Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2019 March 13.

Published in final edited form as:

Phys Med Biol. 2015 July 07; 60(13): 5211–5224. doi:10.1088/0031-9155/60/13/5211.

Predicting location of recurrence using FDG, FLT, and Cu-ATSM PET in canine sinonasal tumors treated with radiotherapy

Tyler Bradshaw[1], **Rau Fu**[2], **Stephen Bowen**[3], **Jun Zhu**[2], **Lisa Forrest**[4], and **Robert Jeraj**[1,5]

[1]Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin, Madison, WI 53705-2275, USA

[2]Department of Statistic, University of Wisconsin, Madison, WI 53705-2275, USA

[3]Departments of Radiology and Radiation Oncology, University of Washington, Seattle, WA 53705-2275, USA

[4]Department of Surgical Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, WI 53705-2275, USA

[5]Department of Human Oncology, School of Medicine and Public Health, University of Wisconsin, Madison, WI 53705-2275, USA

Abstract

Dose painting relies on the ability of functional imaging to identify resistant tumor subvolumes to be targeted for additional boosting. This work assessed the ability of FDG, FLT, and Cu-ATSM PET imaging to predict the locations of residual FDG PET in canine tumors following radiotherapy. Nineteen canines with spontaneous sinonasal tumors underwent PET/CT imaging with radiotracers FDG, FLT, and Cu-ATSM prior to hypofractionated radiotherapy. Therapy consisted of 10 fractions of 4.2 Gy to the sinonasal cavity with or without an integrated boost of 0.8 Gy to the GTV. Patients had an additional FLT PET/CT scan after fraction 2, a Cu-ATSM PET/CT scan after fraction 3, and follow-up FDG PET/CT scans after radiotherapy. Following image registration, simple and multiple linear and logistic voxel regressions were performed to assess how well pre- and mid-treatment PET imaging predicted post-treatment FDG uptake. R^2 and pseudo R^2 were used to assess the goodness of fits. For simple linear regression models, regression coefficients for all pre- and mid-treatment PET images were significantly positive across the population ($P < 0.05$). However, there was large variability among patients in goodness of fits: R^2 ranged from 0.00 to 0.85, with a median of 0.12. Results for logistic regression models were similar. Multiple linear regression models resulted in better fits (median $R^2 = 0.31$), but there was still large variability between patients in R^2. The R^2 from regression models for different predictor variables were highly correlated across patients ($R \approx 0.8$), indicating tumors that were poorly predicted with one tracer were also poorly predicted by other tracers. In conclusion, the high inter-patient variability in goodness of fits indicates that PET was able to predict locations of residual tumor in some patients, but not others. This suggests not all patients would be good candidates for dose painting based on a single biological target.

Keywords

PET; dose painting; Cu-ATSM; modeling

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

The practice of administering biologically conformal radiation therapy, or dose painting, is based on the principle that highly resistant tumor subvolumes can be identified using biological imaging, and thereby targeted for additional boosting. The success of dose painting is contingent on the ability of biological imaging to reliably and consistently identify resistant tumor subvolumes before (or during) radiation therapy. While several dose painting clinical trials have been conducted, and several more are currently underway, the largest uncertainty in dose painting has yet to be resolved—how to identify tumor subregions that are resistant to radiation.

Unlike conventional imaging biomarkers of resistance in which summary statistics from tumor imaging (e.g. maximum standardized uptake value) are correlated with clinical outcome, imaging targets for dose painting should be spatially correlated with spatial patterns of resistance. As tumor radioresistance cannot be directly measured, location of tumor recurrence is often used as a surrogate measure of tumor resistance (Aerts *et al* 2012, Vogelius *et al* 2013, Shusharina *et al* 2014). Pre- and post-treatment images are coregistered, and the locations of tumor recurrence on the post-treatment image are mapped back to the pre-treatment image. Patterns of recurrence are then compared to pre-treatment imaging patterns. For example, using positron emission tomography (PET) imaging in non-small cell lung cancer tumors, Aerts *et al* found high spatial overlap between post-treatment 2-deoxy-2-[^{18}F]fluoro-D-glucose (FDG) uptake and pre-treatment FDG uptake (Aerts *et al* 2009, Aerts *et al* 2012). In head-and-neck tumors, Soto *et al* found that 8 out of 9 patients' recurrence volumes resided primarily within regions of high pre-treatment FDG uptake (Soto *et al* 2008). Results such as these have encouraged the use of FDG as a target for dose painting, as clinical trials of FDG dose painting have been conducted in head-and-neck and lung tumors (Madani *et al* 2007, Duprez *et al* 2011, van Elmpt *et al* 2012). On the other hand, FDG uptake has also been shown to be influenced by other biological properties not directly related to resistance, such as inflammation and perfusion (Bos *et al* 2002), which may confound its utility as a reliable dose painting target.

Besides FDG PET, other potential biological targets for dose painting exist. Hypoxia is a known cause of radioresistance (Nordsmark and Overgaard 2000), making PET imaging of hypoxia a promising method for guiding dose distributions (Chao *et al* 2001). Uptake levels of copper(II)-diacetyl-bis(N4-methylthiosemicarbazone) (Cu-ATSM), a PET radiotracer for hypoxia, have been shown to be prognostic in patients undergoing radiation therapy with cervical (Dehdashti *et al* 2003a, Dehdashti *et al* 2008), head-and-neck (Minagawa *et al* 2011, Sato *et al* 2014), lung (Dehdashti *et al* 2003b), and rectal (Dietz *et al* 2008) cancer. Tumor proliferation maps may also prove useful in guiding dose distributions, especially those acquired during treatment, as accelerated repopulation of irradiated cancer cells has been shown to have deleterious effects on patient outcome (Kim and Tannock 2005, Kishino *et al* 2012).

In addition to PET imaging at a single time point, imaging before and during the course of treatment can provide spatially-resolved biological response information. Patterns of early

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

biological response may correlate with patterns of overall resistance, allowing one to adapt dose distributions according to early changes in PET uptake (Brahme 2003). A comparison of tumor recurrence patterns to patterns of biological response during treatment has yet to be reported in literature.

The goal of this work was to assess how accurately pre-treatment FDG, FLT, and Cu-ATSM PET imaging, as well as mid-treatment FLT and Cu-ATSM PET imaging, predicts the location of residual or recurrent tumor in canine sinonasal tumors. This work builds off of a previous exploratory study, in which we performed voxel regressions of pre-treatment FDG, FLT, and Cu-ATSM uptake values against 3 month post-treatment FDG uptake values in 9 canine patients (Bowen *et al* 2012). In that study, pre-treatment FDG PET was found to perform better than FLT and Cu-ATSM PET at predicting post-treatment FDG PET distributions. The work reported herein builds on these results by increasing the number of patients from 9 to 19, adding mid-treatment FLT and Cu-ATSM PET images as additional predictor variables, and expanding the statistical methods of analysis.

Methods

Patients

The trial included 22 canine veterinary patients referred to the University of Wisconsin Veterinary Medical Teaching Hospital, 19 of whom were evaluable for this study. Three patients were excluded from analysis: two because no recurrent tumor was visible in follow-up PET/CT images, and one because the tumor did not shrink following treatment. The research protocol was approved by the Animal Care and Use Committee of the University of Wisconsin, and all canine owners signed a written informed consent. Patients were diagnosed using CT and biopsied for histopathological evaluation. All dogs had treatment-naive nasal or paranasal sinus tumors with no evidence of distant metastases or intracranial extension. Histopathology results were 10 adenocarcinoma, 7 chondrosarcoma, 1 squamous cell carcinoma, and 1 osteosarcoma.

Imaging and treatment

Patients received intensity-modulated radiation therapy (IMRT) delivered by helical tomotherapy with curative intent. Patients were divided into two treatment groups as part of a larger study investigating dose escalation. The first group received the veterinary standard-of-care prescription at our institution: 10 fractions of 4.2 Gy to the planning target volume (PTV) for a total of 42 Gy. The second group was prescribed 10 fractions of 4.2 Gy to the PTV with an integrated boost of 0.8 Gy per fraction to the gross tumor volume (GTV), for a total of 50 Gy to the GTV.

The imaging and treatment schedules were as follows: Patients had pretreatment FDG, FLT, and [^{61}Cu]Cu-ATSM PET/CT scans on consecutive days beginning 3 d before the onset of IMRT. Patients received a second FLT PET/CT scan after two fractions of IMRT (8.4 or 10 Gy) and a weekend break. Patients received a second Cu-ATSM PET/CT scan following the third fraction of IMRT (12.6 or 15 Gy). One patient missed the midtreatment FLT scan due to equipment failure (resulting in 18 patients with complete data sets). Patients were injected

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with 150–370 MBq of tracer. After injection, patients were kept in a kennel to limit physical activity. PET/CT scans were acquired on a Discovery VCT (GE Healthcare, Waukesha, WI) scanner. FDG and Cu-ATSM PET scans were acquired 60 min and 180 min post-injection, respectively; both were 20 min 3D static acquisitions over a single 15 cm bed position. FLT PET scans were 90 min 3D dynamic acquisitions over a single 15 cm bed position. Patients were anesthetized during imaging and treatment with an initial propofol bolus injection, and then maintained on isoflurane inhalation plus 100% oxygen. Emission data were attenuation, decay, and deadtime corrected and reconstructed using 3D ordered subset expectation maximization (2 iterations, 35 subsets, and 3 mm postfiltering). The image grid was 256 × 256 × 47 with 2.0 × 2.0 × 3.3 mm^3 voxel sizes. Voxel activity measurements were converted to standardized uptake values (SUVs) for analysis. For FLT PET scans, SUVs were calculated by averaging frames between 60–90 min. To achieve reproducible positioning across PET/CT scans and IMRT treatment sessions, patients' maxillae were positioned into custom dental molds that were affixed to the scanner and treatment couches, and patients' bodies were immobilized with vacuum mattresses (Kubicek *et al* 2012).

Following therapy, patients were scheduled for follow-up FDG PET/CT scans at 3 and 6 months, with an additional CT scan at 9 months. Patients without progressive disease within the first 9 months (evaluated according to RECIST criteria) received an additional PET/CT scan whenever they presented with clinical signs suspicious of recurrence (e.g. epistaxis).

Patient CT scans were rigidly registered, and transformations were subsequently applied to their respective PET images. The GTV was segmented by a veterinarian based on the planning CT, and subsequently propagated to the PET images. The ethmoturbinates of the nasal cavity were manually segmented and excluded from analysis due to non-specific uptake of FDG and Cu-ATSM.

Regression analysis

Similar to Bowen *et al* (Bowen *et al* 2012), we used voxel regression methods to assess how well PET uptake patterns predicted the location of tumor recurrence. For each tumor, the predictor variables were pre-treatment FDG (FDG_{pre}), FLT (FLT_{pre}), Cu-ATSM (Cu-$ATSM_{pre}$) and mid-treatment FLT (FLT_{mid}) and Cu-ATSM (Cu-$ATSM_{mid}$) PET voxel SUVs, and the response variable was follow-up FDG PET voxel values (FDG_{post}), which was used as a surrogate for residual/recurrent tumor (for this paper, the terms recurrent tumor and residual tumor will be used interchangeably). Ratio maps of mid-treatment voxel values to pre-treatment voxel values for FLT (R_{FLT}) and Cu-ATSM ($R_{Cu\text{-}ATSM}$) PET images were also used as predictor variables. Two types of regression were considered. In linear regression models, voxel SUVs from patients' post-treatment FDG PET images were used as response variables. In logistic regression models, recurrent tumor was segmented in the post-treatment FDG PET image, and the response variables were converted to binary variables—1 if the voxel was within the segmented volume, and 0 otherwise. The segmentation of post-treatment PET images was performed using an in-house automatic segmentation algorithm that combines gradient, region-growing, and textural feature methods (Galavis 2013). Both simple regression and multiple regression models were created for each patient. For multiple regression models, all 5 PET images were included as predictor variables (ratio variables

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were not included, as the regression model already accounts for absolute differences between pre- and mid-treatment voxel values). As patients had multiple follow-up FDG PET scans, the time point at which the tumor was the smallest was used for the response variable. In most cases ($N = 13$), response was measured at 3 months.

To evaluate how well pre- and mid-treatment PET distributions predicted recurrent tumor, goodness of fits of the regression models were calculated. For linear regression models, the coefficient of determination (R^2) was calculated. For logistic regression models, pseudo R^2 was used. Pseudo R^2 was calculated according to the correlation between the model's predicted response and measured response values:

$$\text{Pseudo}\, R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}, \quad (1)$$

where y_i is the ith response value, $\hat{\pi}_i$ is the model-predicted probability, and $\bar{y}$ is the mean response value. It should also be noted that pseudo R^2 for logistic regression and R^2 from linear regression are not directly comparable, but do have similar interpretations. Significance testing of the regression coefficients was not performed due to the very high spatial correlations between neighboring voxels, which resulted in exaggerated significance values. However, similar to Bowen *et al*, two-tailed t-tests were used to assess if the population of regression coefficients from all patients were significantly different from zero.

Using correlation analysis, we investigated if any clinical measures or imaging biomarkers could predict which patients would have high or low R^2 (i.e. which patients would be good or poor candidates for dose painting). Clinical measures that were tested as predictors of R^2 values included histologic tumor type and dose level. Imaging biomarkers that were tested included tumor volume, maximum and mean SUV from pre- and mid-treatment PET images, relative changes in maximum and mean SUV from pre- to mid-treatment, and relative change in 1D tumor size (i.e. RECIST measurements) from pre- to post-treatment.

Predictive modeling

Given the limitations of linear regression (e.g. assumption of linearity), we also created more advanced prediction models to evaluate how accurately the location of recurrence for a 'new' patient could be predicted given the imaging data of all the other patients. For each patient, 4 prediction models were created/trained based on the other 17 patients' combined data, and then applied to the test patient (i.e. leave-one-out cross validation). One patient was excluded due to a missing FLT_{mid} data. The 4 models included a multiple linear regression (LR) model (as a basis for comparison), a classification and regression tree (CART) model (Breiman *et al* 1984, Loh 2011), a linear mixed-effects regression (LMER) model, and a multiple linear regression model that included neighborhood structures surrounding each voxel as additional covariates to account for autocorrelation between neighboring voxels.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For all predictive models, FDG_{pre}, FLT_{pre}, Cu-$ATSM_{pre}$, FLT_{mid}, and Cu-$ATSM_{mid}$ were used as continuous predictor variables. Patient-level variables of tumor histologic type, tumor volume, and dose level were also used as predictor variables. FDG_{post} was used as the response variable. The multiple linear regression (LR) model was defined according to

$$FDG_i^{post} = \beta_0 + \beta_1 PET_{1i} + \beta_2 PET_{2i} + \ldots + \beta_5 PET_{5i} + \varepsilon_i, \quad (2)$$

where FDG_i^{post} is the post-treatment FDG SUV at voxel i, PET_{mi} is the SUV at voxel i for the mth PET image, β_k are the 6 regression coefficients, including an intercept, and ε_i is the error term. In leave-one-out cross-validation fashion, β_k values were determined by pooling together 17 of the patients' imaging and response data and running regression on the pooled data. Prediction accuracy was then evaluated according to R^2 by fitting the testing data set (i.e. the one patient excluded from the training data set) with the coefficients calculated from the training data set. Each patient was likewise evaluated for prediction accuracy.

For the classification and regression tree (CART) model, the same 5 PET images were used as predictor variables as were used in the LR regression model. Patient-level variables of histologic tumor type, tumor volume, and dose level were also included as patient-level factors. Like the LR model, the model was trained on the pooled image data and tested for each patient in a leave-one-out cross-validation manner.

The linear mixed effects (LME) regression model was similar to the LR model, but with the intercept of different patients defined as a random effect. This assumes the baseline FDG_{post} values for different patients are normally distributed, with the mean of the distribution used as intercept for the pooled data. All other predictors were used as fixed effects. This model and the CART model were implemented in *R* using the packages *lme4* and *rpart*, respectively (R Development Core Team 2014).

The autocorrelation model was defined similar to the LR model, but with added neighborhood structures as additional predictor variables for each PET image:

$$FDG_i^{post} = \beta_0 + \beta_1 PET_{1i} + \beta_2 \sum_{j \epsilon \mathcal{N}_i} PET_{1j} + \ldots + \beta_9 PET_{5i} + \beta_{10} \sum_{j \epsilon \mathcal{N}_i} PET_{5j} + \varepsilon_i, \quad (3)$$

where β_k now include 11 regression coefficients (2 for each PET image), N_i is the neighborhood surrounding voxel i, and PET_{mj} is the SUV of the jth voxel in N_i for PET image m. For this study, we used the first-order neighborhood surrounding voxel i to define N_i, meaning only those voxels sharing a side with voxel i (Fu *et al* 2013).

When pooling the 17 patients' data for training the models, only 249 voxels were randomly sampled and used from each patient so that each tumor in the training set was equally represented in the pooled model, regardless of tumor size (249 voxels was the size of the smallest tumor). Fitted R^2 values were calculated based on the model fit of the 17 patients × 249 voxels = 4 233 voxels. For testing prediction accuracy, prediction R^2 was calculated by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

applying the fitted model to all the voxels of the tested patient. The procedure was repeated 100 times for each patient, and the average of the resulting R^2 values are reported.

In addition to the continuous prediction models, the 4 prediction models were also created for each patient using dichotomized FDG_{post} as the response variable, and using logistic regression instead of linear regression. Results for logistic regression models are reported in terms of pseudo R^2 and sensitivity.

Results

An example patient's PET/CT images and simple linear regression plots are shown in figure 1. The fitted regression coefficients from simple linear and logistic regressions for all patients are shown in figure 2. Overall, linear regression coefficients were significantly different from zero across the population for all pre- and mid-treatment tracers ($P < 0.05$), with FDG_{pre}, FLT_{mid}, and $Cu\text{-}ATSM_{mid}$ being the most significant ($P < 0.005$). R_{FLT} coefficients were significantly less than zero ($P = 0.038$), whereas $R_{Cu\text{-}ATSM}$ coefficients were not significantly different from zero ($P = 0.055$). Likewise, for logistic regression, regression coefficients for all pre- and mid-treatment tracers were significantly different from zero ($P < 0.05$), except for FLT_{pre} ($P = 0.38$). Again, R_{FLT} coefficients were significantly less than zero ($P = 0.028$), whereas $R_{Cu\text{-}ATSM}$ coefficients were not ($P = 0.33$).

Figure 2 also shows R^2 and pseudo R^2 for all patients' linear and logistic simple regression models. There was large variability between patients in goodness of fits, with R^2 and pseudo R^2 values ranging from 0.00 to 0.85, with an overall median of 0.12. For linear regression, the median R^2 for FDG_{pre} was 0.19, while the median R^2 values for FLT_{pre}, $Cu\text{-}ATSM_{pre}$, FLT_{mid}, and $Cu\text{-}ATSM_{mid}$ were all around 0.10. Median R^2 for R_{FLT} and $R_{Cu\text{-}ATSM}$ were less than 0.05. For logistic regression, median pseudo R^2 values for FDG_{pre}, FLT_{pre}, $Cu\text{-}ATSM_{pre}$, FLT_{mid}, and $Cu\text{-}ATSM_{mid}$ varied between 0.06 to 0.10, and R_{FLT} and $R_{Cu\text{-}ATSM}$ had median pseudo R^2 less than 0.03.

Figure 3 shows all patients' R^2 and pseudo R^2 from both linear and logistic multiple regression models, respectively. Multiple regression models had better goodness of fits than simple regression models: the median R^2 for the linear model was 0.31, and median pseudo R^2 for the logistic model was 0.24.

No clinical measures (e.g. dose level and tumor histology) or imaging biomarkers were correlated with patients' respective R^2 ($R < 0.3$). We found that the R^2 of different regression models (for different predictor variables) were correlated across patients ($R > 0.75$). In other words, patients with low R^2 for a single predictor variable had similarly low R^2 for all predictor variables, even for multiple regression models. This is illustrated for all three pre-treatment simple linear regression models in figure 4. Likewise, if a patient had low regression coefficient for a single predictor variables, the other predictor variables were likely to also have low regression coefficients (see figure 4).

Figure 5 shows results from each patient's predictive modeling. For each patient, the fitted R^2 and fitted pseudo R^2 are shown—these indicate how well the models described the *other* 17 patients' data. The prediction R^2 and prediction pseudo R^2 indicate how well the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

prediction models performed when applied to the test patient. In general, CART outperformed the other models in fitting the patients' data, but often had the worst performance in predicting a 'new' test patient. Overall, advanced prediction models did not improve the prediction accuracy when compared to multiple linear regression models.

Discussion

In this study, we used canine sinonasal tumors undergoing radiation therapy as models for evaluating which radiotracer uptake distribution—FDG, FLT, or Cu-ATSM—best correlated with location of tumor recurrence as a surrogate for measuring spatially-resolved resistance to radiation therapy. A characteristic of canine sinonasal tumors that made them especially well-suited for this study was the fact that tumors were enclosed by the bony sinonasal cavity, making image registration very accurate, thus enabling voxel-wise comparisons of PET images across different scans and time points. In our previously published exploratory voxel regression study evaluating 9 of the patient used in this study, pre-treatment FDG uptake distributions were found to be the best predictors of post-treatment FDG uptake distributions (Bowen *et al* 2012). In this study, we made several additions to the methods of analysis, including the addition of mid-treatment FLT and Cu-ATSM PET images as predictor variables, implementing logistic regression with segmented follow-up FDG PET images as binary response variables (removing potential autocorrelation between pre- and post-treatment FDG voxel values), and creating advanced prediction models accounting for non-linearity, patient-level variables, and autocorrelation among neighboring voxels. We also increased the statistical power of the study by increasing the number of evaluable canine patients from 9 to 19.

Overall, we found large variability among patients in how well pre- and mid-treatment PET imaging predicted the locations of tumor recurrence. For simple linear regression models, none of the predictor variables were significantly better than the other variables. FDG_{pre}, however, had the highest median and mean R^2 values of the tested predictor variables, suggesting it might be the most suitable target for dose painting. This is in agreement with the results reported by Bowen *et al* (Bowen *et al* 2012). On the other hand, when FDG_{post} was dichotomized and logistic regression was performed, the R^2 values for FDG_{pre} were nearly identical to those of the other predictor variable. Unsurprisingly, we found that multiple regression models resulted in better goodness of fits (R^2 and pseudo R^2) than simple regression models. However, there was a cluster of 5 patients whose R^2 were at or below 0.20, regardless of regression method. We hypothesized that perhaps smaller tumors or tumors with more pronounced anatomical response were those that had lower R^2 values. This, however, was not the case, as patients' R^2 values were not correlated with any imaging biomarkers or clinical measures that we tested, including dose level, tumor volume, and histologic tumor type. Biomarkers or patient characteristics that could predict which patients would 'behave well' under dose painting would be valuable for selecting patients most likely to benefit from dose painting.

The interpretation and implications of a low or high R^2 for such regression models, or the thresholds at which an R^2 could be considered adequate for dose painting, are not straightforward. In redistribution-based dose painting, total dose to the target volume is not

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

increased but rather redistributed such that more sensitive regions receive less dose and more resistance regions receive more dose. Under such a scenario, an imprecise dose painting target (i.e. low R^2) that is unable to differentiate sensitive and resistant regions could actually result in worse outcomes than standard uniform dose treatments. Bender addressed these issues using a tumor control probability model, and reported that when the voxel-based correlation coefficient between a particular imaging method and the theoretical probability of recurrence risk was greater than 0.45 (corresponding to an R^2 of 0.20), voxel-based dose painting would result in greater tumor control than uniform dose prescriptions (Bender 2012). Clearly, the measure of post-treatment FDG SUV used in our study was not a direct measure of recurrence risk, but if we were to use an R^2 threshold of 0.20 for the simple linear regression models, we would find that roughly half of the 19 patients would benefit from dose painting. For the other half of the population, however, dose painting might actually be worse than uniform dose prescriptions. These results, of course, depend on the tracer used (see figure 2), and regression method (linear vs. logistic versus multivariable), but do suggest that not all patients are good candidates for dose painting. These results highlight the need to identify alternative objective measures that can predict which patients would make good or poor candidates for dose painting based on a given biological imaging target.

Voxel-based predictors based on ratios of PET images from pre- to mid-treatment (R_{FLT} and $R_{\mathrm{Cu\text{-}ATSM}}$) were not good at predicting recurrent tumor location: R^2 and pseudo R^2 values were close to 0 for simple linear and logistic regression models. We originally hypothesized that biological changes occurring early during radiation therapy might correlate with tumor resistance. This hypothesis was supported by recent analysis of clinical outcome for the canine patients included in this study: mid-treatment FLT $\mathrm{SUV}_{\mathrm{max}}$ and relative changes in FLT $\mathrm{SUV}_{\mathrm{max}}$ from pre-treatment to mid-treatment were the most predictive of progression-free survival (Bradshaw *et al* 2015). The caveat, however, was that tumors with large relative decreases in FLT $\mathrm{SUV}_{\mathrm{max}}$ actually had a *worse* prognosis. In this study, we did not find response maps of FLT uptake to be significant spatial predictors of recurrent tumor location. A mathematical problem with ratio maps, however, is that response values can become inflated in regions with initially-low voxel values (i.e. denominators close to zero), so that even small fluctuations in image values due to statistical noise can appear as large response values. This may have confounded the utility of R_{FLT} and $R_{\mathrm{Cu\text{-}ATSM}}$ as predictors. Furthermore, we previously demonstrated in these patients that image patterns for FLT PET, and especially for Cu-ATSM PET, were spatially stable from pre-treatment to mid-treatment (Bradshaw *et al* 2014). Other techniques of quantifying spatially-resolved changes in longitudinal PET images may be more appropriate, such as absolute SUV changes (although these were included in the multiple regression models), or other methods (Kinoshita *et al* 2012).

Advanced predictive models did not provide an improvement over multiple linear regression models at predicting a ‘new’ patient’s post-treatment FDG values. These prediction models were used to account for potentially confounding factors that are not accounted for in linear regression, such as patient categorical variables, non-linearity, and autocorrelation among neighboring voxels. For the models with continuous response data, the fitted R^2 for the pooled data were generally higher than for the multiple linear models shown in figure 3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

However, the better fitting did not result in high prediction accuracy, as median prediction R^2 values for the different models were less than 0.15. CART performed the best at describing the training data (see figure 5), but often having the poorest prediction performance—likely a consequence of over fitting. Overall, prediction R^2 were approximately the same as the regression R^2 shown in figure 3, and contained large interpatient variability.

Several other studies have correlated imaging patterns to recurrent tumor location. The studies by Aerts *et al* and Soto *et al* in lung and head-and-neck tumors, respectively, have already been discussed in the introduction (Soto *et al* 2008, Aerts *et al* 2009, Aerts *et al* 2012). Abramyuk *et al* and Shusharina *et al* performed similar studies in 10 and 19 patients, respectively, who had lung cancer and experienced loco-regional failure following radiation therapy. Both studies found high degrees of spatial overlap between recurrence volumes and pre-treatment FDG uptake (Abramyuk *et al* 2009, Shusharina *et al* 2014). Petit *et al* used logistic voxel regression to relate pre-treatment FDG uptake to post-treatment FDG uptake in 39 NSCLC tumors, and found increased FDG uptake at baseline predicted a greater probability of post-treatment FDG uptake (Petit *et al* 2009). Vogelius *et al* found that recurrent tumor was most likely to originate in the regions of elevated FDG uptake in 39 HN cancer patients undergoing radiation therapy (Vogelius *et al* 2013).

A primary assumption to the methods of this study—which is a common assumption to all of the previously-mentioned spatial analysis studies—is that, following image registration, the tumor cells in a voxel location do not change or move location between pre-treatment imaging and post-treatment imaging. It is unclear how frequently this assumption is valid. Deformable registration methods attempt to account for tumor morphological changes over time, but deformable registration algorithms assume that the tumor's mass is preserved over time, which is not the case for tumors undergoing treatment. For this reason, we used rigid registration. Canine sinonasal tumors benefit from the fact that rigid registration can be very accurate due to the surrounding bony nasal cavity, and (based on our experience) the turbinates inside the nasal cavity help to hold the tumor in its place, restricting tumor movement over time. An additional limitation of the study is the degree to which post-treatment FDG uptake is a valid marker of recurrent or residual tumor, and whether or not it corresponds to resistant tumor. Numerous studies have found that post-treatment FDG uptake does correlate with poor clinical outcome (Inoue *et al* 1995, Levine *et al* 2006, Bollineni *et al* 2012, Machtay *et al* 2013). However, non-specific uptake of FDG, including radiation-induced inflammation, could still be a confounding factor in studies such as this. A further limitation to the study is the uncertainty regarding the specificity of Cu-ATSM uptake as a marker for tumor hypoxia (Hueting *et al* 2014), especially under conditions of anesthesia and oxygen breathing (Kersemans *et al* 2011).

Conclusion

Using multi-tracer PET imaging and voxel regression models, we found that FDG, FLT, and Cu-ATSM PET were able to predict the locations of recurrence in some canine tumors, but they were poor predictors of recurrence locations in other tumors. We also showed that no individual PET tracer significantly outperformed the other PET tracers in how well they

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

predicted patterns of recurrence. No patient characteristics or imaging biomarkers that we tested could predict which patients would be good candidates for dose painting (i.e. have a high prediction R^2). These results highlight the need for methods of pre-selecting patients who would be good candidates for dose painting, and for the development of better molecular imaging probes of resistance.

References

Abramyuk A, Tokalov S, Zophel K, Koch A, Szluha Lazanyi K, Gillham C, Herrmann T and Abolmaali N 2009 Is pre-therapeutical FDG-PET/CT capable to detect high risk tumor subvolumes responsible for local failure in non-small cell lung cancer? Radiother. Oncol 91 399–404 [PubMed: 19168248]

Aerts HJ, Bussink J, Oyen WJ, Van Elmpt W, Folgering AM, Emans D, Velders M, Lambin P and De Ruysscher D 2012 Identification of residual metabolic-active areas within NSCLC tumours using a pre-radiotherapy FDG-PET-CT scan: a prospective validation. Lung Cancer 75 73–6 [PubMed: 21782272]

Aerts HJWL et al. 2009 Identification of residual metabolic-active areas within individual NSCLC tumours using a pre-radiotherapy (18)Fluorodeoxyglucose-PET-CT scan Radiother. Oncol 91 386–92 [PubMed: 19329207]

Bender ET 2012 Using spatial information about recurrence risk for robust optimization of dose-painting prescription functions Med. Phys 39 2713–20 [PubMed: 22559642]

Bollineni VR, Widder J, Pruim J, Langendijk JA and Wiegman EM 2012 Residual ^{8}F-FDG-PET uptake 12 weeks after stereotactic ablative radiotherapy for stage I non-small-cell lung cancer predicts local control Int. J. Radiat. Oncol. Biol. Phys 83 e551–5 [PubMed: 22417800]

Bos R et al. 2002 Biologic correlates of (18)fluorodeoxyglucose uptake in human breast cancer measured by positron emission tomography J. Clin. Oncol 20 379–87 [PubMed: 11786564]

Bowen SR, Chappell RJ, Bentzen SM, Deveau MA, Forrest LJ and Jeraj R 2012 Spatially resolved regression analysis of pre-treatment FDG, FLT and Cu-ATSM PET from post-treatment FDG PET: an exploratory study Radiother. Oncol 105 41–8 [PubMed: 22682748]

Bradshaw TJ, Bowen SR, Deveau MA, Kubicek L, White P, Bentzen SM, Chappell RJ, Forrest LJ and Jeraj R 2015 Molecular imaging biomarkers of resistance to radiation therapy for spontaneous nasal tumors in canines Int. J. Radiat. Oncol. Biol. Phys 91 787–95 [PubMed: 25752393]

Bradshaw TJ, Yip S, Jallow N, Forrest LJ and Jeraj R 2014 Spatiotemporal stability of Cu-ATSM and FLT positron emission tomography distributions during radiation therapy Int. J. Radiat. Oncol. Biol. Phys 89 399–405 [PubMed: 24685446]

Brahme A 2003 Biologically optimized 3D in vivo predictive assay-based radiation therapy using positron emission tomography-computerized tomography imaging Acta Oncol. 42 123–36 [PubMed: 12801131]

Breiman L, Friedman JH, Olshen RA and Stone CJ 1984 Classification and Regression Trees (Monterey, CA: CRC)

Chao KS, Bosch WR, Mutic S, Lewis JS, Dehdashti F, Mintun MA, Dempsey JF, Perez CA, Purdy JA and Welch MJ 2001 A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guided intensity-modulated radiation therapy Int. J. Radiat. Oncol. Biol. Phys 49 1171–82 [PubMed: 11240261]

Dehdashti F, Grigsby PW, Lewis JS, Laforest R, Siegel BA and Welch MJ 2008 Assessing tumor hypoxia in cervical cancer by PET with 60Cu-labeled diacetyl-bis(N4-methylthiosemicarbazone) J. Nucl. Med 49 201–5 [PubMed: 18199612]

Dehdashti F, Grigsby PW, Mintun MA, Lewis JS, Siegel BA and Welch MJ 2003a Assessing tumor hypoxia in cervical cancer by positron emission tomography with 60Cu-ATSM: relationship to therapeutic response—a preliminary report Int. J. Radiat. Oncol. Biol. Phys 55 1233–8 [PubMed: 12654432]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dehdashti F, Mintun MA, Lewis JS, Bradley J, Govindan R, Laforest R, Welch MJ and Siegel BA 2003b In vivo assessment of tumor hypoxia in lung cancer with 60Cu-ATSM Eur. J. Nucl. Med. Mol. Imaging 30 844–50

Dietz DW, Dehdashti F, Grigsby PW, Malyapa RS, Myerson RJ, Picus J, Ritter J, Lewis JS, Welch MJ and Siegel BA 2008 Tumor hypoxia detected by positron emission tomography with 60Cu-ATSM as a predictor of response and survival in patients undergoing Neoadjuvant chemoradiotherapy for rectal carcinoma: a pilot study Dis. Colon Rectum 51 1641–8 [PubMed: 18682881]

Duprez F, De Neve W, De Gersem W, Coghe M and Madani I 2011 Adaptive dose painting by numbers for head-and-neck cancer Int. J. Radiat. Oncol. Biol. Phys 80 1045–55 [PubMed: 20643512]

Fu R, Thurman AL, Chu TJ, Steen-Adams MM and Zhu J 2013 On estimation and selection of autologistic regression models via penalized pseudolikelihood J. Agric. Biol. Environ. Stat 18 429–49

Galavis, PE; Robust tumor segmentation for radiotherapy target definition PhD Thesis University of Wisconsin. 2013.

Hueting R et al. 2014 A comparison of the behavior of (64)Cu-acetate and (64)Cu-ATSM in vitro and in vivo J. Nucl. Med 55 128–34 [PubMed: 24337603]

Inoue T, Kim EE, Komaki R, Wong FC, Bassa P, Wong WH, Yang DJ, Endo K and Podoloff DA 1995 Detecting recurrent or residual lung cancer with FDG-PET J. Nucl. Med 36 788–93 [PubMed: 7738649]

Kersemans V. et al. 2011; Hypoxia imaging using PET and SPECT: the effects of anesthetic and carrier gas on [Cu]-ATSM, [Tc]-HL91 and [F]-FMISO tumor hypoxia accumulation. PLoS One. 6:e25911. [PubMed: 22102855]

Kim JJ and Tannock IF 2005 Repopulation of cancer cells during therapy: an important cause of treatment failure Nat. Rev. Cancer 5 516–25 [PubMed: 15965493]

Kinoshita M et al. 2012 A novel PET Index, ^{18}F-FDG-11C-Methionine uptake decoupling score, reflects glioma cell infiltration J. Nucl. Med 53 1701–8 [PubMed: 23000747]

Kishino T, Hoshikawa H, Nishiyama Y, Yamamoto Y and Mori N 2012 Usefulness of 3′-deoxy-3′-^{18}F-fluorothymidine PET for predicting early response to chemoradiotherapy in head and neck cancer J. Nucl. Med 53 1521–7 [PubMed: 22872738]

Kubicek LN, Seo S, Chappell RJ, Jeraj R and Forrest LJ 2012 Helical tomotherapy setup variations in canine nasal tumor patients immobilized with a bite block Vet. Radiol. Ultrasound 53 474–81 [PubMed: 22731939]

Levine EA, Farmer MR, Clark P, Mishra G, Ho C, Geisinger KR, Melin SA, Lovato J, Oaks T and Blackstock AW 2006 Predictive value of 18-fluoro-deoxy-glucose-positron emission tomography (18F-FDG-PET) in the identification of responders to chemoradiation therapy for the treatment of locally advanced esophageal cancer Ann. Surg 243 472–8 [PubMed: 16552197]

Loh W-Y 2011 Classification and regression trees Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery 1 14–23

Machtay M et al. 2013 Prediction of survival by [^{18}F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/RTOG 0235 trial J. Clin. Oncol 31 3823–30 [PubMed: 24043740]

Madani I et al. 2007 Positron emission tomography-guided, focal-dose escalation using intensity-modulated radiotherapy for head and neck cancer Int. J. Radiat. Oncol. Biol. Phys 68 126–35 [PubMed: 17448871]

Minagawa Y et al. 2011 Assessment of tumor hypoxia by 62Cu-ATSM PET/CT as a predictor of response in head and neck cancer: a pilot study Ann. Nucl. Med 25 339–45 [PubMed: 21327756]

Nordsmark M and Overgaard J 2000 A confirmatory prognostic study on oxygenation status and locoregional control in advanced head and neck squamous cell carcinoma treated by radiation therapy Radiother. Oncol 57 39–43 [PubMed: 11033187]

Petit SF, Aerts HJWL, Van Loon JGM, Offermann C, Houben R, Winkens B, Ollers MC, Lambin P, De Ruysscher D and Dekker ALAJ 2009 Metabolic control probability in tumour subvolumes or

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

how to guide tumour dose redistribution in non-small cell lung cancer (NSCLC): an exploratory clinical study Radiother. Oncol 91 393–8 [PubMed: 19328570]

R Development Core Team 2014 R: A Language and Environment for Statistical Computing. 3.1.1 (Vienna: R Foundation for Statistical Computing)

Sato Y, Tsujikawa T, Oh M, Mori T, Kiyono Y, Fujieda S, Kimura H and Okazawa H 2014 Assessing tumor hypoxia in head and neck cancer by PET with 62Cu-Diacetyl-Bis(N4-Methylthiosemicarbazone) Clin. Nucl. Med 39 1027–32 [PubMed: 25140555]

Shusharina N, Cho J, Sharp GC and Choi NC 2014 Correlation of (18)F-FDG Avid volumes on pre-radiation therapy and post-radiation therapy FDG PET scans in recurrent lung cancer Int. J. Radiat. Oncol. Biol. Phys 89 137–44 [PubMed: 24725696]

Soto DE, Kessler ML, Piert M and Eisbruch A 2008 Correlation between pretreatment FDG-PET biological target volume and anatomical location of failure after radiation therapy for head and neck cancers Radiother. Oncol 89 13–8 [PubMed: 18555547]

Van Elmpt W, De Ruysscher D, Van Der Salm A, Lakeman A, Van Der Stoep J, Emans D, Damen E, Ollers M, Sonke J-J and Belderbos J 2012 The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer Radiother. Oncol 104 67–71 [PubMed: 22483675]

Vogelius IR, Hakansson K, Due AK, Aznar MC, Berthelsen AK, Kristensen CA, Rasmussen J, Specht L and Bentzen SM 2013 Failure-probability driven dose painting Med. Phys 40 081717 [PubMed: 23927314]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

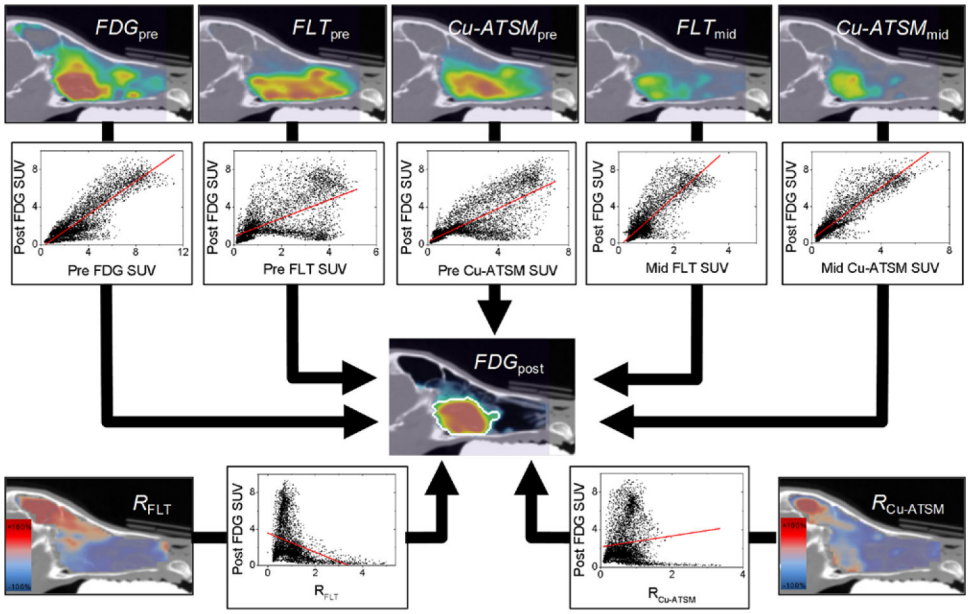

Figure 1.

An example patient is shown, illustrating the simple linear voxel regression method. Pre- and mid-treatment PET voxel values (*top row*) were regressed against follow-up FDG voxel values. Response maps (R_{FLT} and $R_{Cu\text{-}ATSM}$; bottom row) were also regressed against follow-up FDG. For logistic regression, post-treatment voxel values were dichotomized using an in-house segmentation method (outlined in white).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

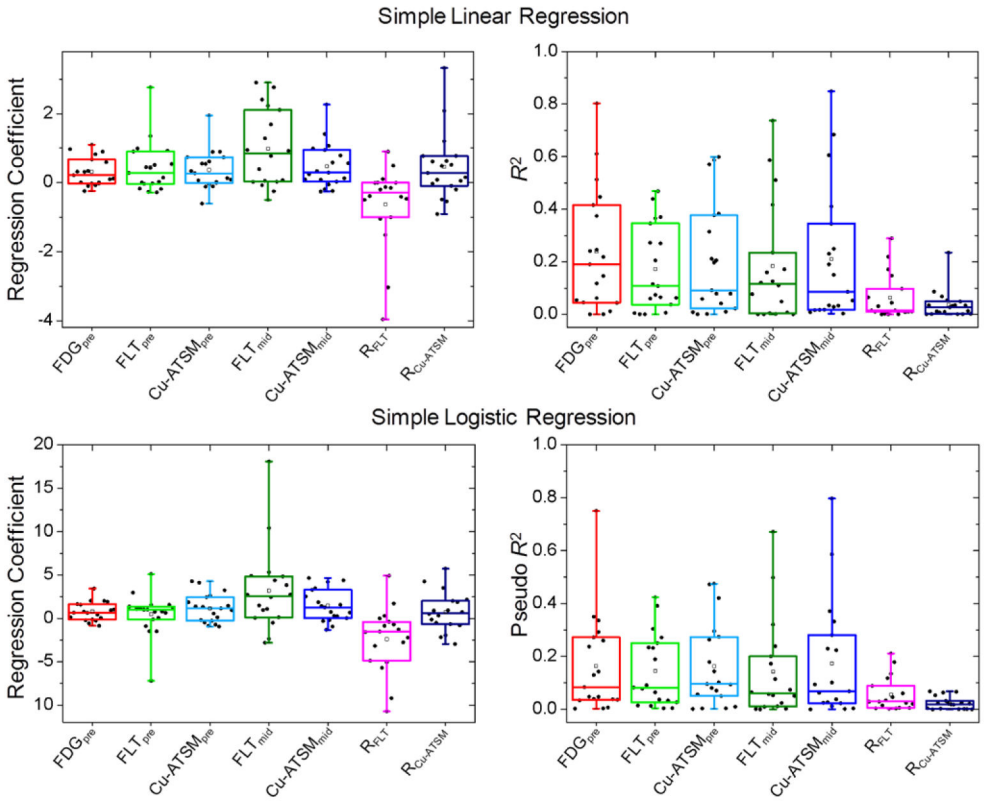

Figure 2.

Results from simple linear (top row) and logistic (bottom row) voxel regressions for all patients. Regression coefficients (left) and goodness of fits (right) are shown.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

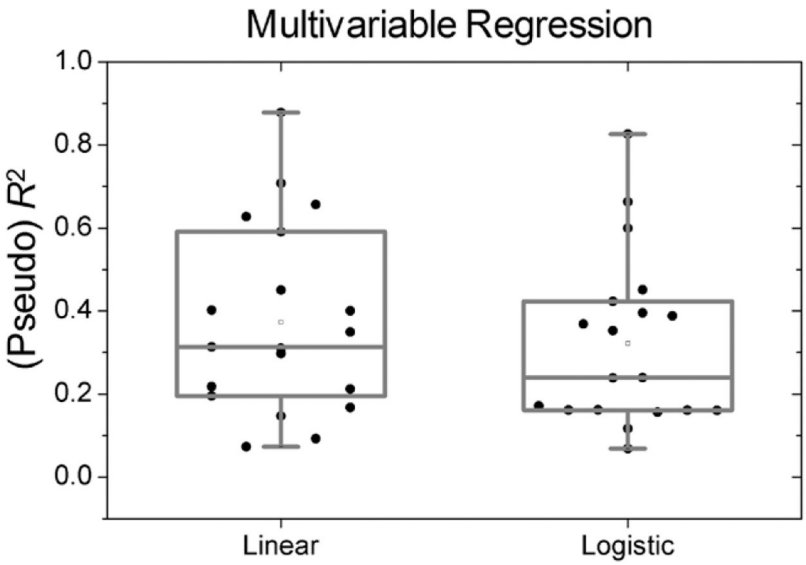

Figure 3.
R^2 and a pseudo R^2 for each patient's linear and logistic multiple regression models, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

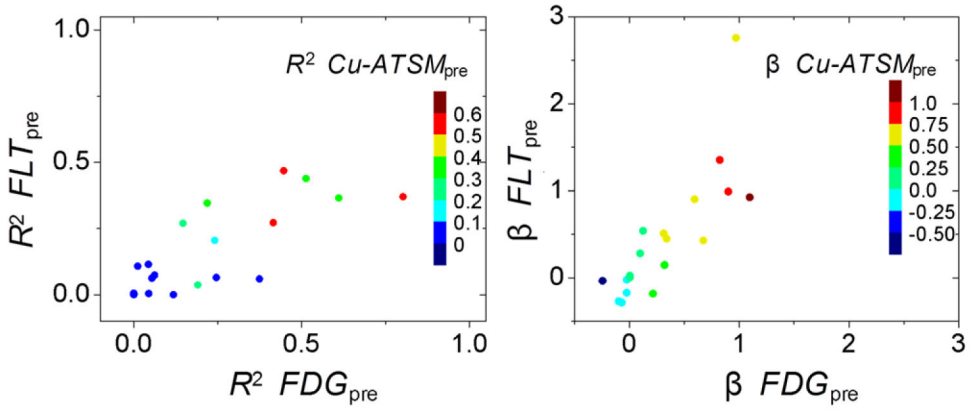

Figure 4.

Scatter plots illustrating that different predictor variables' R^2 were correlated across patients (left). Likewise, regression coefficients for different predictor variables were correlated across patients (right). Only results from pre-treatment simple linear models are shown here.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

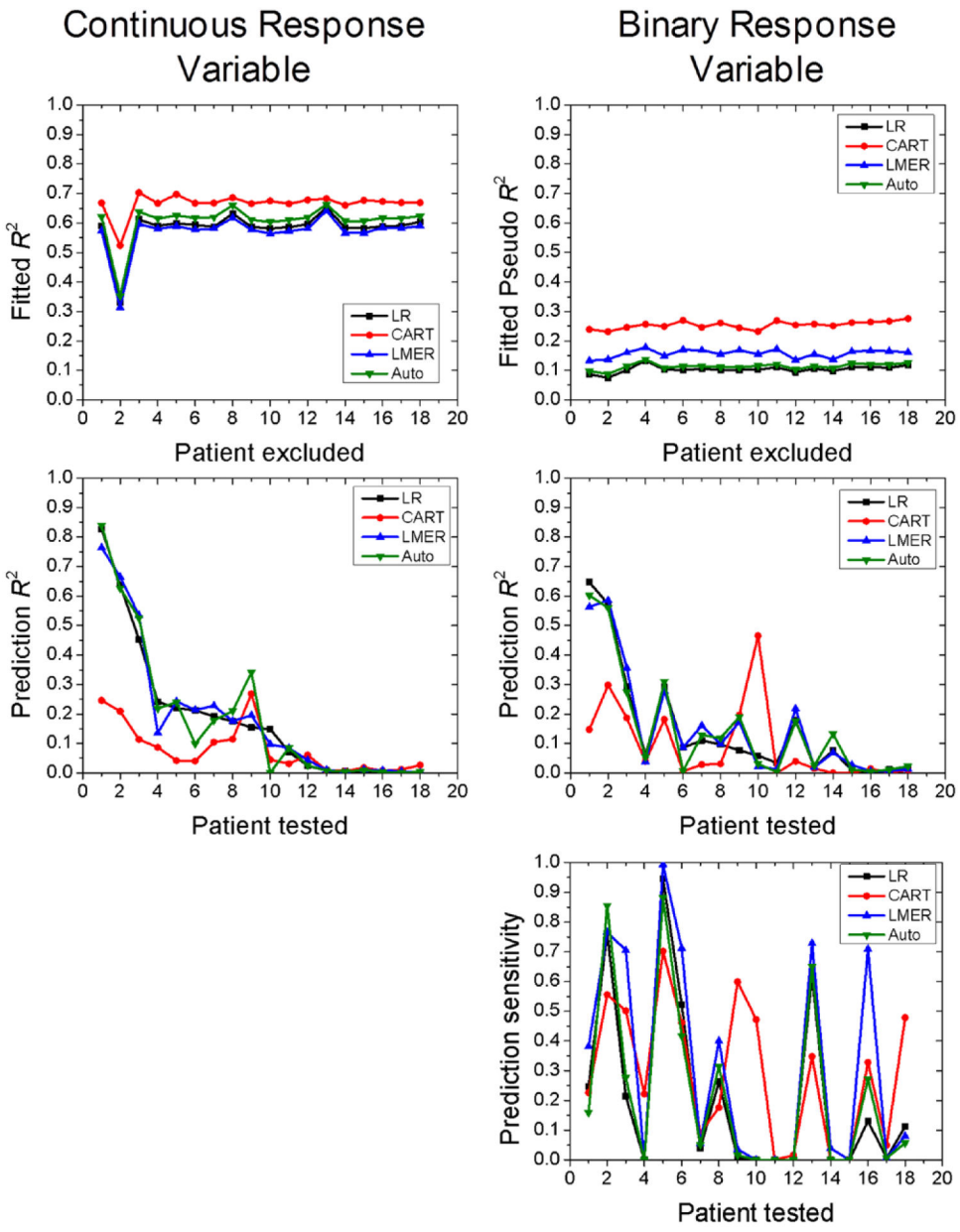

Figure 5.

Prediction modeling results for each patient, for both continuous response variables (left column) and binary response variables (right column). For each patient, fitted R^2 are shown, indicating how well the four models fit the training data set, which consisted of all patients but one. Prediction R^2 are also shown, indicating how well the trained models predicted the response values for the test patient. For binary response models, the prediction sensitivity is also shown. Patients are sorted according to the prediction R^2 for the linear regression model, from high to low. LR = multiple linear regression, CART = classification and regression tree, LMER = linear mixed effects regression, Auto = regression with neighborhood structures.