Tissue oxygen saturation predicts response to breast cancer neoadjuvant chemotherapy within 10 days of treatment

Jeffrey M. Cochran
David R. Busch
Anaïs Leproux
Zheng Zhang
Thomas D. O’Sullivan
Albert E. Cerussi
Philip M. Carpenter
Rita S. Mehta
Darren Roblyer
Wei Yang
Keith D. Paulsen
Brian Pogue

Shudong Jiang
Peter A. Kaufman
So Hyun Chung
Mitchell Schnall
Bradley S. Snyder
Nola Hylton
Stefan A. Carp
Steven J. Isakoff
David Mankoff
Bruce J. Tromberg
Arjun G. Yodh
Tissue oxygen saturation predicts response to breast cancer neoadjuvant chemotherapy within 10 days of treatment

Jeffrey M. Cochran, a, * David R. Busch, b Anaïs Leproux, c Zheng Zhang, d Thomas D. O’Sullivan, c Albert E. Cerussi, e Philip M. Carpenter, e Rita S. Mehta, f Darren Roblyer, f Wei Yang, g Keith D. Paulsen, i Brian Pogue, i Shudong Jiang, i Peter A. Kaufman, h So Hyun Chung, a Mitchell Schnall, i Bradley S. Snyder, i Nola Hylton, m Stefan A. Carp, n Steven J. Isakoff, o David Mankoff, o Bruce J. Tromberg, c and Arjun G. Yodha a

1 Introduction

Neoadjuvant chemotherapy (NAC) is a widely used treatment method for breast cancer that permits increased conservation of breast tissue during tumor resection and limits the need for axillary node treatment and surgery. 1 In addition, pathologic complete response (pCR) to NAC, defined as no residual invasive carcinoma, has been correlated with improved survival compared to incomplete response. 2, 3 Unfortunately, this assessment occurs after the completion of NAC. The ability to predict response to NAC at an earlier timepoint during chemotherapy, by contrast, could enable physicians to dynamically optimize the treatment regimen, thereby avoiding unnecessary therapy doses, reducing tissue damage, and improving patient outcomes.

NAC response is typically evaluated with physical exams and radiologic imaging in current clinical practice. Unfortunately, these methods are inadequate predictors of pCR. 4– 6 Magnetic resonance imaging (MRI) provides better correlation with pathology than mammography or ultrasound. 7 Broadly, functional monitoring techniques offer significantly improved correlation with response relative to structural imaging modalities. Magnetic resonance spectroscopy (MRS), 8 contrast-enhanced
MRI, and positron emission tomography (PET) have predictive value with respect to pCR, but MRI, MRS, and PET all have practical constraints, which limit the frequency of monitoring in clinical care. These limitations include cost, the use of contrast agents, and ionizing radiation for PET.

The present contribution investigates the utility of diffuse optical monitoring for prediction of pCR during NAC and adds an analysis to prior reports of a multicenter trial. Briefly, diffuse optical techniques measure functional hemodynamic properties of tissue with nonionizing near-infrared radiation. These optical methods are relatively low cost and can be employed at the bedside. Furthermore, the technology offers a quantitative tool to predict treatment outcome based on longitudinal measurements during therapy. Diffuse optical spectroscopic imaging (DOSI) and tomography (DOT) probe deeply, i.e., several centimeters, into tissue and provide information about tissue optical absorption ($\mu_a$) and reduced optical scattering ($\mu_s$), from which deoxygenated-(HHb) and oxygenated-hemoglobin (HbO$_2$) concentrations, as well as lipid and water (H$_2$O) concentrations can be deduced. The concentrations of HHb and HbO$_2$, in turn, are readily utilized to calculate total hemoglobin concentration (HbT) and tissue oxygen saturation (StO$_2$). These parameters have been shown to discriminate malignant from healthy tissue in the breast and several studies have employed DOSI techniques to explore functional changes in malignancies during NAC and have correlated these changes with response to therapy.

We recently reported the first results of ACRIN-6691, an American College of Radiology Imaging Network (ACRIN) multicenter clinical trial of patients monitored longitudinally by DOSI throughout their NAC regimen. The primary aim of ACRIN-6691 was to evaluate whether a change in a particular DOSI endpoint, the tissue optical index (TOI), could be used to predict a clinical endpoint, pCR, by the midpoint of NAC, ~2 to 3 months after the first infusion. The TOI combines tissue deoxyhemoglobin concentration (HHb), water, and lipid into a single index (see Sec. 2). In that initial study, we reported significant reductions in tumor to normal (T/N) TOI ratios for pCR subjects. A 40% or greater change in this parameter at midpoint, combined with baseline tumor StO$_2$ greater than median values (77%) was shown to be a promising predictor of pCR ($AUC = 0.83; 95\% CI: 0.63$ to $1.1$).

In this study, we explore the ACRIN-6691 secondary aim of predicting pCR much earlier in the 3- to 6-month NAC cycle by examining DOSI response parameters within 10 days of therapy initiation. To address this goal, we develop and retrospectively apply z-score normalization and a logistic regression algorithm to correlate DOSI-measured parameters of malignant breast lesions to subjects' posttherapy pathologic response status. Our hypothesis is that identification and optimization of this z-score DOSI index could predict pCR to NAC at an early timepoint in the course of therapy, providing significant potential for clinical utility.

### 2 Materials and Methods

#### 2.1 Trial Design and Subjects

Data for this study were collected during the ACRIN 6691 multisite trial using a DOSI instrument developed at the University of California, Irvine. Subjects provided written informed consent, and the HIPAA-compliant protocol and

<table>
<thead>
<tr>
<th>Molecular subtype, n (%)</th>
<th>pCR (N = 15)</th>
<th>Non-pCR (N = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Her2 positive</td>
<td>3 (20)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Luminal B</td>
<td>6 (40)</td>
<td>11 (61)</td>
</tr>
<tr>
<td>Luminal A</td>
<td>0 (0)</td>
<td>3 (17)</td>
</tr>
<tr>
<td>Triple negative</td>
<td>4 (27)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (13)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
informed consent were approved by the American College of Radiology Institutional Review Board, the NCI Cancer Therapy Evaluation Program, and each site’s Institutional Review Board. All 60 enrolled subjects were females between the ages of 28 and 67 with biopsy-confirmed invasive ductal carcinomas and/or invasive lobular carcinomas of at least 2 cm in length along the greatest dimension. For each subject, the chemotherapy regimen was determined by the subject’s physician. Chemotherapy type was not controlled in this study, except that regimens were required to include at least one cytotoxic chemotherapeutic agent.

pCR to therapy was defined as no residual invasive primary carcinoma without regard to residual lymph node disease and was determined for each subject from postsurgery pathology reports. Subjects that achieved partial response were not distinguished from nonresponders because of statistical considerations, i.e., sample size, and due to the previously reported correlation between complete response and improved survival. Table 1 contains demographic information, as well as tumor histology and immunohistochemistry for complete and noncomplete responders.

A number of enrolled subjects were excluded from the final data set. Of these, three subjects withdrew from the study. An additional 13 subjects were not included in the imaging analysis because of the following DOSI scan issues: mandatory baseline DOSI was not performed ($n = 1$), baseline DOSI was nonevaluable ($n = 8$), mandatory midtherapy DOSI was not performed ($n = 3$), or too few normal region points were available ($n = 1$). A DOSI scan was considered nonevaluable in case of unrealistic physiological values or incorrect instrument configuration. This decision was made on blinded, deidentified data using instrument calibration and raw data QC reports. A flowchart for this exclusion process can be found in Fig. 6 in the Appendix.

Fig. 1 Timeline and schematic of DOSI measurement during NAC. (a) Each enrolled subject underwent NAC for a period of 4 to 6 months. DOSI measurements were made at four timepoints throughout the course of therapy: (1) baseline—prior to the administration of therapy, (2) early—5 to 10 days after the first dose of therapy, (3) midpoint—the midpoint of the therapy regimen, (4) final—at least 7 days after the final dose of therapy and prior to tumor resection. Note that some subjects are missing data at one or more of the nonbaseline timepoints, and the measurements at the final timepoint were not used due to their limited predictive utility. (b) Top left: DOSI instrument and probe. Right: a grid of points, over a surface area ranging from $7\,\text{cm} \times 7\,\text{cm}$ to $15\,\text{cm} \times 16\,\text{cm}$, were measured on the lesion-bearing breast. This grid was chosen to encompass both the tumor and a portion of the surrounding healthy tissue. The grid of points was marked using a transparency, which was then used to mirror the grid for measurements on the contralateral breast. The transparency was also used to ensure consistent measurement locations across all timepoints. The tumor region was chosen to be all contiguous points with magnitude greater than half of the maximum TOI measurement. The tumor-bearing breast normal region was defined as all points outside the tumor region and areola, excluding a 1-cm margin around both the tumor and areola. The contralateral breast normal region was defined as all measured points, excluding the areola and a 1-cm margin around the areola. Bottom left: a sample DOSI image of the TOI contrast mapped onto a 3-D breast surface (see Sec. 2).
2.2 Optical Imaging Methods

The DOSI instrument used in this study combines multispectral frequency-domain and broadband diffuse optical spectroscopy to measure tissue concentrations of oxygenated hemoglobin (HbO2), deoxygenated hemoglobin (HHb), water (H2O), and lipid, as well as the tissue scattering amplitude (A) and power (b), as defined by a simplified Mie scattering model.37 The combination of these measured parameters permits calculation of total tissue hemoglobin concentration (HbT = HbO2 + HHb), tissue oxygen saturation (StO2 = HbO2/HbT), and the tissue reduced scattering coefficient (\(\mu'_r\)). For a full description of the DOSI method and instrument performance in the multicenter ACRIN-6691 trial, see Ref. 38.

The DOSI instrument measured subjects using a handheld probe (handpiece) placed in contact with the patient’s breast. Four timepoints were acquired throughout the course of each subject in order to guide DOSI handpiece placement using a hand-marked transparency film that was produced for breast (see Fig. 1). These measurement grids were recorded on the contralateral normal tissue was measured on the lesion-bearing breast. A mirrored grid of points was measured on the contralateral normal tissue on the lesion-bearing breast. These margins were not included in the normal region of the lesion-bearing breast, as defined as all points outside the tumor region excluding the areola and 1-cm margins around the areola and tumor region (see Fig. 1). These margins were not included in the normal region to limit signal contamination due to the partial volume effect.

In practice, significant inter- and intrasubject variation in optically measured physiological parameters of the breast can arise,21,30 and these systemic variations can bias the logistic regression. Moreover, the optically measured tissue parameters are not normally distributed (see Fig. 2).

To remedy these issues, we introduce and employ a z-score normalization method to define target variables for prediction of pathologic response. Briefly, the natural logarithm of each data point is first taken. Then, the mean and standard deviation of a given parameter in healthy tissue. In addition to transforming all parameters to be approximately the same magnitude, this method better accounts for the intersubject systemic variations by finding the difference of each parameter from the mean value of the normal (healthy) tissue. It also more fully accounts for intrasubject variation in healthy tissue by normalizing with the healthy tissue standard deviation. A concrete example of this statistical transformation scheme is shown in Fig. 2 for early timepoint tissue oxygen saturation. In this study, normal (healthy) region of tissue are used to transform raw tumor data into z-score data as in

\[
Z_j = \frac{\ln X_j - \langle \ln X_{\text{Norm}} \rangle}{\sigma[\ln X_{\text{Norm}}]}. \tag{1}
\]

Here, \(X_j\) is the unnormalized \(j\)th measured parameter in the tumor region, \(X_{\text{Norm}}\) is the unnormalized \(j\)th measured parameter in the normal (healthy) region of the tumor-bearing breast. \(\langle \ln X_{\text{Norm}} \rangle\) represents the mean over all points in the normal (healthy) region, and \(\sigma[\ln X_{\text{Norm}}]\) represents the standard deviation over all points in the normal (healthy) region. \(Z_j\) is then the tumor region z-score relative to the healthy tissue for the \(j\)th parameter. Each \(Z_j\) parameter was averaged over all spatial points in the tumor region for a given subject and timepoint. As a result, the logistic regression algorithms can utilize a single tumor quantity for each subject, for each timepoint, and for each measured parameter.

Thus every predictor data point used in the regression model is measured in units of standard deviations from the mean of a given parameter in healthy tissue. In addition to transforming all parameters to be approximately the same magnitude, this method better accounts for the intersubject systemic variations by finding the difference of each parameter from the mean value of the normal (healthy) tissue. It also more fully accounts for intrasubject variation in healthy tissue by normalizing with the healthy tissue standard deviation. A concrete example of this statistical transformation scheme is shown in Fig. 2 for early timepoint tissue oxygen saturation. In this study,
we explored z-score normalization schemes that defined the normal region as either the healthy breast (excluding the areola) or all tissue on the lesion breast outside a certain margin of the tumor region (excluding the areola). See Fig. 1 for a graphical representation of these different normalization regions.

In the logistic regression framework, a response parameter for a given model is defined as

$$R_i = \beta_0 + \sum_{j=1}^{N_j} \beta_j \cdot Z_j^i,$$

(2)

Here, $R_i$ is the given model’s log odds of response for the $i$th subject, $\beta_0$ is the intercept term of the fitted weight vector, $\beta_j$ is the weighting term for the $j$th measured parameter used in the model, $Z_j^i$ is the z-score for the $j$th measured parameter of the $i$th subject, and $N_j$ is the number of parameters used in a particular model. The full weight vector $\tilde{\beta}$ is

$$\tilde{\beta} = [\beta_0, \beta_1, \ldots, \beta_{N_j}].$$

(3)

The $\tilde{\beta}$ weight vector is fit using MATLAB’s native logistic regression function, mnrfit. The response parameter $R$ can then be transformed into a probability of response parameter $P_R$ using a logistic function

$$P_R = \frac{1}{1 + e^{-R}}.$$

(4)

The parameter $P_R$ represents the probability that a subject will achieve pCR. It has a range from 0 to 1, and it can readily be used to predict each subject’s status as either a pathologic complete responder, or noncomplete responder, depending on threshold levels.

Because we are working with a small dataset, we employed a leave-one-out validation protocol to test the regression model. Briefly, a series of logistic regression models are created for each parameter set we wish to test, and each of these models leaves one of the subjects out of the dataset (see Fig. 3). The weight vector created by each of these models $\tilde{\beta}_i$ is the weight vector created when the $i$th subject is left-out; it is used to produce a probability of response prediction for the $i$th subject $P_{R_i}$, which is independent of the $\tilde{\beta}_i$ model. This well-known
approach provides the most robust and least biased validation given our sample size, which precludes the use of a significantly large independent test set.4 For completeness, we compared the leave-one-out protocol to other methods. For example, we also tested k-fold cross validation with k = 3, 5, and 10; these schemes produced \( \hat{\beta} \) vectors and AUC values that were similar to those of the leave-one-out protocol.

The quality of the resultant models was empirically determined using DeLong’s method for the area under the curve (AUC) and 95% confidence interval of a receiver-operating-characteristic (ROC) analysis graph for the \( P_R \) parameter.42 The ROC analysis is performed using each of the individual leave-one-out models, and the reported weight vector \( \hat{\beta} \) will be the median \( \hat{\beta} \) from the series of models created for each parameter set, with the interquartile range (IQR) of these models reported as uncertainty. We also calculated a single-logistic regression model run across the entire dataset; it produced very similar \( \hat{\beta} \) vectors to the median \( \hat{\beta} \) vector approach described above.

Models based on the z-scores of each single measured parameter (HbB, HbO2, lipid, H2O, HbT, S2O2, and TOI) at the baseline, early, and midpoint timepoints were tested, and the most predictive models were chosen using the AUC value as a criterion. To explore any additional benefit from multivariate models, combinations of two and three measured parameters were also evaluated. Models with more than three parameters were not tested to avoid overfitting. Data from the final timepoint were not used because our focus in this work is on early diagnosis.

Other, more commonly used, normalization methods were also tested to demonstrate the improvement in predictive ability provided by z-score normalization. These comparisons included tumor-to-normal ratio normalization without information about the normal tissue heterogeneity, as well as raw tumor physiological values without normalization, and baseline-normalized values, which represent changes in the measured parameters over the course of the therapy regimen. All statistical analysis was performed using MATLAB R2015a (The Mathworks, Inc., Natick, Massachusetts, USA).30

### 3 Results

The final data set was derived from \( n = 33 \) subjects who had complete data sets at the baseline and midpoint timepoints. For models that used measured parameters from the early timepoint, slightly fewer subjects were used (\( n = 29 \)) due to missing data at this timepoint. All subjects had biopsy-confirmed invasive carcinomas and underwent an NAC regimen determined by their physicians.13

For the logistic regression algorithm, z-score normalization to the healthy tissue on the lesion breast, as opposed to normalization to the contralateral breast, produced more predictive models. Recall that we derive z-score data for multiple data types (HbO2, HbB, HbT, S2O2, H2O, and lipid) at multiple timepoints (baseline, early, and midpoint) (all data available in Table 3 in the Appendix). The single best regression model used only the early timepoint tissue oxygen saturation (eS2O2). The weight vector for this model was \( \hat{\beta} = [\hat{\beta_o} = 0.79 \pm 0.09, \hat{\beta_{eS2O2}} = 2.29 \pm 0.04] \). This finding suggests that, at early timepoints, tumors that are not hypoxic relative to the surrounding normal tissue, or tumors that are only slightly hypoxic and within the normal region’s confidence interval, are more likely to be pathologic complete responders to NAC. By contrast, tumors that were significantly hypoxic relative to the normal tissue were likely to be nonresponders (see Fig. 4 for data summary in traditional units). When ROC analysis was performed, this model produced an AUC = 0.92 with a 95% confidence interval of AUC = 0.82 to 1 (see Fig. 5). Additionally, the small uncertainties of the \( \hat{\beta} \) components, relative to the median, indicate that the fitted \( \hat{\beta} \) did not vary significantly across the leave-one-out validation protocol.

Two- and three-parameter models did not improve upon the single-parameter model AUC. Higher-order models, e.g., four-parameter, were not considered in order to avoid overfitting of the data.

Notably, in addition to the early timepoint oxygen saturation, a two-parameter model using only baseline data provided an AUC = 0.83 with a 95% confidence interval of AUC = 0.70 to 0.97, thus enabling an even earlier prediction of a subject’s pCR status, albeit with lower accuracy than the early timepoint oxygen saturation. This two-parameter model incorporated the baseline oxygen saturation (bS2O2) and water concentration (bH2O), and the median weight vector was \( \hat{\beta} = [\hat{\beta_o} = 0.14 \pm 0.09, \hat{\beta_{bS2O2}} = 1.69 \pm 0.06, \hat{\beta_{bH2O}} = 0.65 \pm 0.03] \). Again, the uncertainties in the \( \hat{\beta} \) components for H2O and S2O2 are small, signifying a consistent fitted model across

![Fig. 4 Tumor and normal S2O2 versus probability of response. This graph shows the probability of response predicted by the regression model using only early timepoint S2O2 (see Fig. 5). Contour lines of constant probability are also included. The probability of response (shading) is plotted versus the difference between the absolute tumor region percent oxygen saturation and the absolute normal region percent oxygen saturation (horizontal axis), and the size of the confidence interval for the absolute normal region oxygen saturation, corresponding to one standard deviation in the log-transformed data (vertical axis). Note that the oxygen saturation in this figure is not log-transformed or z-score normalized. Each cross represents a subject that was a pathologic complete responder while each circle indicates a nonresponding subject. All subjects that had tumor regions with absolute oxygen saturations that were higher than their normal regions achieved pCR. Subjects whose tumor regions were only slightly hypoxic relative to their normal regions were more likely to achieve pCR if the subjects’ normal regions had larger confidence intervals. These observations indicate that a subject is likely to be a pathologic complete responder if the oxygen saturation of the tumor region is either higher than that of the normal region or well within the normal region’s confidence interval. A subject whose tumor was significantly hypoxic relative to the normal tissue was likely to be a nonresponder.](https://biomedicaloptics.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics on 09 Jun 2019 Terms of Use: https://biomedicaloptics.spiedigitallibrary.org/terms-of-use)
the leave-one-out validation procedure. The fact that $b_{bStO2} > b_{bH2O}$ indicates that the oxygen saturation is a more significant predictor of pCR than water concentration at the baseline timepoint. As with the early timepoint model, subjects with hypoxic tumors were less likely to achieve pCR.

For comparison, additional models were produced that (1) used the contralateral breast for z-score normalization, (2) used tumor-to-normal ratio normalization, i.e., with no information about the standard deviation of the normal region, and (3) used no normalization. With contralateral z-score normalization, instead of z-score normalization to the healthy tissue on the tumor-bearing breast, the aforementioned early ($eStO2$) and baseline ($bStO2$ and $bH2O$) models had AUC values of 0.67 and 0.64, respectively. With simple tumor-to-normal ratio normalization, the same two-parameter sets produced AUC values of 0.80 and 0.67, and when completely unnormalized data were used, the $eStO2$ model produced an AUC = 0.75 while the $bStO2$ and $bH2O$ model produced an AUC = 0.68. Thus, for these parameter sets, z-score normalization to the healthy tissue in the tumor-bearing breast provided the best results.

4 Discussion

By application of a logistic regression model using z-score normalized DOSI measurements, we derived a robust predictor of response (AUC = 0.92; 95% CI: 0.82 to 1) within the first 10 days after a subject’s initial chemotherapy dose. Using an optimally chosen cutoff value of $P_{CR} = 0.50$, which maximizes the sum of the sensitivity and specificity, this model provided an overall classification accuracy of 86% (25 of 29 subjects), including a positive predictive value of 79% for subjects predicted to achieve pCR (11 of 14), and a negative predictive value of 93% for subjects predicted to not achieve pCR (14 of 15). Prediction of response at this therapy timepoint was a secondary aim of the ACRIN 6691 protocol\textsuperscript{29} and could, with further validation, enable clinicians to modify the patient’s therapeutic plan after a single dose. This ability holds potential to improve patient outcomes and prevent unnecessary side effects from ineffective treatments.

The best model indicated that low $S\text{O}_2$ at the early timepoint relative to the surrounding normal tissue was predictive of nonresponse to chemotherapy. This observation suggests that tumors that are well-perfused in the early stages of treatment, and therefore are not hypoxic relative to healthy tissue, may receive chemotherapy more efficiently. Such tumors are also typically more responsive to therapy than hypoxic tumors, which often exhibit resistance to treatment.\textsuperscript{45,44} Additionally, the lack of hypoxia in complete responders could indicate a decreased oxygen demand due to suppression of tumor metabolism, a condition previously shown to be correlated with response to therapy.\textsuperscript{45}

Additionally, the two-parameter model using only the baseline ($S\text{O}_2$) and water concentration (AUC = 0.83; 95% CI: 0.70 to 0.97) also indicated that higher $S\text{O}_2$ is correlated with pCR. Though the AUC value is lower for this model compared to the early timepoint $S\text{O}_2$ model, prediction of response prior to the initiation of therapy offers additional clinical utility. These models are also consistent with previous studies, which have observed correlation between pCR and optically measured tissue oxygen saturation prior to the start of therapy\textsuperscript{31} and after the first dose.\textsuperscript{29}

Previous diffuse optical studies of response to breast cancer NAC have correlated temporal changes in measured physiological parameters with response to treatment.\textsuperscript{24,30,32} We compared our technique to this approach in the current study. However, even the most predictive of the models derived in this analysis that used the change in DOSI physiological parameters between the baseline and early timepoints only produced an AUC = 0.63. The temporal change models of $S\text{O}_2$, in particular, could be limited by the large intersubject dispersion of the baseline oxygen saturation; this large dispersion prevents the change in $S\text{O}_2$ from the baseline to early timepoint from accurately reflecting the oxygenation state of the tumor relative to the normal region. By contrast, the model we have presented in this contribution does not depend on the baseline $S\text{O}_2$ and, as such, is not affected by intersubject baseline variation.

Z-score normalization was implemented to place all parameters on the same magnitude scale, which mitigates systemic physiological differences among the subject population and accounts for the systemic effects of chemotherapy. For comparison, we also investigated models that used fully unnormalized data and tumor-to-normal ratio normalization. However, since neither model incorporates healthy tissue standard deviation, neither model accounts for the heterogeneity of normal tissue. With tumor-to-normal ratio normalization, a one-parameter model with early timepoint $S\text{O}_2$ produced an AUC = 0.80, and the two-parameter model with baseline timepoint $S\text{O}_2$ and $H_2O$ produced an AUC = 0.67. The AUC values for the same models but with no normalization were even lower (AUC = 0.75 and AUC = 0.67, respectively). Thus z-score normalization improves the predictive power of the tissue oxygen saturation logistic regression models.
For completeness, several other models were explored that did not incorporate the baseline or early $S_{O_2}$. Some of these produced predictions of response to therapy that were significant (AUC ≈ 0.75 to 0.80). However, in addition to having lower AUC values, these other models relied on data from the midpoint timepoint, which increases the time-to-prediction of response by ~2 months. Furthermore, the early timepoint measurements typically occur before significant anatomic changes in tumor size arise. This feature enables the DOSI measurement to sample known tumor tissue more easily; at the midpoint of therapy, by contrast, the tumor size has decreased and signal contamination between the malignant and healthy tissue can occur and limit the ability of DOSI to determine tumor physiological parameters accurately. Note also that the physiological predictions of these other models were consistent with our two primary prediction models.

Another interesting and potentially important finding of the present work is that the best models used $z$-score normalization to the normal tissue on the lesion breast rather than the contralateral breast. If, instead, the contralateral breast was used, our one-parameter early $S_{O_2}$ model produced an AUC = 0.67, and the two-parameter model with baseline $S_{O_2}$ and $H_2O$ produced an AUC = 0.64. The comparatively better quality of the tumor breast $z$-score normalized models suggests that measurement of the contralateral breast is less important for early prediction of response to therapy than previously thought. If this is true, then the paradigm could eliminate the need for contralateral measurement and reduce imaging time.

The results we have presented provide evidence for early prediction of response with AUC values that are comparable to other modalities, such as MRI, FDG-PET, and biomarker analysis. Some of these studies produced predictions prior to or within the first 10 days of treatment initiation, whereas other approaches relied on imaging that occurred either after 6 weeks of NAC at the midpoint of therapy, or after the completion of NAC. The potential advantage of the logistic regression DOSI model is premised on its unique combination of accurate prediction at an early timepoint in therapy and its portability, low cost, and lack of ionizing radiation.

The primary limitations of this study are the relatively small number of subjects and the highly variable chemotherapy regimens across the subject population. Additionally, the initial study had a fairly high dropout rate introducing a potential bias into the statistical analysis. The dropout rate is likely to be artificially elevated in this study due to the difficulties inherent in translating an experimental imaging technique into a multisite setting for the first time. We do not anticipate that these issues will affect the DOSI technique moving forward. Finally, although the initial ACRIN 6691 trial was a prospective study, this $z$-score parameter imaging metric was retrospectively optimized using a standard leave-one-out protocol for multiple potential models. The leave-one-out technique limits overfitting and enhances the generalizability of the prediction metric; it has been extensively used by the cancer community. Of course, a fully prospective validation of this single prediction model, as opposed to the series of models tested here, will be necessary prior to clinical adoption.

Per the first limitation noted above, application of this model to a prospective study with a larger subject population is a natural course of action. Importantly, because the DOSI instrumentation has been shown to provide consistent performance over time, across multiple instruments, and across multiple measurement sites, we anticipate that the weight vector derived for the early timepoint $S_{O_2}$ (see Fig. 5) could be used with $z$-score normalized measurements in future DOSI studies to calculate a probability of response, i.e., without creating a logistic regression model for each population. In this case, the future study would serve as a direct, independent test set for the results obtained by our current model. Additionally, a logistic regression could also be run on this larger data set to derive an improved prediction model based on a larger training set. If a future study was performed on a significantly different patient population, e.g., patients with tumors in nonbreast tissue, then deriving a weight vector via logistic regression would likely be beneficial.

In addition to providing evidence to further corroborate the results of this pilot investigation, the larger subject population may enable stratification of the subject population by tumor subtype and/or chemotherapy regimen. Our current results are reported for a diverse patient population, various tumor molecular subtypes, and an assortment of chemotherapy regimens (see Table 1). Tumor subtypes may have different levels of tissue oxygen saturation and may respond to chemotherapy differently. The physiological mechanisms of chemotherapy regimens also vary. Thus, especially for parameters at the early timepoint, response prediction might be improved by creating individual models for different classes of chemotherapy and/or different tumor subtypes. Also, independent hypoxia biomarkers, such as carbonic anhydrase IX, and measurements of vascular density, such as CD31 staining or DCE-MRI, can be collected at similar timepoints and may enable better understanding of the mechanisms responsible for correlations between tissue oxygen saturation and response. Exploration of these questions should be possible in a larger study.

5 Conclusion

Logistic regression modeling of $z$-score normalized physiological parameters measured by DOSI was presented and found to predict pCR to NAC. The best model successfully predicted pCR (AUC = 0.92; 95% CI: 0.82 to 1) using tumor and normal tissue oxygen saturation measured within the first 10 days after the initial dose of therapy based on data from the ACRIN 6691 clinical trial. This model suggests that if tumors are hypoxic relative to the surrounding normal tissue, then they are less likely to achieve pCR. These early predictions of therapeutic efficacy are based on quantitative DOSI measurements of tumor (and normal) tissue functional parameters, rather than changes in tumor size, and the $z$-score normalization of the tumor physiological data yielded improved prediction models compared to tumor-to-normal ratio or unnormalized data. Prospective validation is still needed to confirm these promising results. With this validation, DOSI and logistic regression methods could be used early in NAC to optimize treatment outcomes for individual patients.

Appendix

This appendix contains details about the subject exclusion criteria (Fig. 6) and a more complete accounting of measured $z$-score parameters for responders and nonresponders at all timepoints (Tables 2 and 3).
Fig. 6 Subject exclusion chart. Of the 60 subjects accrued for this study, \( n = 3 \) withdrew consent, \( n = 1 \) did not have central pathology data, and \( n = 10 \) were excluded for lack of normal tissue measurement. The other \( n = 13 \) subjects were excluded due to lack of baseline DOSI measurement \( (n = 1) \), baseline DOSI measurements that were not evaluable \( (n = 8) \), lack of midpoint DOSI measurement \( (n = 3) \), or too few normal region points were available \( (n = 1) \). This subject population is identical to the population used in the initial ACRIN 6691 study\(^ {13} \) except that one fewer subject was used. This additional excluded subject did have a normal tissue measurement but not a sufficient number of spatial points in the normal region to perform the necessary standard deviation calculation [see Fig. 2 and Eq. (1)].

Table 2 Definitions of measured DOSI parameters and their methods of calculation. HHb, HbO\(_2\), lipid, and H\(_2\)O concentration are all fit directly using the measured intensities throughout the wavelength range. Hb\(_T\), S\(_{O_2}\), and TOI are all derived from the fit parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Calculation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHb</td>
<td>Deoxy-hemoglobin concentration</td>
<td>Multispectral fit of absorption</td>
</tr>
<tr>
<td>HbO(_2)</td>
<td>Oxy-hemoglobin concentration</td>
<td>Multispectral fit of absorption</td>
</tr>
<tr>
<td>Lipid</td>
<td>Lipid concentration</td>
<td>Multispectral fit of absorption</td>
</tr>
<tr>
<td>H(_2)O</td>
<td>Water concentration</td>
<td>Multispectral fit of absorption</td>
</tr>
<tr>
<td>Hb(_T)</td>
<td>Total hemoglobin concentration</td>
<td>( Hb_T = HHb + HbO_2 )</td>
</tr>
<tr>
<td>S(_{O_2})</td>
<td>Tissue oxygen saturation</td>
<td>( S_{O_2} = HbO_2/Hb_T )</td>
</tr>
<tr>
<td>TOI</td>
<td>Tissue optical index</td>
<td>( TOI = (HHb - H_2O)/lipid )</td>
</tr>
</tbody>
</table>

Disclosures

B. J. Tromberg reports intellectual property related to DOSI breast imaging assigned to UC Regents and licensed by Infit, Inc., D. R. Busch and A. G. Yodh have two pending patent applications and A. G. Yodh has two other patents relevant to this work (United States patents 8,082,015 and 6,076,010).

Some of these patents have been licensed to private companies, but neither Dr. Yodh and Dr. Tromberg nor any of the other authors have a financial interest in common with these entities. The other authors have indicated no financial conflicts of interest.

Acknowledgments

Funding for this study was provided through grants from the American College of Radiology Imaging Network, which receives funding from the National Cancer Institute (Nos. U01-CA079778 and U01-CA080098); the National Institutes of Health (Nos. P41-EB015893, R01-NS060653, R01-EB002109, R01-CA142989, P41-EB015890, U54-CA136400, T32-HL007915, R01-NS072338, and R01-NS082309-01A1); the Chao Family Comprehensive Cancer Center (No. P30-CA62203); the Thrasher Research Foundation; the Arnold and Mabel Beckman Foundation; and the June and Steve Wolfson Family Foundation. The diffuse optical spectroscopic imaging instrumentation used in this study was constructed in a university laboratory using federal grant support (NIH). The authors thank the entire ACRIN staff for their generous support in completing this study, including Donna Hartfeil, Sharon Mallet, and Dunstan Horng; UCI coordinators Montana Compton, Erin Sullivan, and Jennifer Ehren; UCI engineers Amanda Durkin and Brian Hill; UPenn coordinators Ellen Foster, Madeline Winters, and Sarah Grundy, Dr. Angela DeMichele and Dr. Julia Tchou, clinical coordinators at all sites, all clinicians who contributed to subject recruitment, and the patients who generously volunteered their time for this study.
## Table 3  Median z-score values with IQRs for each measured parameter and timepoint, separated by pCR status.

<table>
<thead>
<tr>
<th></th>
<th>Baseline median z-score (IQR)</th>
<th>Early median z-score (IQR)</th>
<th>Midpoint median z-score (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbO₂</td>
<td>1.5 (1.1, 1.8)</td>
<td>1.0 (0.4, 1.9)</td>
<td>1.4 (0.8, 1.7)</td>
</tr>
<tr>
<td>HbH</td>
<td>2.2 (1.5, 2.7)</td>
<td>1.9 (1.5, 4.0)</td>
<td>1.5 (1.0, 1.9)</td>
</tr>
<tr>
<td>HbT</td>
<td>1.8 (1.5, 2.4)</td>
<td>1.4 (0.8, 2.2)</td>
<td>1.4 (0.8, 1.8)</td>
</tr>
<tr>
<td>S𝑂₂</td>
<td>−0.2 (−0.8, 0.1)</td>
<td>−1.8 (−2.3, −0.8)</td>
<td>0.4 (−0.3, 0.5)</td>
</tr>
<tr>
<td>H₂O</td>
<td>2.1 (1.5, 2.8)</td>
<td>1.8 (1.2, 3.1)</td>
<td>1.5 (1.3, 2.3)</td>
</tr>
<tr>
<td>Lipid</td>
<td>−1.8 (−2.9, −1.4)</td>
<td>−1.4 (−2.5, −0.1)</td>
<td>−1.0 (−1.8, −0.6)</td>
</tr>
<tr>
<td>TOI</td>
<td>3.0 (1.7, 3.8)</td>
<td>2.4 (1.6, 4.1)</td>
<td>1.7 (1.2, 2.4)</td>
</tr>
</tbody>
</table>

## References

Jeffrey M. Cochran is a postdoctoral fellow in the Department of Physics and Astronomy, University of Pennsylvania. He received his BS degree in physics from Denison University and his PhD in physics from the University of Pennsylvania. His research is focused on developing noninvasive imaging biomarkers of disease using diffuse optical techniques.

David R. Busch is an assistant professor in the Department of Anesthesiology and Pain Management and Department of Neurology and Neurotherapeutics at the University of Texas Southwestern Medical Center. His work focuses on developing noninvasive and minimally invasive optical tools to monitor deep tissues in health and disease, as well as throughout therapy.

Thomas D. O’Sullivan is an assistant professor in the Department of Electrical Engineering, University of Notre Dame. He received a PhD in electrical engineering from Stanford University in 2011 and then was a US Department of Defense Breast Cancer Research Program postdoctoral fellow at the University of California, Irvine Beckman Laser Institute. While at BLI, he also served as a director of the Diffuse Optical Spectroscopy and Imaging Laboratory.

Rita S. Mehta is a health science clinical professor in the Department of Hematology/Oncology, University of California, Irvine. She led a phase-III trial of anastrozole versus anastrozole/fulvestrant in advanced breast cancer (S0226, NEJM-2012) and served as an expert member for the development of ASCO guidelines for advanced breast cancer. She completed four phase-II investigator-initiated trials including the use of trastuzumab in HER-2-positive and dose-dense therapy in triple-negative and inflammatory breast cancer.

Darren Roblyer is an assistant professor of biomedical engineering at Boston University. He received his BS degree in biomedical engineering from Johns Hopkins University in 2004 and received his PhD in bioengineering from Rice University in 2009. He did his postdoctoral work at the Beckman Laser Institute, University of California, Irvine. He is a senior member of SPIE.

Arjun G. Yodh is the James M. Skinner professor of science in the Department of Physics and Astronomy, University of Pennsylvania. He directs the Laboratory for Research on the Structure of Matter, an interdisciplinary materials institute that hosts the Penn MRSEC. His biomedical research is oriented toward diffuse optical imaging and monitoring, aiming to demonstrate the potential of these tools for functional imaging/monitoring in brain, breast, and muscle.

Biographies for the other authors are not available