Using Machine Learning Techniques to Determine Dose Thresholds Predictive of Grade >= 2 Acute Rectal Toxicity in Prostate Cancer Patients treated with Radiation Therapy

Jiuyun Hu[1] , Jing Li[1] , Sujay Vora[2] ,Steven Schild[2] ,William Wong[2] , Wei Liu[2] ,Mirek Fatyga[2]

[1]Industrial Engineering , Arizona State University, Tempe, AZ; [2]Department of Radiation Oncology, Mayo Clinic

Background and Motivation

- Modern treatment planning techniques rely on Dose Volume Histogram (DVH) constraints to drive the optimizer.
- DVH constraints must be obtained from clinical data, by constructing Normal Tissue Complication Probability (NTCP) models..
- Modern databases allow us to search the entire DVH to find the most predictive DVH index. Such searches raise Multiple Comparisons (False Discovery) concerns, however.

➢ Purpose

- To design a method to search for most predictive DVH indices while addressing Multiple Comparison concerns.

Materials and Methods

Clinical data:

- Grade >=2 acute rectal toxicity in conventionally fractionated prostate cancer treatments for 79 IMRT patients. 20% (16) patients with toxicity.

Conventional, "common sense" approach to NTCP modeling.

- We used Univariate Logistic Regression to fit rectal toxicity to D_V% index, where D_V% signifies the lowest dose found in the V% percentage of rectal volume exposed to the highest dose.
- We fit a family of ULR models, each model with a single D_V% index, changing V% between 10% and 90%.
- We evaluated each model using Receiver Operating Characteristics (ROC) analysis and fit quality metric (p value).
- We pick the model with highest Area Under the Curve (AUC) and lowest p-value.

Machine Learning Approach

- We used a linear combination of equispaced V%_D indices, with 1Gy step, as an input to Multivariate Linear Regression model .The V%_D index is a percentage of rectal volume exposed to dose 'D' or greater.
- We used Fused Lasso Operator (FLO) to account for intrinsic correlations between V%_D indices
- We modified FLO with additional requirement that V%_D coefficients were non-decreasing as 'D' increased.
- This modification can be summarized as a hypothesis that there exists a threshold in 'D' beyond which dose volume relationship (expressed by V%_D) becomes predictive for toxicity.
- Since the coefficients can increase, multiple thresholds can be detected.

Results: Univariate Linear Regression

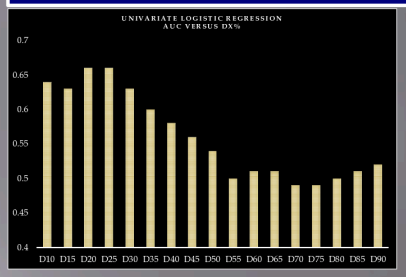

Fig 1

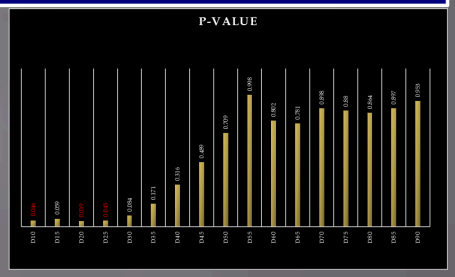

Fig 2

- Figure 1 and Figure 2 show univariate analysis, AUC in Fig.1 and p-value in Fig.2, plotted against D_V% indices. Note the non-random structure in these plots, with p-values monotonically decreasing while AUC values are monotonically increasing. We hypothesize that this structure is caused primarily by a combination of two factors: the dose threshold and strong correlations between D_V% indices.
- The maximum AUC in ULR is 0.67, for D_20%, and is consistent with values found in literature for NTCP modeling of rectal toxicity with dosimetry only.
- The minimum p-value is 0.04 for D_20%

Influence of Correlations between DVH indices on ULR analysis:

- The Fig.3 (below) shows the correlation between ULR coefficients in models shown in Figures 1 and 2, and Spearman correlation coefficients between D_X% indices, relative to D_20% index, in the D_20% to D_50% range.
- Note that ULR coefficients (horizontal axis) express the correlation between scored toxicity and DVH indices, while Spearman correlation coefficients (vertical axis) do not involve toxicity at all, just correlations between DVH indices in patient population.
- The strong correlation between these two sets of numbers suggest that ULR analysis is strongly affected by correlations between DVH indices.

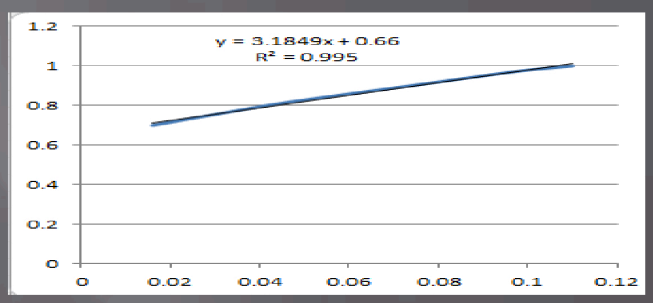

Fig.3 – correlation between ULR coefficients (horizontal axis) and Pearson correlation coefficient (vertical axis), relative to D_20% index in the 20% - 50% range, for ULR models from Fig.1

Results: Multivariate Regression with Knowledge Constrained Fused Lasso Operator

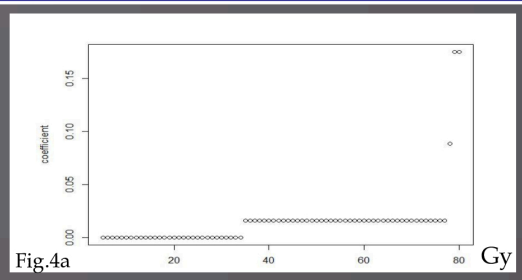

Fig.4a

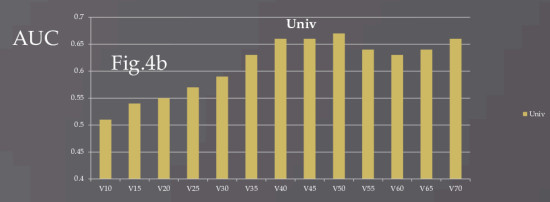

Fig.4b

Fig.4a (top panel) – Coefficients of Multivariate Logistic Regression (MLR) model with Knowledge Constrained Fused Lasso Operator, using V_%D indices with 1Gy step (2Gy dose equivalent).
Fig. 4b (bottom panel) Corresponding ULR analysis with AUC as an evaluation criterion. Please note a hint of two dose thresholds in the ULR analysis, also visible in Fig.2

Note:

- Best p-value for ULR model =0.04
- Best p-value for MLR model = 0.008
- The AUC is similar for MLR and the best ULR model (AUC ~ 0.7)

Summary and Conclusions

- We created an NTCP model based on Multivariate Logistic Regression (MLR) with machine learning augmentations. This model uses the entire DVH simultaneously (an array of V%_D indices).
- The model finds dose thresholds above which dose volume indices are predictive for clinical toxicity.
- The MLR model has much better fit quality (p=0.008) than the best ULR model (p=0.04), which suggests that the MLR model is more generalizable than the ULR models.
- The MLR model can not be used directly in commercial optimizers because of the limitations of clinically practical optimization algorithms.
- The best clinical use of the model is to identify dose ranges in which DVH indices are predictive for toxicity. A "common sense" approach with ULR can then be used to derive NTCP models that, though less accurate, can be used to derive optimization constraints that can be used in optimization.
- If one observes dose thresholds with statistical significance in the MLR model, one can use the "common sense" ULR technique without concerns for the effects of Multiple Comparisons.
- A monotonic change in fit quality p-values and in AUC of ULR models may be a signature of a dose threshold combined with correlations between DVH indices.