Improving the accuracy of DCE-MRI-based prediction of bevacizumab- and FOLFOX6-induced CRC liver metastasis shrinkage

Saada A M Abujarada[1,2], James P B O'Connor[1,3], and Chris J Rose[1,2]

[1]*University of Manchester Biomedical Imaging Institute, Manchester, Greater Manchester, United Kingdom,* [2]*University of Manchester Academic Health Science Centre, Manchester, Greater Manchester, United Kingdom,* [3]*Department of Radiology, Christie Hospital NHS Trust, Manchester, Greater Manchester, United Kingdom*

INTRODUCTION There is emerging evidence that biomarkers derived using dynamic contrast-enhanced (DCE-) MRI and measured prior to treatment commencement are prognostic—and may be predictive—of tumor response to anti-VEGF therapy[1,2]. Because anti-VEGF and vascular disrupting agents are relatively expensive, and patient response can be variable, an ability to stratify patients according to expected response to a given treatment could allow healthcare providers to both improve patient outcomes and control costs[3]. Typically, regression analyses are used to evaluate any correlation between imaging biomarkers and outcome measurements, and to parameterize the fitted model. The most common approach is to attempt to fit a linear function—i.e., a line in the case of a single independent variable, and a hyperplane in the case of multiple independent variables—to the data. Stepwise procedures are often used to find a minimal subset of the independent variables that are both statistically significant within the model, and which maximize the total variance in the dependent variable explained by the model. However, while linear regression is well understood, there is no reason to believe that a linear function is the best way to model such data. One approach is to visually inspect the data and transform the independent variables (e.g., using logarithms) to improve linearity. Another problem is that imaging biomarkers vary in their reliability (i.e., scan-rescan repeatability)—e.g., enhancing fraction may be more reliable than the volume transfer coefficient, K^{trans}. Regression analyses should therefore take the reliability of each independent variable into account. If multiple pre-treatment scans are performed, the reliability of each variable can be estimated and used in an errors-in-variables regression. However, in this approach the tasks of assessing reliability and regression are distinct; it would be preferable if both were done simultaneously within a single conceptual framework. In this work we explore an alternative to the above approaches, in which multiple simple nonlinear functions are fitted to predict tumor shrinkage within an ensemble learning framework that explicitly models the reliability of the independent variables. We apply this to the data used in Refs. 1 & 2, and demonstrate that this nonlinear modeling approach is able to predict tumor shrinkage with greater accuracy compared to linear modeling.

METHOD A "regression stump" is a nonlinear function, f, that takes one of two values on the basis of its argument. It can be written as

$$f(x_i) = \begin{cases} c_1 & x_i < s_i \\ c_2 & \text{otherwise} \end{cases}, \quad (1)$$

where x_i is a value of the i^{th} independent variable, c_1 and c_2 are the two values the function can return, and s_i is the value of the "split point" for the i^{th} independent variable. The terms "stump" and "split point" are used because regression stumps are minimal versions of tree structures; an example is shown in figure 1, with the values c_1 and c_2 shown in the terminal nodes. Given a data set, algorithms exist to determine the optimal values of i, c_1, c_2, and s_i that minimize the error of the stump's predictions[4]. Used alone, stumps are not very accurate predictors. However, they can be used as "weak learners" in ensemble learning algorithms that combine many poor predictors into a single good one. We use bootstrap aggregation[4] ("bagging"): multiple bootstrap samples, each with the same number of observations as in the original data set, are drawn from that data set, and a stump is trained on each one. The final predictor's output is the mean of each stump's. To model biomarker reliability, each bootstrap sample is drawn in such a way that the value chosen for each of the multiple imaging biomarkers (independent variables), for each tumor, may come from any of the pre-treatment scans. For a given individual, values chosen from a reliable biomarker will have smaller variance between bootstrap samples, than values chosen from an unreliable biomarker. This approach should penalize biomarkers that are both unreliable and not predictive. Missing data (e.g., due to a patient failing to attend) are imputed within each bootstrap sample.

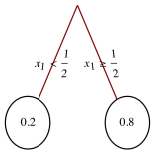

Fig. 1 *A single stump.*

Ten patients with 26 analyzable liver metastases from histologically-confirmed colorectal cancer received single agent bevacizumab (10mg/kg) followed every two weeks by combined bevacizumab and FOLFOX6 (oxaliplatin/5FU/leucovorin) for 5 cycles (10 weeks)[5]. High-resolution T_1- and T_2-weighted imaging and DCE-MRI were performed at two pre-treatment sessions on a 1.5T Philips Intera system (allowing tumor location and volume to be determined, and biomarker reliability to be estimated). DCE-MRI time series were modeled using the extended Kety model[6] (using in-house software), providing voxel-wise estimates of K^{trans}, v_p and v_e. All DCE-MRI data underwent a thorough quality assurance procedure. X-ray CT was performed at the end of cycle 5 (EC5) to measure clinical response in terms of change in tumor volume. All scanning was performed according to local research ethics committee approval; written informed consent was obtained from all patients.

Pre-treatment biomarkers of tumor volume, microvascular function (median K^{trans}, mean v_p, median v_e), enhancement (enhancing fraction, EF), heterogeneity in parameter distributions (standard deviation of K^{trans}, v_p, v_e) and spatial heterogeneity (box dimension of the map of enhancing voxels, d_0; and information and correlation dimensions[7] of the K^{trans}, v_p and v_e parameter maps) were computed for each of the two pre-treatment visits. In one patient (3 tumours), data were available only for one pre-treatment scan. Tumor response was quantified as percentage of tumor volume remaining at EC5 relative to baseline. Leave-one-out analysis was used to estimate how well our regression method would perform on unseen data; 1000 bootstrap samples were used each time. Prediction error was quantified in terms of absolute difference between the actual and predicted values. Our regression method was implemented using Java (Oracle Corp., Redwood City, CA). Results were analyzed using Excel (Microsoft Corp., Redmond, WA) and Mathematica (Wolfram Research Inc., Champaign, IL).

RESULTS Using bagged regression stumps, 50% of tumor shrinkage predictions from the pre-treatment DCE-MRI parameterizations have error ≤12% and 80% of predictions have error ≤20%. In comparison, when using linear errors-in-variables regression, 50% of predictions have error ≤12% and 80% of predictions have error ≤31%. While nonlinear modeling does not improve the median error, fewer large errors are made. However, the Bland-Altman plot (Fig. 2) shows a clear relationship between the magnitude and direction of prediction error and remaining tumor volume. The variables most frequently selected for use in the stumps (expressed as a % of the total number of bootstrap samples) were: information or correlation dimensions for v_p (42%), EF (40%), median K^{trans} (8%) and d_0 (6%), indicating that the spatial arrangement of DCE-MRI information is at least as important as the DCE parameterizations themselves in this experiment.

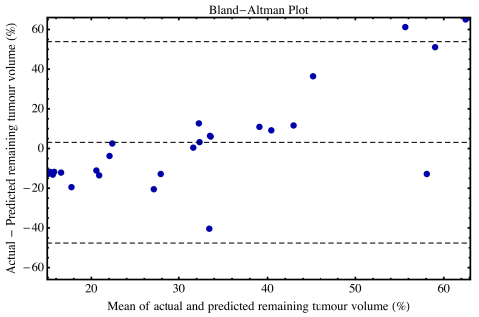

Fig. 2 *Bland-Altman plot showing agreement between actual and predicted values of tumor shrinkage.*

CONCLUSIONS We have developed and evaluated a nonlinear regression method for modeling and predicting tumor response from MR imaging biomarkers that simultaneously accounts for biomarker reliability and missing data. We have demonstrated that the method is more accurate than a simpler linear modeling method and in particular that it makes fewer large prediction errors.

REFERENCES **1** Rose 2011, Proc. ISMRM #338. **2** O'Connor 2011, Brit J Cancer. **3** Meyer 2009, Transl Oncol, 2:198–210. **4** Hastie 2003, The Elements of Statistical Learning, Springer. **5** O'Connor 2009, Clin Cancer Res, 51:6473–5. **6** Tofts 1997, Mag Reson Imag, 7:91–101. **7** Rose 2009, Magn Reson Med, 62:488–499.