.

For the primary and secondary analyses, p-values less than 0.05 were considered significant. Exploratory analyses also used an alpha of 0.05 to identify potential effects. All statistical data analyses were performed with SAS version 9.4 (SAS Institute, Cary, North Carolina, USA).

Results

Inclusion and Data Quality

119 subjects were enrolled in the extension trial. Seven subjects withdrew before completing the study, 3 were ineligible due to development of metastases, 2 received a non-compliant treatment regimen, and 5 were imaged at incorrect timepoints. MR spectroscopy data at either the first or second time-points was not acquired (n=20) or lost (n=15) in 35 of the remaining 102 cases, leaving 67 cases with MRS data available at both TP1 and TP2. Of these, 16 had poor quality MRS at either TP1 or TP2, and 18 had no measurable tCho at either TP1 (n=15) or TP2 (n=3), leaving 33 cases with analyzable MRS. In addressing the primary outcome of pathologic response, 4 subjects did not have final pathology data available, leaving 29 cases in the analysis set with a pCR prevalence of 28% (8/29). In addressing the secondary outcome of radiologic response, 5 did not have acceptable volume data at either TP1 or TP4, leaving 28 cases in the analysis set. Table 1 presents the participant characteristics of the study population and these two analysis sets. No significant differences were observed between any characteristics of the excluded vs. the included cases for each set ($p>0.05$).

Pathologic Response

Table 2 and Figure 2a–b present the descriptive statistics of [tCho] at the first two time points for the 29 subjects, stratified by pCR status. They show the general trend that [tCho] decreases from TP1 to TP2, with little difference between the two response groups. The AUC for using $\Delta[tCho]_{1\rightarrow2}$ to predict pCR was 0.53 (95% CI: [0.27, 0.79]), indicating poor predictive accuracy. The optimal cutpoint for dichotomizing $\Delta[tCho]_{1\rightarrow2}$ was −0.475 mmol/kg, with sensitivity = 62.5%, specificity = 57.1%, positive predictive value 80.0%, and negative predictive value = 35.7%. No significant difference in pCR rates was found between subjects with $\Delta[tCho]_{1\rightarrow2}$ above or below this cutpoint using Fisher's exact test ($p=0.43$). The univariate logistic model using $\Delta[tCho]_{1\rightarrow2}$ to predict pCR response showed no significant effect ($p=0.83$, OR=0.918, 95% CI: [0.42, 2.02]).

The relative change in tCho (%$\Delta[tCho]_{1\rightarrow2}$) also showed no significant associations with pCR. The AUC for using %$\Delta[tCho]_{1\rightarrow2}$ to predict pCR was 0.51 (95% CI: [0.24, 0.78]), indicating poor predictive accuracy. The optimal cutpoint for %$\Delta[tCho]_{1\rightarrow2}$ was −47.7%, giving a sensitivity = 37.5%, specificity = 80.9%, positive predictive value 42.9%, and negative predictive value = 77.3%. No significant difference in pCR rates was found (p=0.36) between subjects with relative changes (%$\Delta[tCho]_{1\rightarrow2}$) above or below the optimal cutpoint. The univariate logistic model using %$\Delta[tCho]_{1\rightarrow2}$ showed no significant effect (p=0.99, OR=0.98, 95%CI: [0.103, 9.416]).

Radiologic Response

The median change in functional tumor volume from TP1 to TP4 (%$\Delta FTV_{1\rightarrow4}$) was −94.75%; subjects with a greater decrease than this median value were considered radiologic responders. Table 2 and Figure 2c–d show the values of [tCho] at the first two time points, shown in separate groups by radiologic response status. The AUC analysis indicated that $\Delta[tCho]_{1\rightarrow2}$ was a poor predictor of %$\Delta FTV_{1\rightarrow4}$ (AUC = 0.51, 95% CI: [0.27, 0.75]). The optimal cutpoint for $\Delta[tCho]_{1\rightarrow2}$ in this analysis set was 0.562 mmol/kg. Dichotomizing at this threshold gave a significant difference (p=0.04) between responders and non-responders using Fisher's exact test, with sensitivity = 100%, specificity = 35.7%, positive predictive value 60.9%, and negative predictive value = 100%. While the overall AUC was poor, all 5 subjects with a [tCho] increase greater than 0.562 mmol/kg were non-responders, leading to the high sensitivity at the optimal cutpoint. However, the logistic regression using continuous $\Delta[tCho]_{1\rightarrow2}$ found no significant association with radiologic response using a univariate (p=0.73, OR=0.91, 95% CI: [0.53,1.57]) model. No significant correlation was found between $\Delta[tCho]_{1\rightarrow2}$ and %$\Delta FTV_{1\rightarrow4}$, as shown in Figure 3 (Spearman $\rho = -0.07$; $p = 0.72$).

Repeating the above analyses using the relative change in tCho (%$\Delta[tCho]_{1\rightarrow2}$) showed that it was a poor predictor of %$\Delta FTV_{1\rightarrow4}$ (AUC = 0.52, 95% CI: [0.29, 0.74]) and showed no significance using either Fisher's exact test (p=0.38) at the optimal cutoff (20.6%), univariate logistic regression (p=0.42, OR=0.52, 95% CI: [0.11, 2.57]), or Spearman's correlation (Spearman $\rho = -0.04$; $p = 0.84$).

Unmeasurable Pre-treatment tCho

The inability to measure the tCho level at the pre-treatment time-point reduced the amount of analyzable data for all outcomes tested. Of the 15 potential explanatory factors listed above, only 4 were significant when tested with univariate logistic regression: lipid fraction (p=<.001), voxel size (p=0.04), water line width (p=0.002), and field strength (p=0.006). No tCho was detected in voxels with lipid fraction > 33%; for voxels with lipid fraction ≤ 33% the detection rate was 66%. The tCho detection rate was lower in voxels larger than the median size (50% for voxels ≥ 7.7 mL; 71% < 7.7 mL), and in voxels with broader water line widths (33% for water line width > 0.25ppm; 73% for water line width ≤ 0.25 ppm) The detection rate at 1.5T (53%) was lower than that at 3T (93%).

Exploratory Analyses: Early changes in Water T2 and FTV

Larger analysis sets were available for assessing two exploratory predictors of pCR, $\Delta T_2w_{1\rightarrow 2}$ and $\%\Delta FTV_{1\rightarrow 2}$. Good quality T_2w measurements were available for both TP1 and TP2 in 60 cases (59%, 60/102), and 56 of these also had a known pCR status. The change in water T_2 ($\Delta T_2w_{1\rightarrow 2}$) showed a significant association (p=0.004) with pCR when dichotomized at the optimal cutpoint (cutoff = −6.9 ms), with AUC = 59.7%. FTV measurements were available for both TP1 and TP2 in 68 cases (67%. 68/102), and 64 of these had a known pCR status. The early decrease in FTV ($\%\Delta FTV_{1\rightarrow 2}$) showed a significant association (p=0.007) with pCR when dichotomized at the optimal cutpoint (cutoff = −24.7%), with AUC = 64.3%.

Discussion

The ACRIN 6657 extension trial was designed to estimate the accuracy of predicting pathologic and radiologic response to NACT using a change in tCho concentration measured by MR spectroscopy within the first 1–4 days after beginning chemotherapy. Due primarily to the technical challenges of acquiring quantitative breast MRS in a multi-site clinical trial setting, only 28% (29/102) of the eligible cases were usable for the primary analysis, leading to substantially reduced statistical power. Thus, this study was not able to measure the accuracy of predicting pathologic response with MRS with sufficient precision to assess its clinical utility. There was some evidence of a relationship between tCho change and radiologic response: while the estimate of the predictive value was not significantly different than 0.5, there was a statistically significant difference in radiologic outcome when dichotomizing tCho change at the optimal cutpoint. This is attributable to those cases where tCho increased the most: the five subjects with the largest increase in tCho were radiologic non-responders. While a positive finding this should be interpreted cautiously, as there was no overall correlation between tCho change and radiologic response, and ROC analyses can lead to overoptimistic results with small sample sizes (30). Previous studies using tCho to predict treatment response in breast cancer have generally found that a decrease in tCho indicates a favorable response to chemotherapy. An important difference between this trial and the majority of previous studies is the timing of the first post-treatment MRS measurement. This trial measured patients 1–4 days after the first treatment, which reflects a more acute response than studies that measured tCho after completion of the first cycle of chemotherapy (2–4 weeks) (9–14, 31). The only other study that looked at a comparably acute post-treatment timing was Meisamy et al. (17), which had a post-treatment (TP2) timing window of approximately 12–36 hours and reported 100% accuracy for predicting radiologic response (defined by tumor longest diameter measurements). This trial used a wider TP2 window to ease patient scheduling and allow analysis of time variation within that window (which was not possible with the limited data available). It may be that the acute timing used here is too early to reliably detect metabolic changes, or that the optimal TP2 timing depends on the therapeutic agent.

Our trial also found a relationship between tCho change and radiologic response based on functional tumor volume, but the predictive value was poor. An alternate analysis of this trial’s data using the lesion longest diameter (measured at each site) as outcome also showed

no significant findings (data not shown). The primary difference between these two studies was the environment: Meisamy et al. was a small (n=13), single-site pilot study using a non-standard 4 T MR system and methods, whereas this trial was performed across nine sites with standard clinical 1.5 or 3.0 T MR systems, coils, and software from multiple vendors. Along with the lower field strengths, the heterogeneity in systems, sequences, and operators may have contributed to lower performance of this trial.

The low yield of analyzable data was the primary limitation of this study. This led to wide confidence intervals and made relevant subgroup analyses (e.g., grouping by receptor status, timing of TP2, field strength, highest data quality, post-biopsy timing) impractical. Three factors accounted for the majority of this data loss, each of which is attributable to the technical challenge of acquiring quantitative MRS data, despite protocol-specific site qualification and training. First, sites did not submit MRS data for both TP1 and TP2 in 35 of 102 eligible cases (34% loss). This is attributable to the difficulty of acquiring and managing the MRS data within the clinical workflow. In 20 cases the MRS acquisitions were not completed at either the TP1 or TP2 timepoints, due primarily to operator error and patient compliance. In 15 cases the MRS data were acquired but lost before submission, which is partly attributable to insufficient scanner and PACS support for MRS data formats. Additionally, the strict timing required for the TP2 exam (20–96 hrs) may have impacted protocol compliance. Second, poor quality MRS data eliminated 24% (16/67) of cases from the analysis sets. The low quality was attributed to multiple factors, including errors in scanner adjustments (transmit power, B0, and frequency), poor voxel placement and/or localization quality, and patient motion. Five cases were eliminated due to a single system that produced inconsistent data, as determined retrospectively from analysis of the phantom quality control data. Third, a tCho resonance was not detected using the specified objective criteria for either TP1 or TP2 in 37% (18/49) of the cases otherwise eligible for analysis. This non-detection rate is comparable to the 40% reported by Bathen et al. (31). Large voxel sizes, high lipid fractions, and large water line widths (indicating poor B0 shimming) were associated with non-detection of tCho, suggesting that voxel placement and avoidance of adipose tissues were challenging. The presence of biopsy markers, which were used in all patients, may also have complicated voxel placement and/or impacted the spectral quality. The tCho detection rate was higher at 3T than 1.5T, suggesting that the higher signal-to-noise and spectral resolution improved detectability.

The two exploratory metrics, $\Delta T_2w_{1\rightarrow2}$ and $\%\Delta FTV_{1\rightarrow2}$, were more robustly measurable than $\Delta[tCho]_{1\rightarrow2}$ and thus allowed for substantially larger analysis sets. Decreases in T_2w and FTV measured at TP2 were both found to have a significant association with pathologic response. These findings indicate that MR-measurable changes in tumor vascularity (reflected in FTV) and microstructure (reflected in T_2w) occurring within 1–4 days of initiating chemotherapy can assess drug efficacy, and may be useful for adapting chemotherapy regimens.

The findings of this trial suggest several recommendations for future development and assessment of breast MRS. First, real-time placement of the acquisition voxel is challenging and operator-dependent, and likely contributed to data loss and variability between sites. Options to reduce this variability include automated voxel placement, increased training and

oversight of operators, and/or spectroscopic imaging methods that allow for retrospective region selection (14, 32–34). Second, the inclusion of variable MR systems in this study required flexibility in the acquisition protocol and contributed to variability between sites, and likely gave lower performance than a study using controlled scanning platforms. This heterogeneous design is widely used successfully for imaging studies, and at the time of study design the number of encouraging single-site breast MRS studies led expert groups to recommend multi-site, multi-vendor studies of breast MRS (35). In retrospect, future studies would benefit by more rigorous standardization of acquisition methods, including limiting the eligible systems to high-field (3 T) scanners, which gave better detection performance and are more widely available today. While this trial did implement a quality control program using initial qualification and ongoing monitoring (see \appendix), future trials would benefit from stronger quality controls, including more timely feedback to enforce data quality and protocol compliance. Third, the results from this study do not provide evidence that acute post-treatment timing is better than later timepoints. While it is preferable to detect non-response as early as possible, study designs that acquire MRI/MRS after a full cycle of therapy may provide better performance while simplifying patient scheduling. Finally, while a decrease in tCho could not identify patients who would achieve pCR, the lack of an early decrease in tCho was associated with a lack of change in tumor size. While this study was too small to analyze specific cancer subtypes (e.g., HR+, HER+, TN), it is possible that tCho changes may provide more meaningful information in specific subgroups. Our results could be used to design further studies using more controlled MRS methods to assess response in specific breast cancer subgroups.

Acknowledgments

The authors acknowledge those individuals who have contributed substantially to the work reported in the manuscript, including the ACRIN 6657 Trial Team, the I-SPY TRIAL Investigators Network, the patients who participated in the study, and the staff members who contributed to the conduct of the study at the University of California at San Francisco, the University of Minnesota, the University of Pennsylvania, the University of North Carolina at Chapel Hill, Georgetown University, Memorial Sloan-Kettering Cancer Center, the University of Texas Southwestern Medical Center, the University of Washington, and the University of Chicago. The authors also gratefully acknowledge Ralph Noeske at General Electric and Marcus Alley at Stanford University for their assistance and software.

Grant Support

This trial was supported by the National Cancer Institute's grants to ACRIN (U01 CA079778, U01 CA080098), ECOG-ACRIN (U10 CA180794), and CALGB/ISPY (CA31964, CA33601, CA58207). Additional support was provided by NCI R01 CA120509, NCRR P41 RR00879, and NIBIB P41 EB015894.

References

1. Buchholz TA, Lehman CD, Harris JR, et al. Statement of the Science Concerning Locoregional Treatments After Preoperative Chemotherapy for Breast Cancer: A National Cancer Institute Conference. J Clin Oncol. 2008; 26:791–797. [PubMed: 18258988]
2. Killelea BK, Yang VQ, Mougalian S, et al. Neoadjuvant Chemotherapy for Breast Cancer Increases the Rate of Breast Conservation: Results from the National Cancer Database. J Am Coll Surg. 2015; 220:1063–1069. [PubMed: 25868410]
3. 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]

4. 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]
5. Cho N, Im S-A, Park I-A, et al. Breast Cancer: Early Prediction of Response to Neoadjuvant Chemotherapy Using Parametric Response Maps for MR Imaging. Radiology. 2014; 272:385–396. [PubMed: 24738612]
6. Podo F. Tumour phospholipid metabolism. NMR Biomed. 1999; 12:413–439. [PubMed: 10654290]
7. Glunde K, Bhujwalla ZM, Ronen SM. Choline metabolism in malignant transformation. Nat Rev Cancer. 2011; 11:835–848. [PubMed: 22089420]
8. Kvistad KA, Bakken IJ, Gribbestad IS, et al. Characterization of neoplastic and normal human breast tissues with in vivo (1)H MR spectroscopy. J Magn Reson Imaging JMRI. 1999; 10:159–164. [PubMed: 10441019]
9. Baek HM, Chen JH, Nie K, et al. Predicting Pathologic Response to Neoadjuvant Chemotherapy in Breast Cancer by Using MR Imaging and Quantitative 1H MR Spectroscopy. Radiology. 2009; 251:653. [PubMed: 19276320]
10. Tozaki M, Sakamoto M, Oyama Y, Maruyama K, Fukuma E. Predicting pathological response to neoadjuvant chemotherapy in breast cancer with quantitative ^{<FONT SIZE='-1'>1</FONT>}H MR spectroscopy using the external standard method. J Magn Reson Imaging. 2010; 31:895–902. [PubMed: 20373434]
11. Tozaki M, Oyama Y, Fukuma E. Preliminary study of early response to neoadjuvant chemotherapy after the first cycle in breast cancer: comparison of 1H magnetic resonance spectroscopy with diffusion magnetic resonance imaging. Jpn J Radiol. 2010; 28:101–109. [PubMed: 20182844]
12. Jacobs MA, Stearns V, Wolff AC, et al. Multiparametric Magnetic Resonance Imaging, Spectroscopy and Multinuclear (Na-23) Imaging Monitoring of Preoperative Chemotherapy for Locally Advanced Breast Cancer. Acad Radiol. 2010; 17:1477–1485. [PubMed: 20863721]
13. Baek H-M, Chen J-H, Nalcioglu O, Su M-Y. Proton MR spectroscopy for monitoring early treatment response of breast cancer to neo-adjuvant chemotherapy. Ann Oncol Off J Eur Soc Med Oncol ESMO. 2008; 19:1022–1024.
14. Danishad KKA, Sharma U, Sah RG, Seenu V, Parshad R, Jagannathan NR. Assessment of therapeutic response of locally advanced breast cancer (LABC) patients undergoing neoadjuvant chemotherapy (NACT) monitored using sequential magnetic resonance spectroscopic imaging (MRSI). NMR Biomed. 2010; 23:233–241. [PubMed: 20175134]
15. Jagannathan NR, Kumar M, Seenu V, et al. Evaluation of total choline from in-vivo volume localized proton MR spectroscopy and its response to neoadjuvant chemotherapy in locally advanced breast cancer. Br J Cancer. 2001; 84:1016–1022. [PubMed: 11308247]
16. Morse DL, Gray H, Payne CM, Gillies RJ. Docetaxel induces cell death through mitotic catastrophe in human breast cancer cells. Am Assoc Cancer Res. 2005; 4:1495–1504.
17. Meisamy S, Bolan PJ, Baker EH, et al. Neoadjuvant chemotherapy of locally advanced breast cancer: predicting response with in vivo (1)H MR spectroscopy--a pilot study at 4 T. Radiology. 2004; 233:424–431. [PubMed: 15516615]
18. Bolan, PJ., Garwood, M., Rosen, MA., et al. Design of Quality Control Measures for a Multi-Site Clinical Trial of Breast MRS - ACRIN 6657. Proc 16th Annu Meet ISMRM; Toronto. 2008. p. 1588
19. Bolan PJ, DelaBarre L, Baker EH, et al. Eliminating spurious lipid sidebands in 1H MRS of breast lesions. Magn Reson Med. 2002; 48:215–222. [PubMed: 12210929]
20. Bottomley PA. Spatial Localization in NMR Spectroscopy in Vivo. Ann N Y Acad Sci. 1987; 508:333–348. [PubMed: 3326459]
21. 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. 2015:150013.
22. Bolan PJ, Meisamy S, Baker EH, et al. In vivo quantification of choline compounds in the breast with 1H MR spectroscopy. Magn Reson Med. 2003; 50:1134–1143. [PubMed: 14648561]

23. Bolan PJ, Henry PG, Baker EH, Meisamy S, Garwood M. Measurement and correction of respiration-induced B0 variations in breast 1H MRS at 4 Tesla. Magn Reson Med. 2004; 52:1239–45. [PubMed: 15562472]
24. Baik HM, Su MY, Yu H, Mehta R, Nalcioglu O. Quantification of choline-containing compounds in malignant breast tumors by 1H MR spectroscopy using water as an internal reference at 1. 5 T. Magma. 2006; 19:96–104. [PubMed: 16779565]
25. Bakken IJ, Gribbestad IS, Singstad TE, Kvistad KA. External standard method for the in vivo quantification of choline-containing compounds in breast tumors by proton MR spectroscopy at 1. 5 Tesla. Magn Reson Med Off J Soc Magn Reson Med Soc Magn Reson Med. 2001; 46:189–192.
26. Tan, PC., Lowry, M., Manton, DJ., Turnbull, LW. Evaluation of choline concentrations in malignant breast lesions in predicting response to neoadjuvant chemotherapy. Proc 14th Annu ISMRM; Seattle. 2006. p. 574
27. Li Y, Srinivasan R, Ratiney H, Lu Y, Chang SM, Nelson SJ. Comparison of T1 and T2 metabolite relaxation times in glioma and normal brain at 3T. J Magn Reson Imaging. 2008; 28:342–350. [PubMed: 18666155]
28. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]
29. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983; 148:839–843. [PubMed: 6878708]
30. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. Clin Chem. 2008; 54:729–737. [PubMed: 18258670]
31. Bathen TF, Heldahl MG, Sitter B, et al. In vivo MRS of locally advanced breast cancer: characteristics related to negative or positive choline detection and early monitoring of treatment response. Magma N Y N. 2011; 24:347–357.
32. Sijens PE, Dorrius MD, Kappert P, Baron P, Pijnappel RM, Oudkerk M. Quantitative multivoxel proton chemical shift imaging of the breast. Magn Reson Imaging. 2010; 28:314–319. [PubMed: 20071119]
33. Gruber S, Debski B-K, Pinker K, et al. Three-dimensional proton MR spectroscopic imaging at 3 T for the differentiation of benign and malignant breast lesions. Radiology. 2011; 261:752–761. [PubMed: 21998046]
34. Jacobs MA, Barker PB, Argani P, Ouwerkerk R, Bhujwalla ZM, Bluemke DA. Combined dynamic contrast enhanced breast MR and proton spectroscopic imaging: a feasibility study. J Magn Reson Imaging JMRI. 2005; 21:23–28. [PubMed: 15611934]
35. Evelhoch J, Garwood M, Vigneron D, et al. Expanding the use of magnetic resonance in the assessment of tumor response to therapy: workshop report. Cancer Res. 2005; 65:7041–4. [PubMed: 16103049]

Appendix

MRS Quality Control

As this trial was designed to test breast MRS in a heterogeneous clinical setting, which included different MR scanners, field strengths, breast coils, and software capabilities, several quality control measures were used to maximize data consistency. These measures were described prospectively in the trial protocol and a conference abstract (18).

A pair of trial-specific test phantoms were produced in a single batch and distributed to all sites. The *normal* phantom consisted of a 2-liter leak-proof bottle containing vegetable oil and a 40 mm outer diameter plastic sphere mounted on a post ~2″ above the bottom. The sphere contained 1 mM phosphocholine, 0.2 mM Gd-DTPA, 10 mM deuterated TSP as a frequency reference, and 0.1% sodium azide. The *control* phantom was identical to the

normal phantom except it contained no phosphocholine. Prior to enrolling human subjects, each site was required to submit an acceptable *Entry QC* scan of the normal and control phantoms for each *system* that was to be used in the trial. A system was defined as the combination of MR scanner, breast coil and software. For example, when a scanner received a software update it was treated as a new system, and a new Entry QC submission was required prior to acquiring additional in vivo data.

The Entry QC scan consisted of six scans: a water reference scan, a metabolite scan, and a metabolite scan without water suppression acquired from a 20 × 20 × 20 mm voxel placed in the sphere of both the normal and control phantoms. To be considered acceptable the data were required to show a water-fat ratio > 2 (indicating acceptable localization performance), water linewidth < 0.25 ppm (indicating system shimming performance), a detectable tCho resonance in the normal phantom, and no detectable tCho resonance in the control phantom. Additionally, the system sidebands were measured using the metabolite scan without water suppression; if the water sidebands were >= 0.002% in the 1.2 – 2.7 ppm range around the water peak, then either TE averaging or lipid suppression would be required to avoid the possibility of a sideband from a large lipid peak interfering with the tCho measurement (19). The Entry QC testing was used iteratively to determine which acquisition variations (+/− fat suppression, single TE vs TE averaging) gave the best performance on a per-system basis. Details for each system used in the trial are provided in table A.2.

After qualifying for the study, regular biweekly quality control scans were acquired for each system to measure within-system phantom reproducibility, which reflects the phantom measurement precision for a longitudinal study. The mean and coefficient of variation (standard deviation / mean) of the biweekly [tCho] measurements for each system are also provided in table A.2.

In addition to phantom measurements, in vivo measurements were assessed with subjective quality scoring. All spectra for the study were assessed during the final analysis in random order by an experienced spectroscopist blinded to imaging and outcome information. Spectra were assigned scores of good, fair, or poor based on the presence of artifacts, the quality of the data (linewidth, lineshape, and baseline), and the quality of the spectral fitting. Specific findings, such as poor baseline, poor fit, etc., were recorded. All spectra with “poor” scores were eliminated from analysis sets, while both “good” and “fair” datasets were included. Of the 165 spectral datasets submitted for TP1 and TP2, 47 (28%) were assigned good quality, 85 (52%) fair, and 33 (20%) poor quality. The poor quality scores were attributed to an unstable software upgrade on a single system (12), a RF transmit calibration error on a single system (9), inconsistent adjustments between water and metabolite scans (5), incomplete acquisitions (3), poor shim and baseline distortions (3), and motion (1).

Table A.1

Individual participant characteristics by site.

	Site	Field Strength	TP2 Timing	Tumor LD baseline (cm)	FTV baseline (cm^3)	Receptor Status	Mass/NMLE	in pCR analysis set	in FTV Analysis set
1	**A**	3 T	20–28 hours	7.5	30.46	HER2+	NMLE	Yes	Yes
2	**A**	3 T	20–28 hours	8.1	45.47	HER2+	NMLE	Yes	No
3	**A**	3 T	20–28 hours	5.5	29.86	TN	Mass	Yes	Yes
4	**A**	3 T	20–28 hours	4.7	35.39	HR+	Mass	Yes	Yes
5	**A**	3 T	40–96 hours	7.1	40.17	HER2+	Mass	Yes	Yes
6	**A**	3 T	40–96 hours	10.1	N/A	N/A	Mass	Yes	No
7	**B**	1.5 T	20–28 hours	9.6	40.68	HER2+	Mass	Yes	Yes
8	**B**	1.5 T	40–96 hours	4.2	N/A	HR+	Mass	Yes	No
9	**C**	1.5 T	40–96 hours	5.9	9.46	TN	Mass	Yes	Yes
10	**D**	1.5 T	20–28 hours	3.3	38.99	HR+	Mass	Yes	Yes
11	**D**	1.5 T	20–28 hours	3.1	6.54	TN	Mass	Yes	Yes
12	**E**	3 T	40–96 hours	3.6	13.20	HER2+	Mass	No	Yes
13	**E**	3 T	40–96 hours	4.5	15.23	HR+	Mass	No	Yes
14	**F**	1.5 T	40–96 hours	5.0	122.35	TN	Mass	Yes	Yes
15	**F**	1.5 T	20–28 hours	3.7	42.84	HER2+	Mass	Yes	Yes
16	**F**	1.5 T	40–96 hours	3.8	84.74	HR+	Mass	Yes	Yes
17	**F**	1.5 T	40–96 hours	7.3	229.48	HR+	Mass	Yes	Yes
18	**F**	1.5 T	40–96 hours	3.4	14.44	HER2+	Mass	Yes	Yes
19	**F**	1.5 T	40–96 hours	3.5	12.08	TN	Mass	Yes	Yes
20	**G**	1.5 T	20–28 hours	4.8	2.03	HR+	Mass	Yes	Yes
21	**G**	1.5 T	20–28 hours	6.4	15.29	HER2+	Mass	Yes	Yes
22	**G**	1.5 T	40–96 hours	2.8	3.77	HER2+	Mass	Yes	Yes
23	**G**	1.5 T	40–96 hours	4.0	19.54	N/A	Mass	Yes	Yes
24	**G**	1.5 T	20–28 hours	3.0	2.81	HER2+	Mass	Yes	Yes
25	**G**	1.5 T	40–96 hours	2.8	7.28	TN	Mass	Yes	Yes
26	**G**	1.5 T	20–28 hours	3.6	12.48	TN	Mass	Yes	Yes
27	**G**	1.5 T	40–96 hours	2.7	6.18	TN	Mass	Yes	Yes
28	**G**	1.5 T	20–28 hours	9.2	9.57	HR+	NMLE	Yes	Yes
29	**G**	1.5 T	20–28 hours	6.7	43.66	HR+	NMLE	Yes	Yes
30	**G**	1.5 T	40–96 hours	3.3	6.18	TN	Mass	Yes	Yes
31	**G**	1.5 T	40–96 hours	4.0	12.96	HR+	Mass	Yes	No
32	**G**	1.5 T	40–96 hours	3.1	0.78	TN	Mass	No	Yes

HR+: estrogen and/or progresterone receptor positive and HER2 negative

HER2+: estrogen and/or progresterone receptor negative and HER2 positive

TN: estrogen and/or progresterone receptor negative and HER2 negative

N/A: not available

NMLE: Non-mass localized enhancement

Table A.2

Characteristics of MR Systems used to acquire in vivo data for the trial.

Site	System	Field Strength	Scanner Vendor	Software Version	Breast Coil	Sequence Variant	QC Mean [tCho] (mmol/Kg)	QC CV (%)
A	**A.1**	3 T	Siemens	vb15	In Vivo 4ch	Fixed TE, fat sat	0.97	15.8%
A	**A.2**	3 T	Siemens	vb17	In Vivo 4ch	Fixed TE, fat sat	0.90	14.4%
B	**B.1**	1.5 T	Siemens	vb17	In Vivo 7ch	Fixed TE, fat sat	0.69	25.9%
C	**C.1**	1.5 T	GE	HDx 14.3	HD Breast 8ch	TE Ave, no fat sat	0.92	14.1%
D	**D.1**	1.5 T	Siemens	vb15	In Vivo 7ch	Fixed TE, fat sat	0.72	7.9%
D	**D.2**	1.5 T	Siemens	vb17	Sentinelle 8ch	Fixed TE, fat sat	0.76	8.8%
E	**E.1**	3 T	Philips	2.5.1	In Vivo 7ch	Fixed TE, fat sat	0.96	12.1%
E	**E.2**	3 T	Philips	2.6.1	In Vivo 7ch	Fixed TE, fat sat	0.97	25.1%
F	**F.1**	1.5 T	Philips	2.1.3	In Vivo 7ch	Fixed TE, fat sat	0.68	31.5%
F	**F.2**	1.5 T	Philips	2.6.1	SENSE-16M	Fixed TE, fat sat	0.78	24.5%
F	**F.3**	1.5 T	Philips	2.6.3	SENSE-16M	Fixed TE, fat sat	1.01	29.4%
F	**F.4** *	1.5 T	Philips	3.2.1	SENSE-16M	Fixed TE, fat sat	3.38	83.2%
G	**G.1**	1.5 T	GE	HDx 14.3	HD Breast 8ch	TE Ave, no fat sat	0.65	-
G	**G.2**	1.5 T	GE	HDx 14.3	Sentinelle 8ch	TE Ave, no fat sat	0.87	24.7%
G	**G.3**	1.5 T	GE	HDx 14.3	Sentinelle 8ch	TE Ave, no fat sat	0.80	14.5%
G	**G.4**	1.5 T	GE	HDxt 15.0	Sentinelle 8ch	TE Ave, no fat sat	0.80	13.8%

*System F.4 was retrospectively found to perform inconsistently and its data were excluded from analysis sets

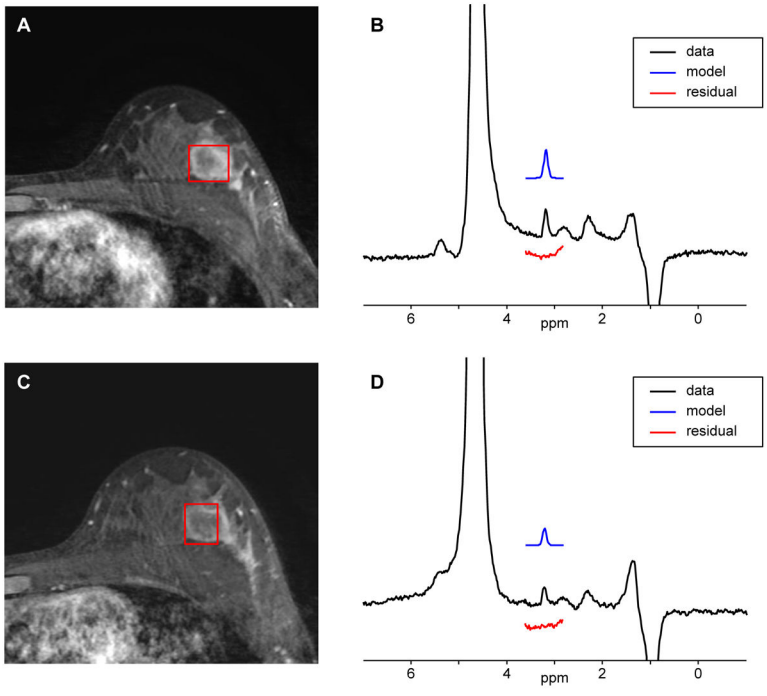

Figure 1.

Representative case showing imaging and MRS data. Voxel positioning for a) TP1 (pre-treatment) and c) TP2 is shown on the first post-contrast T1-weighted image. The corresponding spectra are shown in b) and d) show a decrease in [tCho] from 1.81 mmol/kg to 1.17 mmol/kg. This patient was considered a non-responder by both pathologic and radiologic criteria.

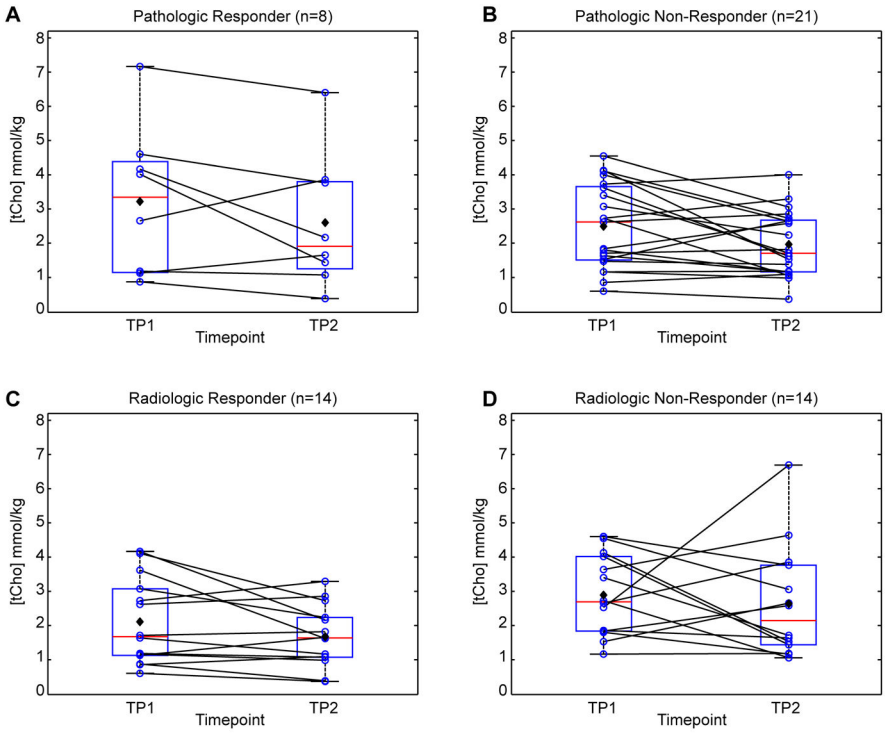

Figure 2.

Concentrations of tCho at time points 1 and 2, separated out by pathologic response (a,b) and radiologic response (c,d). Box plots indicate median (red line), mean (black diamond), and 25th and 75th percentiles (box limits).

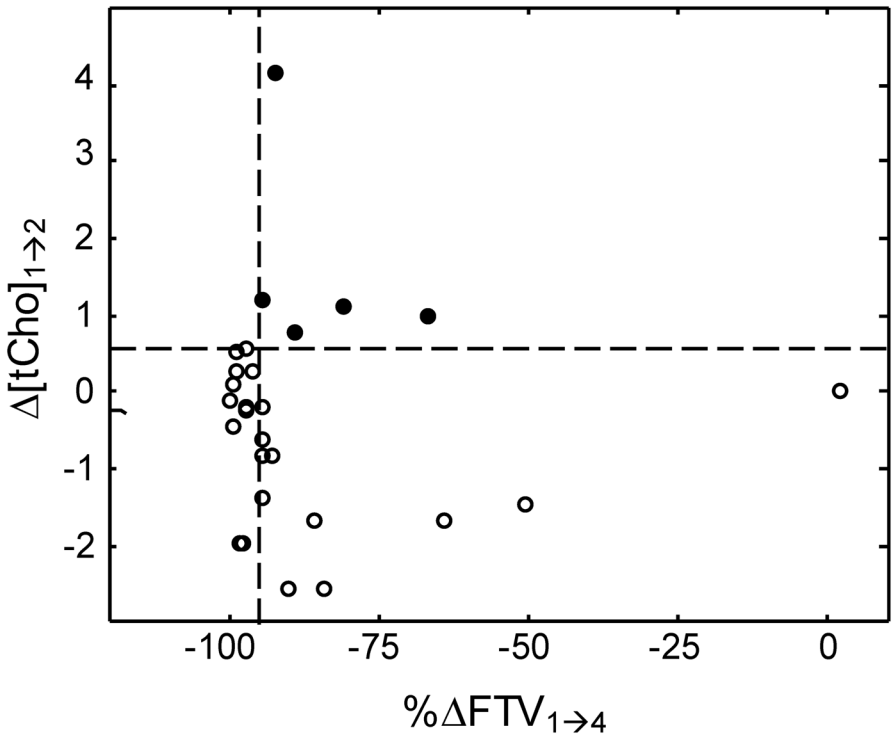

Figure 3.

Relationship between $\Delta[tCho]_{1\rightarrow2}$ and $\%\Delta FTV_{1\rightarrow4}$, showing no significant correlation (Spearman $\rho = -0.07$; $p = 0.72$). The vertical dotted line is the median volume change (−94.75%); the horizontal dotted line is the optimal cutoff for $\Delta[tCho]_{1\rightarrow2}$ (0.562 mmol/kg). The five cases with the largest increase in tCho are indicated by filled circles (see text for details).

J Magn Reson Imaging. Author manuscript; available in PMC 2018 July 01.

Table 1

Participant Characteristics for pathologic and radiologic response analysis sets.

Characteristic	Eligible set (N=119)		pCR Response Analysis set (N=29)		Radiologic Response Analysis set (N=28)	
Age						
Median (Range)	49 (23–76)		49 (24–76)		50 (24–76)	
Mean (SD)	48.8 (11.5)		47.2 (11.0)		48.9 (12.1)	
Age P-value *			***0.3793***		***0.9639***	
Race						
White	74	(62%)	17	(59%)	17	(61%)
Non-White	45	(38%)	12	(41%)	11	(39%)
Race P-value *				***0.6652***		***1.0000***
Ethnicity						
Not Hispanic or Latino	106	(89%)	27	(93%)	26	(93%)
Hispanic or Latino	10	(8%)	-	-	1	(4%)
Unknown	3	(3%)	2	(7%)	1	(4%)
Ethnicity P-value *				***0.1136***		***0.4490***
Hormonal Status						
HER2+	26	(22%)	9	(31%)	9	(32%)
HR+ & HER2-	40	(34%)	9	(31%)	8	(29%)
Triple Negative (HR- & HER2-)	25	(21%)	9	(31%)	10	(36%)
Unknown/Missing	28	(24%)	2	(7%)	1	(4%)
Hormonal Status P-value *				***0.4335***		***0.1784***
Invasive histology						
Not Applicable	30	(25%)	9	(31%)	9	(32%)
Ductal carcinoma	65	(55%)	16	(55%)	17	(61%)
Lobular carcinoma	7	(6%)	3	(10%)	1	(4%)
Mixed ductal/lobular carcinoma	1	(1%)	-	-	-	-
Other	2	(2%)	1	(3%)	1	(4%)
Unknown/Missing	14	(12%)	-	-	-	-

Characteristic	Eligible set (N=119)		pCR Response Analysis set (N=29)		Radiologic Response Analysis set (N=28)	
Invasive histology P-value*				***0.6078***		***0.7554***
DCIS Present						
Yes	54	(45%)	13	(44%)	13	(46%)
No	52	(44%)	16	(55%)	15	(54%)
Unknown/Missing	13	(11%)	-	-	-	-
DCIS Present P-value*				***0.5157***		***0.6616***
Surgical Pathology						
Complete Responder	26	(22%)	8	(28%)	7	(25%)
Residual Disease	71	(60%)	21	(72%)	18	(64%)
Unknown/Missing	22	(18%)	-	-	3	(11%)
Surgical Pathology P-value*				***1.000***		***1.000***
Surgery Type						
Mastectomy, NOS	55	(46%)	12	(41%)	10	(36%)
Lumpectomy	51	(43%)	17	(59%)	18	(64%)
No Surgery	13	(11%)	-	-	-	-
Surgery Type P-value*				***0.1988***		***0.0510***
Lesion Type						
Mass	94	(79%)	25	(86%)	25	(89%)
Non-Mass Localized Enhancement	22	(19%)	4	(14%)	3	(11%)
Unknown	3	(3%)	-	-	-	-
Lesion Type P-value*				***0.5857***		***0.2728***

* P-values reported are for the comparison of the distribution of the analysis set cases versus the excluded cases (eligible set - analysis set), using t-test for age and Fisher's Exact Test for all categorical measures.

Table 2

Descriptive statistics for the tCho concentration stratified by response status.

	Pathologic Response		Radiologic Response	
	Responder (n=8)	**Non-responder (n=21)**	**Responder (n=14)**	**Non-responder (n=14)**
$[tCho]_1$ (mmol/kg)	3.22 ± (2.18) [0.9 – 7.2]	2.50 ± (1.22) [0.6 – 4.5]	2.11 ± (1.26) [0.6 – 4.2]	2.89 ± (1.16) [1.2 – 4.6]
$[tCho]_2$ (mmol/kg)	2.60 ± (1.96) [0.40 – 6.4]	1.96 ± (0.94) [0.4 – 4.0]	1.68 ± (0.90) [0.4 – 3.3]	2.64 ± (1.63) [1.1 – 6.7]
$\Delta[tCho]_{1\rightarrow2}$ (mmol/kg)	−0.63 ± (1.23) [−2.6 – 1.2]	−0.54 ± (1.00) [−2.6 – 1.1]	−0.43 ± (0.84) [−2.0 – 0.6]	−0.25 ± (1.82) [−2.6 – 4.2]

Note: Values in each category show mean tCho concentration ± (standard deviation), and the range [min – max] shown in brackets.