# Bonferroni correction is not applicable to the problem of dose volume predictors of clinical outcomes. An example of alternative statistical approaches.

Mirek Fatyga, PhD

Mayo Clinic, Arizona

MEDICAL PHYSICS COMMUNITY INSPIRATION INNOVATION IMPACT
2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

## ABSTRACT

:Aim: Dose Volume Histogram (DVH) indices remain a backbone of treatment planning in radiation therapy. Many outcomes studies treat DVH indices as independent candidate predictors and retain only indices that meet the statistical significance cutoff of p=0.05. Such a procedure raises multiple comparisons concerns (Bonferroni correction). We aim to show that Bonferroni Correction does not apply to this problem because of strong correlations between indices. Systematic scanning of the index space is more appropriate, combined with statistical methods that account for correlations

M&M : We used two patient outcomes studies: (1) grade >= 2 acute rectal toxicity in a cohort of 79 patients treated with conventionally fractionated radiation therapy (RT) for prostate cancer, and (2) cardiac toxicity in a cohort of 134 locally advanced, Non-small Cell Lung Cancer (LA-NSCLC) patients treated with conventionally fractionated radiation therapy, with overall survival (OS) as a clinical endpoint. Both cohorts were treated at Mayo Clinic Arizona. All patients had their treatment plans in Eclipse TPS (Varian, Inc). ESAPI interface was used to extract arrays of V%_D indices computed at 1Gy and 5Gy intervals. The V%_D index is the percentage of OAR volume subjected to dose “D” or greater. We first fit a family of NTCP models, each using one DVH index as a potential predictor. Subsequently, we fit generalized linear model with Fused Lasso constraint to the same data..

Results: p-values associated with NTCP models are not randomly distributed, but show systematic structure suggesting predictive and transition regions, with the transition region dominated by correlations with indices in the predictive region. The GLM model finds distinct dose thresholds

Conclusions: Bonferroni correction is not applicable to this problem. Systematic scanning of DVH index space, combined with advanced statistical methods, needs to be used instead

## CONTACT

fatyga.mirek@mayo.edu

## INTRODUCTION

Modern Treatment Planning Systems (TPS) save full dosimetry in their databases. This creates an opportunity to search for most predictive DVH indices by scanning their entire space. Arguably, modern day clinical outcomes studies should always use such systematic searches. How do we find the most predictive index in a systematic search?

## Data Sets

To answer the above question, we used two data sets:

- Grade >=2 acute rectal toxicity in 79 prostate cancer patients treated with conventionally fractionated radiation therapy. All patients had implanted fiducial markers for precise IGRT. *This data set was used to develop the model*.
- 134 locally advanced Non-small Cell Lung Cancer (NSCLC) patients treated with conventionally fractionated RT. We examined the correlation between cardiac dosimetry and Overall Survival (OS*). We applied the model to this data set.*

## Initial “common sense” approach

In the initial, common sense, approach we used logistic regression to model acute rectal toxicity. We inserted one D_X% index at a time into logistic regression and calculated the p-value for each model. Results are shown in **Fig.1**. We observe that D_X% indices are not independent predictor candidates, as there is a clear structure to this plot. Where is this structure coming from?

### Hypothesis

There are three distinct regions in Fig.1:

1. Non-predictive (rightmost, flat region)
2. Transition region with monotonically declining p-values. This region is dominated by correlations with indices from the predictive region.
3. True predictive region (leftmost).

The true biological threshold separates regions (3) and (2)

### Can we find a proof?

**Fig.2** plots the logistic regression coefficient (horizontal axis) for D_X% indices in the transition region, versus Spearman correlation coefficient with respect to the first index of the predictive region (D_20%). What is remarkable about this plot? The horizontal axis contains all the complexities of the clinical response. The vertical axis is just playing with treatment planning numbers in a computer. And yet, these two variables are nearly perfectly correlated. This is a strong indication that the transition region (2) is dominated by correlations with the predictive region.

### How can we capture the hypothesis in a model?

## KC-LASSO model.

A GLM model with Fused Lasso and radiobiologically motivated Knowledge Constraints.

- LASSO reduces dimensionality with small sample sizes
- Fused penalizes differences between adjacent coefficients to account for correlations.
- Knowledge Constraints stabilize the model and make it more interpretable for radiation therapy.

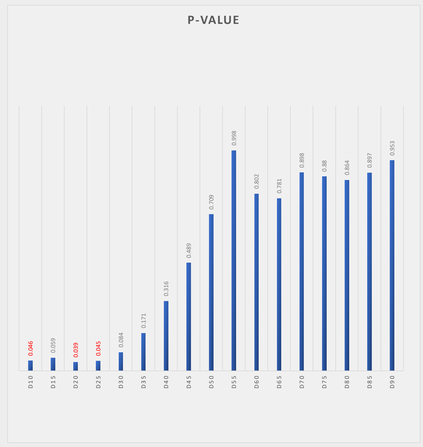


**Fig. 1**

P-values for a family of logistic regression models, with one rectal D_V% index at a time. Each bar corresponds to a model.

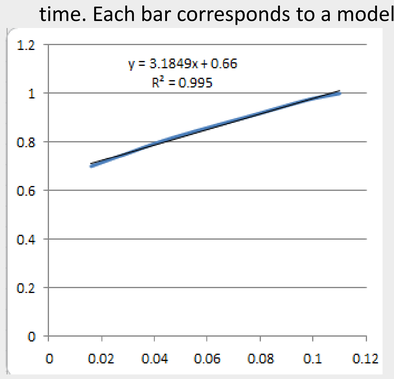


**Fig. 2**

X-axis: logistic regression coefficient in the transition region
Y-axis: Spearman correlation coefficient with respect to D_20%

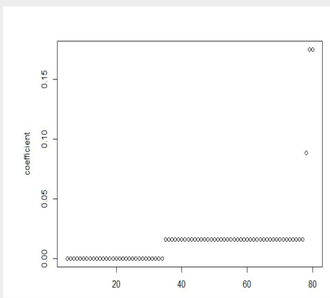


**Fig. 3**

KC-Lasso coefficients as a function of dose, grade 2 acute rectal toxicity

KC-LASSO detected two thresholds in model coefficients (Fig.3). This is encouraging, since existing literature suggests two predictive regions: high doses for rectal bleeding and intermediate doses for the remainder of grade 2 toxicity.

### Modelling Cardiac Toxicity in locally advanced (stage III) Non-small Cell Lung Cancer

We model Overall Survival with Cox Model and heart DVH. We first use single V%_D index per model. The p-value for each model is plotted against Dose in V%_D in Fig.4. Note pronounced minimum at approximately 55Gy (p=0.035) with a monotonic descent similar to Fig.1. The subsequent rise in p-values is caused by the loss of data past the prescription dose. ***The KC-LASSO model detected a single predictive threshold at approximately 51 Gy***.

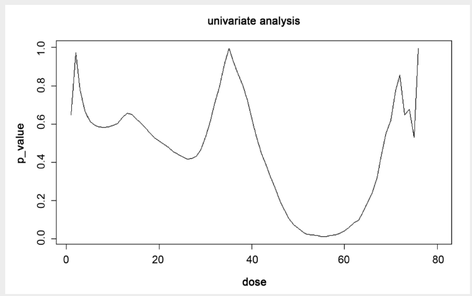


**Fig.4**

P-values for a family of Cox models, each model with one heart V%_D index, plotted as a function of Dose. Note pronounced minimum at ~55Gy, p=0.035

### Summary and clinical use

Ddosimetric indices are strongly correlated and thus cannot be treated as independent predictor candidates. One should scan the index space to look for patterns identifying the truly predictive index range, or use more advanced statistical analysis, like KC-Lasso. A single index from the predictive range can be used to construct a predictive model which is used to establish treatment planning constraints.

**Base Model:**

$$\theta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$X_i = V\%_{D_i}$$

Fused Lasso Operator

**Knowledge Constraints:**

*The positivity constraint*: $\beta_i \geq 0$

*The monotonicity constraint*: $\beta_{i+1} \geq \beta_i$