Modeling Renal Outcomes from Blood Pressure Variability During Non-Cardiac Surgeries

Department of Statistical Science

Duke University

Author: Callie Mao

Collaborators: Willem van den Boom, David B. Dunson

Clinical Collaborators: Rebecca Schroeder, Tracy L. Setji, Michael W. Manning

December 2016

Abstract

The consequences of short-term blood pressure variability, specifically renal damage, during non-cardiac surgeries are understudied. We investigate whether intraoperative mean arterial blood pressure variability during non-cardiac procedures influences renal outcomes, specifically percent changes in creatinine. We utilize generalized additive models to model this relationship and find a positive, clinically significant nonlinear relationship between generalized average variability for ranges outside of healthy norms and percent change in creatinine after taking into account demographic and other intraoperative confounders. Another model with a logarithmic transformed response fit model assumptions of normally distributed errors better than the first, but did not find a significant relationship. However, because not all modeling assumptions are upheld in the first model, we do not yet have enough evidence to verify this relationship without more advanced analysis or more advanced functional feature extraction for variability.

Table of Contents

1. Introduction

Long-term blood pressure variation has been considered a risk factor for cardiovascular disease [1, 2] and previous studies have found that long-term blood pressure variability is correlated with target-organ damage independent of mean blood pressure [3]. However, consequences of short-term blood pressure variability during non-cardiac surgeries are understudied. To discern whether prompt intraoperative interventions to combat short-term intraoperative blood pressure variability is necessary, it is important to determine the significance of volatile blood pressure deviations in the operating room.

Renal end-organ damage specifically is a complication of interest. Previous studies [4] have demonstrated that "surgical patients experiencing Acute Kidney Injury (AKI) postoperatively are eight times more likely to die within 30 days of surgery." Such studies have also found an association between hypotension from sustained mean arterial pressure of less than 55-60 mmHg during operations and postoperative acute kidney injury [4]. However, the effect of blood pressure variability during such operations is not as clear. AKI is a condition that is commonly measured from levels of creatinine, which the kidneys normally filter out. Hospitalized patients experiencing an increase of over 50% in creatinine are 6.9 times more likely to die [4]. Because of this, creatinine is a measure of interest for quantifying renal changes and potential complications following a surgical procedure.

We thus look to examine the effects of intraoperative blood pressure variability on changes in creatinine and attempt to capture accurate predictors of blood pressure variability in a low signal-to-noise environment. We focus specifically on non-cardiac procedures and consider the blood pressure trajectory curve, type of surgery, surgery time, intraoperative drug-classes, lab results, and demographic information to determine if there is a clinically significant relationship between various functions of blood pressure variability and changes in creatinine.

2. Data

2.1 Overview

Lab, demographic, and intraoperative information for over 431,480 different surgical cases performed at the Duke University Hospital were retrieved from Duke's electronic heath records and the Innovian Database of the Department of Anesthesiology stored within the Duke Hospital Information System. However, only 80,065 of the noted cases have both demographic information and mean arterial blood pressure measurements recorded over the course of a surgical procedure.

The 80,065 cases with both blood pressure (BP) and demographic information were taken from as early as June 19, 2000 and to as late as December 8, 2014. Each case has on average approximately 560.49 and a median of 525 measurements of MAP over time. Consecutive measurements are roughly half a minute apart, for a total of 44,875,910 data points measured for MAP.

2.2 Non-Cardiac Procedures

Since cardiac surgeries are often considered separately from non-cardiac surgeries, we filter the 80,065 cases with MAP measurements to 40,900 non-cardiac cases based on ICD-9 codes parsed from the procedure type dataset. We matched these cases with cardiac codes from the *2014 ICD-9-CM for Hospitals* [5] and utilized the remaining non-cardiac subset.

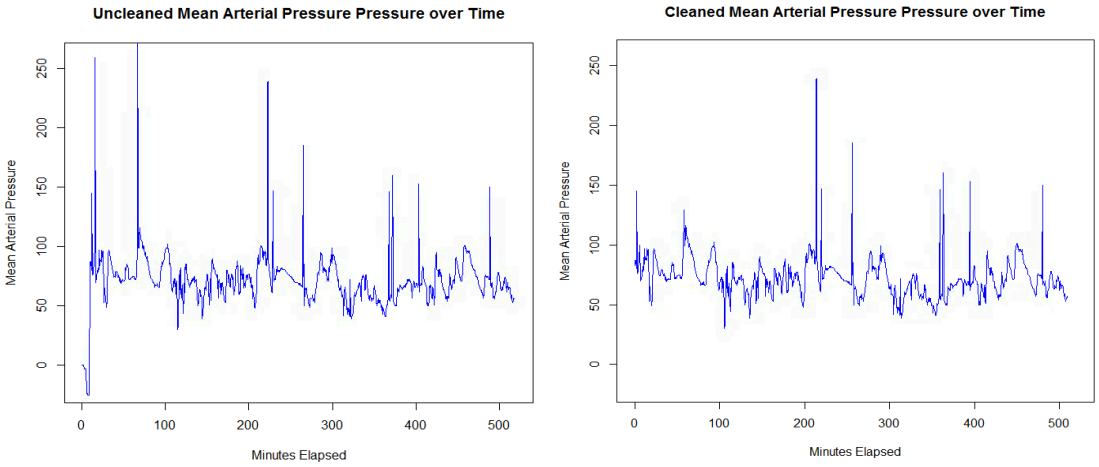
2.3 Cleaning

Many patients' BP measurements displayed extreme spikes to ranges outside of what is considered normal for human blood pressure. Because intraoperative blood pressure measurements usually

utilize invasive approaches during a surgical procedure, we may observe spikes in the data caused by flushing, which passes fluid externally through the measuring medical device and can cause dramatic and short-lived deviations that are not accurate readings of a patient's blood pressure. These artifacts are common in electronic anesthesia records and we clean these by removing improbable mean with MAP ranges below 10mmHg and above 250mmHG as suggested by our clinical collaborators.

Both an uncleaned and cleaned blood pressure trajectory curve are in the figure below, where extreme spikes in improbable ranges have been removed.

Figure 1. Uncleaned BP Trajectories vs. Cleaned BP Trajectories



After cleaning, we remove cases where no blood pressure measurements were remaining (likely due to measurement errors or lack of data) and are left with 38,799 cases in the final dataset of blood pressure measurements. The cleaned MAP curves of these non-cardiac cases form the basis of our analysis in extracting features that best model variability affecting changes in creatinine.

2.4 Datasets

In addition to the blood pressure data for each case, supplementary datasets were utilized to account for additional factors. A selected summary of datasets utilized for each patient and referenced later in the report are compiled in Table 1.

Table 1. List of Datasets

Label	Data Type	Use
Mean arterial pressure	Continuous measurements of blood pressure over time	Dataset that measures of variability are generated from
Procedure Type	Categorical labels for ICD-9 codes	Dataset that non-cardiac procedures are based on
Intraoperative fluids	Different types of intraoperative fluids. For each type, there is a continuous value depicting its dosage	Dataset used to determine whether a patient received insulin intraoperatively
Case Demographics	Different types of demographic information. For each demographic, there is continuous data describing patient characteristics matching the surgical case in question	Dataset used to determine age, height, and weight
Case Demographics: Gender	Categorical data describing patient gender	Dataset determining gender
Creatinine	Continuous data describing creatinine lab results with time labels	Dataset used to determine the response variable
Lab results	Fifferent type of lab results. For each lab, there is a time label with a	Dataset used to filter by HDL cholesterol, fasting glucose,

	continuous value quantifying a lab result	and A1c (glycated hemoglobin) results
Admission/Discharge Date	Timestamp denoting admission and discharge dates.	Dataset used in calculating differences in creatinine and preoperative lab results
Postoperative Outcomes	Continuous data denoting length of stay in hours and binary data indicating death	Dataset used for length of stay, which is used for exploratory data analysis

2.5 Subsetting the Data

We are interested in obtaining a more homogeneous population of patients at greater risk of kidney complications. This would reduce heterogeneity that could affect our modeling and potentially allow us to see more signal suggesting increased changes in percent creatinine difference. We include some considerations below.

2.5.1 Patients at Risk of Metabolic Syndrome

Our clinical collaborators suggested looking at a subset of patients with metabolic syndrome. Metabolic syndrome is a risk factor for diabetes and cardiovascular disease and is also associated with increased risk for chronic kidney disease in nondiabetic adults. [5]

The dataset that we are utilizing does not include labeling for metabolic syndrome, so we instead consider a possible subset of patients who are at risk for metabolic syndrome. Using guidelines set by the International Diabetes Federation (IDF) and the American Heart Association (AHA), we selected 24,419 patients that we refer to as our $atRisk$ subset who matched at least one of the

following that are used in diagnosis or that are acknowledged as causative factors of metabolic syndrome:

- Obesity: BMI > 30 (IDF)
- HDL cholesterol < 40 mg/dL in males (IDF)
- HDL cholesterol < 50 mg/dL in females (IDF)
- Fasting plasma glucose > 100 mg/dL (IDF, AHA)

These selection criterion do not involve physician interpretation and risk being too conservative in subsetting patients that are potentially at risk.

2.5.2 Patients with A1c Measurements

Both to address the concerns of the subset discussed in 2.4.1 and to utilize A1c90day as a predictor, we also considered a subset of 3,323 patients who had A1c lab exams requested within 90 days before surgery. This measurement is referred to as $A1c90day$ which is discussed in more detail in Section 3.2.5. The presence of *A1c90day* could be an indicator of additional risk factors that physicians find suspicious, allowing for a more selective and homogenous dataset.

As we can see in the boxplots of Appendix A, patients who had A1c measurements tended to be older, more overweight, and more often male. They also tended to have more repeat surgeries, longer postoperative stays, higher glucose average from the day of surgery and the first post-operative day ($GlucosePOD1$), and higher $creatinineDiff$ and $percentRecovery$, which suggest adverse postsurgical outcomes. Thus, we have reason to believe that subsetting by patients with A1c measurements may allow us to observe patients who are more at risk.

3. Predictors

This section describes the various predictors extracted from the data discussed in Section 2.

3.1 Variability Feature Extraction

Quantifying variability across time series in medical data can be challenging. It can be difficult to determine whether several severe spikes, prolonged variability above a certain threshold both in time and amplitude, circadian patterns, or a countless number of other combinations may significantly affect outcomes. Because of this, simple measures like standard deviation may not be able to capture enough information within widely volatile and disparate blood pressure curves, and thus we look to other measures of quantifying variability.

3.1.1 Generalized Average Realized Variability (*gARV*)

We utilize a variability metric researched in detail in Hansen et. al [10], which found that average realized variability (ARV), which measures short term blood pressure variation given ordered measurements taken equal time intervals apart and estimates variability better than standard deviation in times-series data. ARV can be formalized as

$$ARV = \frac{1}{T}\sum_{k=1}^{N-1} t|BP_{k+1} - BP_k|$$

where N is the number of blood pressure measurements, T is the total time from the first to the N^{th} blood pressure reading, BP_k is the blood pressure reading at the k^{th} measurement, and t is the time interval between each reading.

However, Sun et. al notes that ARV will overestimate the variability of steep changes in blood pressure for uneven time intervals and proposes utilizing generalized average realized variability ($gARV$), which is robust to varying distances between readings [2]. Generalized ARV is defined by

$$gARV = \frac{1}{T}\sum_{k=1}^{N-1}|BP_{k+1} - BP_k|$$

After cleaning the dataset to remove improbable values, our dataset does not involve readings at equal time intervals apart, and thus we utilize $gARV$ as a measure of blood pressure variability.

3.1.2 *gARV* Outside of Healthy Ranges (*gARVOutsideRange*)

Because of the complex nature of blood pressure, we developed our own extensions of $gARV$ and extracted additional variability features that would be more relevant to renal outcomes. We explored individual case BP curves with a focus on features quantifying measures of MAP variability outside ranges deemed healthy by clinicians. In cardiac surgery patients, past research has found a relationship between intraoperative blood pressure deviations outside of a targeted range and mortality [11,12]. These deviations are measured by magnitude and duration and provide impetus to explore blood pressure excursions out of known healthy ranges, especially in models looking at potential outcomes affecting the kidneys.

Previous medical literature has supported the kidney's ability to auto-regulate within a given range [13] and our clinical collaborators suggested that blood pressures below 70 and above 110 were considered potentially unhealthy and outside such ranges. These numbers were used as thresholds to derive additional measurements of variability potentially having greater impact on kidney complications.

While $gARV$ taken for an entire BP curve is a measure of global variability, we define an alternative version. We instead take an aggregate of $gARV$'s computed on subsets of BP trajectories that are outside of the healthy range. Measurements of $gARV$ outside the range are considered by determining the variability of "hills" and "dips" outside the healthy range. The $gARV$ is calculated for each individual hill, which is defined by a consecutive sequence of measurements above the healthy range, and each individual dip, which is defined by a consecutive sequence of measurements below the healthy range. The mean for the resulting $gARV$'s weighted by the time spent in each hill or dip is then calculated. This can be formalized as

$$gARVOutsideRange = \frac{1}{nT}\sum_{i=0}^{n} t_i gARV(i)$$

where n represents the total number of hills and dips, i represents the i^{th} hill or dip in the blood pressure measurements, $gARV(i)$ represents the $gARV$ of the i^{th} hill or dip, t_i represents the time spent outside the range for the i^{th} hill or dip, and T represents the total amount of time spent in surgery. This feature measures $gARV$ outside the healthy range, which we will refer to as $gARVOutsideRange$.

3.1.3 Relative Time Spent Outside of Healthy Range (*relTime)*

Another feature of interest is the relative time a patient's MAP spent outside of the healthy range compared to the total time that the patient was operated on during the procedure. Using the same notation as above, this can be represented as the sum of times for all the existing hills and dips divided by the total amount of time in surgery.

$$relTime = \frac{\sum_{i=0}^{n} t_i}{T}$$

This predictor provides another frame of reference on how severe a patient's variability may be. For example, two patients with the same $gARV$ may be differentiated by the relative time spent

outside of the healthy range. The patient who had a higher relative time would be hypothesized to be more at risk of developing complications.

3.1.4 *gARV* Above and Below Healthy Ranges (*gARVAboveRange* and *gARVBelowRange*)

We also considered differences in being considered outside of a healthy range by recalculating $gARVOutsideRange$ but only utilizing one of the two original ranges: above the healthy range, which included blood pressure measurements above 110 inclusive, or below the range, which included blood pressure measurements below 70 inclusive. The same calculations as in Section 3.1.2 were performed on only hills for obtaining $gARVAboveRange$ and only dips for obtaining $gARVBelowRange$.

3.1.5 Relative Time Above and Below Healthy Ranges (*relTimeAbove* and *relTimeBelow*)

Similarly to $gARVAboveRange$ and $gARVBelowRange$, relative time was calculated for exclusively blood pressure measurements made above 110 inclusive to obtain the relative time spent above healthy ranges, which we refer to as $relTimeAbove$, and for exclusively blood pressure measurements below 70 inclusive to obtain the relative time spent below healthy ranges, which we refer to as $relativeTimeBelow$.

3.1.6 Mean Blood Pressure Above and Below Healthy Ranges (*meanBPAbove* and *meanBPBelow*)

Mean blood pressure was also calculated for only blood pressure readings above 110 inclusive to obtain $meanBPAbove$ and below 70 inclusive to obtain $meanBPBelow$. These provide additional information on the average magnitude and amplitude of blood pressure variability deviations above or below the healthy range.

Thus, we consider and test $gARV$, $gARVOutsideRange$, $gARVBelowRange$, and $gARVAboveRange$, $relTimeAbove$, $relTimeBelow$, $meanBPAbove$, and $meanBPBelow$ as predictors that measure variability.

3.2 Confounders

3.2.1 Mean Blood Pressure (*meanBP*)

The relationship between mean blood pressure and organ damage has already been well-studied. Previous studies have found that mean blood pressure that deviates outside of a targeted range is known to relate to more adverse outcomes. [1, 7-9]

The mean of the cleaned MAP curves was calculated for each patient and utilized as an aggregate metric for the patient. While this value does not provide much information about variability, it can be useful in differentiating patients from each other or provide supplemental information to later measures of variability.

3.2.2 Demographics

Additional variables were added to the model to control for confounding factors. General demographic features were considered in formulation of the model and these included gender, age, and BMI. Gender is coded as one if female and zero if not, while age and BMI are both continuous.

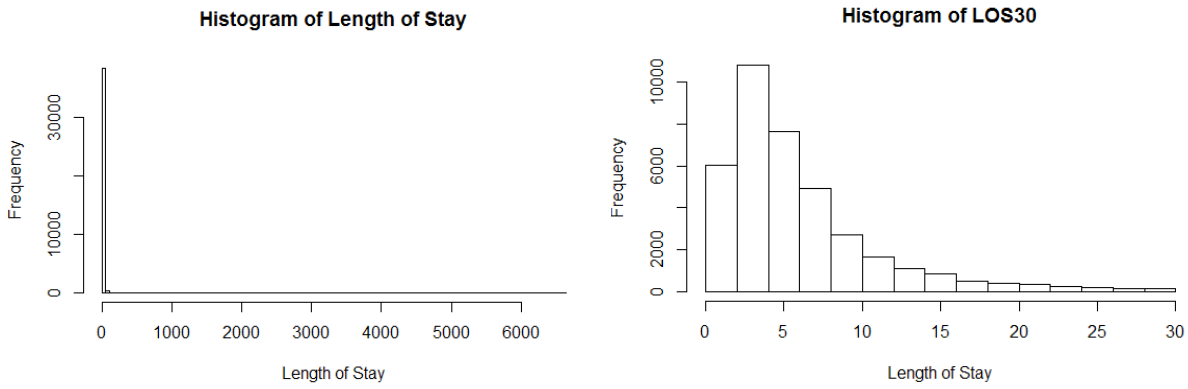
3.2.3 Length of Stay

We considered length of stay in our exploratory data analysis. While it could potentially be utilized as an outcome variable to model complications during surgery, it does not directly measure renal

outcomes and thus is just used to determine differences in patients who have and do not have A1c measurements in Appendix A.

We use the variable LOS30, which is a continuous variable that lists the length of stay in minutes. We delete any cases with a length of stay greater than 30, since these are likely outliers or in areas of sparse data that are difficult to include and interpret. A histogram of length of stay shows elimination of extreme outliers on the right tail after applying this correction.

Figure 2. Comparison of Histograms of Length of Stay and LOS30



3.2.4 Insulin

We observed a binary variable classifying whether a patient received insulin intraoperatively during a procedure. The variable we utilize in our models is coded as true if a patient received insulin via intraoperative fluids during a procedure and false if not.

3.2.5 Glycated Hemoglobin (*A1c90day*)

As mentioned in Section 2.5.2, we also observed the mean of glycated hemoglobin or A1c lab results between 90 days pre-surgery and the day of surgery, which we refer to as $A1c90day$. We

are suspicious of patients at risk of metabolic syndrome and we have reason to think that A1c results (and insulin) may affect the outcome.

However, as in the case with any elective medical lab, A1c measurements only exist for a portion of patients. In fact, only 8.6% of the cleaned 38,799 cases had measurements available for A1c. Including this variable as a predictor would subset the data further to only patients who had A1c measurements within 90 days before their surgery, which could be suggestive of an at-risk patient who a physician would request the elective lab for.

3.2.6 Plasma Glucose and HDL Cholesterol

We utilized lab results on plasma glucose and HDL cholesterol to filter the dataset down into the $atRisk$ subset described in Section 2.5.1, but these values were not considered as predictors.

3.2.7 Total Time

We also considered looking at total time spent during a procedure, but discovered that it was collinear with predictors of interest. Additionally, because total time is not flexible enough to be controlled in an operating room, we decide to not include this variable in later analysis.

3.3 Response Variables (*creatinineDiff* and *percentRecovery*)

Medical classification and staging systems like AKIN and RIFLE classify acute kidney injury based on ranges of percent increases in creatinine [17]. We evaluated the data to determine an appropriate response variable to reflect changes in creatinine during a procedure.

We compared the creatinine measurement prior and nearest to the start of surgery, C_0, to the peak creatinine between the end of surgery and discharge, C_f . The percentage difference $creatinineDiff$ was then calculated and utilized as the response variable.

$$creatinineDiff = \frac{C_f - C_0}{C_0}$$

As suggested by our clinical collaborators, we also considered percent recovery as a response variable, which we refer to as $percentRecovery$. This was defined as the percent difference between the maximum creatinine reached post-surgery, C_f , and the creatinine measured right before discharge, C_d.

$$percentRecovery = \frac{C_f - C_d}{\mathrm{C_f}}$$

4. Modeling Approaches

4.1 Generalized Linear Model

We performed our preliminary, exploratory analysis using linear models, which involve some combination of our measures of BP variability and confounders.

A generalized linear model (GLM) is composed of linear predictors

$$g\big(E(Y_i)\big) = \beta_0 + \beta_1 x_{1i} + \cdots \beta_p x_{pi}$$

where $E(Y_i)$ is the response variable for the i^{th} patient, $g(Y)$ is a link function that allows a function of the response variable to vary linearly with the linear predictor, and x_{ji} are the corresponding data associated with the j^{th} predictor (which could be a measure of BP variability or a confounder in our application), and p is the total number of parameters in our model excluding the intercept.

4.2 Generalized Additive Model

Because of the complexity involved in modeling variability, we turn to an approach allowing more flexibility than GLMs that removes the restriction that each predictor's marginal effect is linear. Generalized additive models (GAM) allow for more adaptable predictor functions that are still relatively interpretable and can reveal nonparametric trends that linear models are incapable of detecting.

GAMs can be used to model relationships between predictors and response variables that follow smooth functions that can be either linear or nonlinear. These smooth relationships can be estimated simultaneously and the response can be determined additively. [14]

A generalized additive model can be written as

$$g\big(E(Y_i)\big) = a + s_1(x_{1i}) + \ldots + s_p\big(x_{pi}\big)$$

where s_k denotes the smooth, nonparametric function for the k^{th} predictor variable.

GAMs allow us to determine the marginal additive impact of a single variable while also giving us the ability to regularize the smoothness of the nonparametric functions to avoid overfitting. We can thus observe either parametric and non-parametric effects depending on which variables we add in that include a smoothing function *s*.

4.3 Model Fitting

The most basic linear model utilizes an identity link function with normally distributed errors. We continue to utilize this link function and assumption of normally distributed errors in our GAMs. We utilize the R package 'mgcv' [15] for fitting GAMs, which adaptively smooths and utilizes a Bayesian approach to variance estimation. We are particularly interested in the resulting coefficients or smooth functions for the variables discussed in Section 3.1.

5. Results

We built several generalized additive models after selecting differing combinations and measures of changes in creatinine, MAP variability, subsets, and confounders of the variables described in Section 3.

We chose to utilize $creatinineDiff$ as the response variable in our final model due to easier medical interpretability over $percentRecovery$. The calculated values of $percentRecovery$ contained a disproportionately large number of zeroes in the response (35%) likely due to only one creatinine observation post-operation, which in fact would not provide us any information about recovery. Additionally, modeling these non-normally distributed errors, even with transformations, added additional complexity that could be minimized with the use of $creatinineDiff$.

We found greater significance in models utilizing $gARVOutsideRange$ and $relTime$ than when variability measures were split into $gARVAboveRange$ and $gARVBelowRange$ or $relTimeAbove$ and $relTimeBelow$, which can be seen in Appendix B. This suggests that with the approaches that we used, differentiating between hypotensive and hypertensive variability may not be beneficial in modeling changes in creatinine.

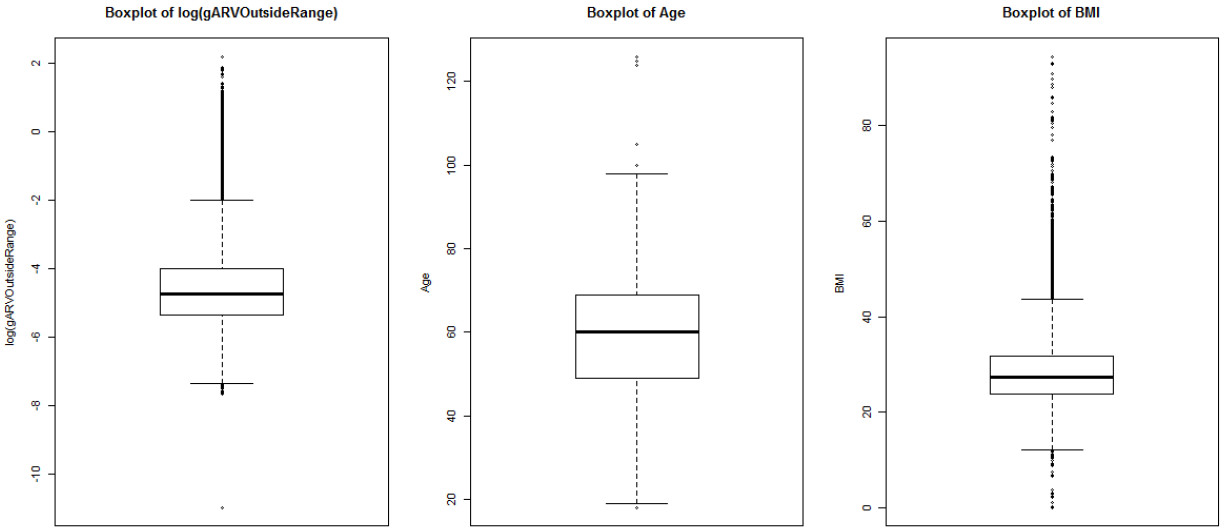
We examined the subset of patients at risk of metabolic syndrome, which as defined in Section 2.5.1 was determined by having at least one of the mentioned risk factors. However, models built utilizing this dataset did not show any significant results, most likely because the selection criteria were too conservative leading to too large and heterogeneous of a population.

Instead, we examined the subset of patients with A1c measurements, which resulted in more promising figures. This again is likely due to physician involvement, since physicians request A1c

labs for patients who they determine may need A1c monitoring after holistic examination, as opposed to using a crude, quantitative method which is only based on numbers.

Using these datasets, we observe heavy skew in several variables as can be seen in Figure 3.

Figure 3. Boxplots of log($gARVOutsideRange$), Age, and BMI



Due to the heavy skew, we decide to remove cases with $gARVOutsideRange$ values that are greater than 0.5 ($log(0.5) = -0.7$). Most of these are due to a single short spike outside of the healthy range, which can be seen in Appendix C, resulting in a high $gARVOutsideRange$. These curves are difficult to compare to those with more data and contain remnants of the artifacts mentioned in Section 2.3 that were not fully removed.

Additionally, due to medical differences and the unlikelihood of the following occurrences, we make conservative removals of patients with age greater than 110 years, BMI greater than 70, and BMI less than 5 before proceeding with the analysis. In the following sections, we present several final selected models of interest for discussion.

5.1 Variability Outside Healthy Ranges on Patients with A1c Measurements

We fit the following GAM with confounders

$$creatinineDiff = \beta_0 + s(meanBP) + s(gARVOutsideRange) + s(relTime) + s(BMI) + s(age) + gender + insulin + s(A1c90day) + \varepsilon$$

5.1.1 Model Results and Analysis

The results of the GAM are displayed in Table 2 and the nonparametric component smoothing functions on the scale of the linear predictor are displayed in Figure 4.

Table 2. Parametric and Nonparametric Effects of Model 5.1

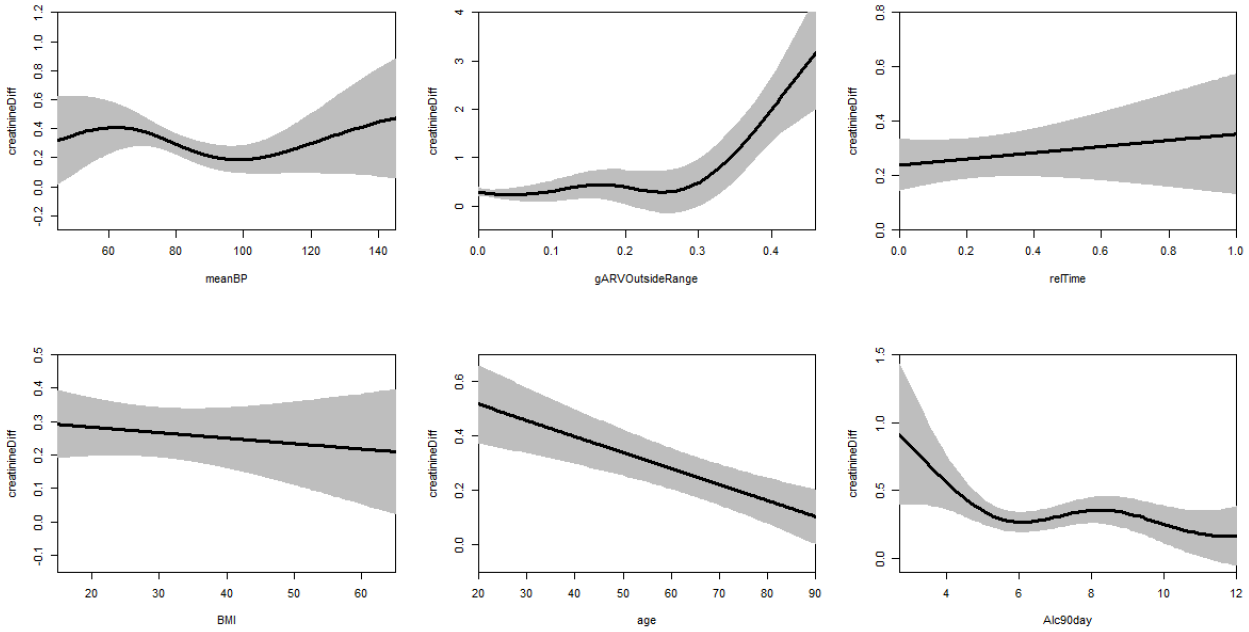
Sample size = 2477

Parametric Effects	Estimate	*t* value
Intercept	$3.2 * 10^{-1}$	12.20 ***
Gender	$4.5 * 10^{-2}$	1.17
Insulin	$2.5 * 10^{-1}$	3.50 ***

Nonparametric Effects	*F* statistic
s(***meanBP***)	2.23 *
s(***gARVOutsideRange***)	5.20 ***
s(***relTime***)	0.69
s(***BMI***)	0.45
s(***age***)	18.16 ***
s(***Alc90day***)	1.94 *

* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001

Figure 4. Plots of the Component Smooth Functions of Model 5.1



The model detects the parametric effect of receiving insulin during the procedure as well as the nonparametric effects of $meanBP$, $gARVOutsideRange$, age, and $Alc90day$ as significant. We observe the marginal plots for their marginal effects and see that $meanBP$ seems to have a slight dip in $creatinineDiff$ at blood pressure values closer to the mean, which we would expect based on literature mentioned in Section 3.2.1. For example, we see a difference of 20 percentage points between a $meanBP$ of 70 ($creatinineDiff = 0.40$) and a $meanBP$ of 95 ($creatinineDiff = 0.20$).

It also appears that with increased relative time outside of healthy ranges, which is utilized as a measure of variability, $creatinineDiff$ increases slightly (a difference of 28 percentage points in $creatinineDiff$ between the 1st and 3rd quartile of $relTime$); however, this nonparametric effect was not found to be significant. The nonparametric effect of age was also found to be significant, and we see a difference of 11 percentage points between the 1st and 3rd quartile of age. $Alc90day$

too was found to have significant nonparametric effects, and we see higher $creatinineDiffs$ from A1c90day approximately below 6, although the relationship is less clear and supported by less data.

Additionally, the parametric effect for receiving insulin during a procedure is statistically significant and patients receiving insulin tended to have $creatinineDiffs$ that are 25 percentage points higher than those without. Insulin breaks down glucose in patients with hyperglycemia, and receiving insulin may be related to patients that are more at risk, so the result that we see may be suggestive that patients more at risk may have more adverse outcomes.

We can see that the relationship between $gARVOutsideRange$ and $creatinineDiff$ is positive and appears to increase linearly for $gARVOutsideRange$ values approximately past 0.27. Before that point, it is difficult to quantify any clinical significance in the given relationships.

We can estimate $creatinineDiff$ using this model while keeping all variables other than $gARVOutsideRange$ at their mean values. We then find the estimated $creatinineDiff$'s at the listed percentiles for $gARVOutsideRange$ in Table 3.

Table 3. $creatinineDiff$ values with changing $gARVOutsideRange$ percentiles.

$gARVOutsideRange$ Percentile	Value	Estimated $creatinineDiff$
25^{th}	0.0047	0.29
50^{th}	0.0085	0.29
75^{th}	0.017	0.27
95^{th}	0.062	0.26
99^{th}	0.19	0.42
99.5^{th}	0.27	0.34
99.9^{th}	0.41	2.11

For cases with $gARVOutsideRange$ at the 99.5th percentile, or 0.27, patients experienced a 34% increase in creatinine at the peak post-operation level relative to the pre-operative baseline. For cases with $gARVOutsideRange$ at the 99.9th percentile, or 0.41, patients experienced a 211% increase. An additional 0.27 added 177 percentage points, which is very clinically significant. According to both the RIFLE and AKIN class classification of staging systems for AKI [17], a creatinine increase of 1.5-fold to 2.0-fold is considered at risk for AKI and a creatinine increase of 2-fold to 3-fold is considered in the stage of kidney injury.

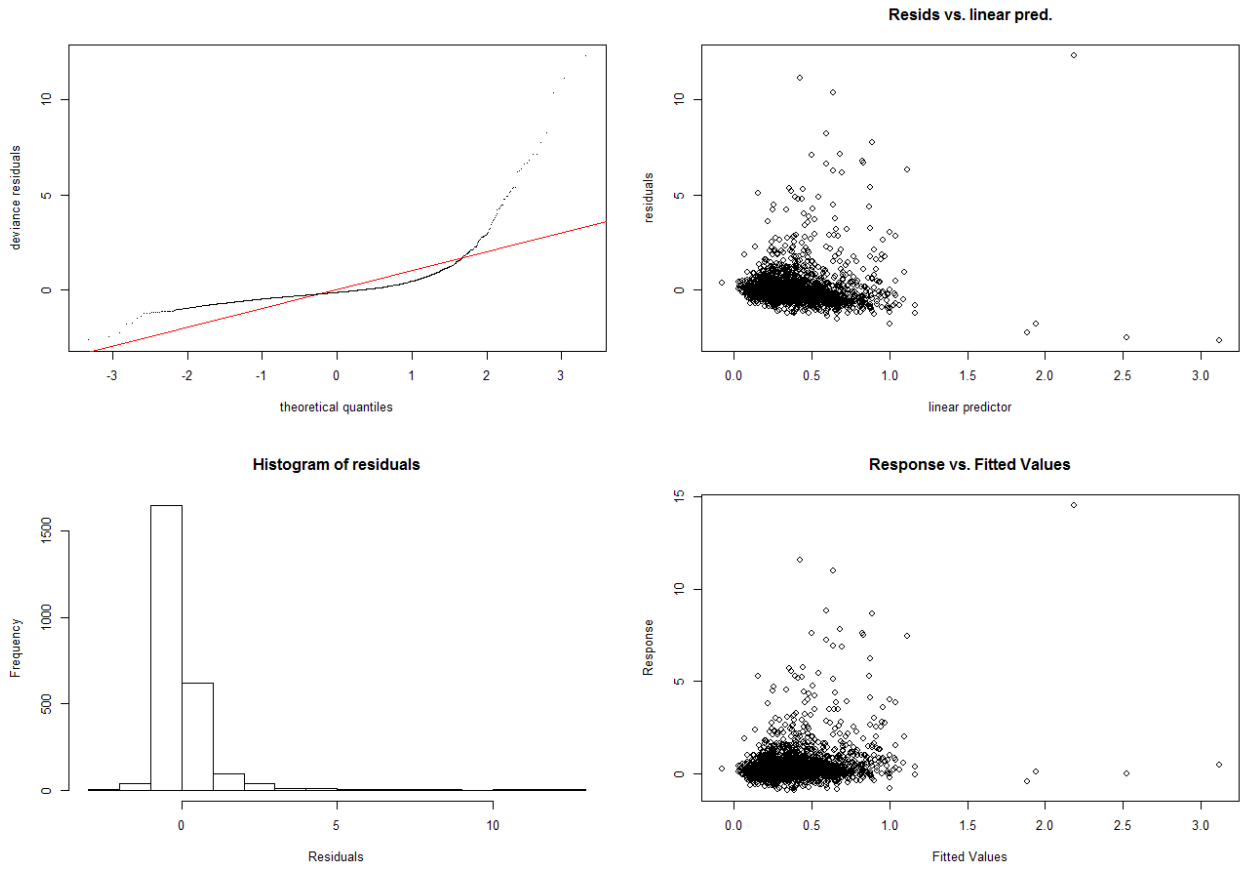
We notice that the relationship does not start until some of the highest percentiles, specifically after 0.27 where there are only 13 data points out of the 2477 in the model, which makes up 0.5% of the dataset. While this is a very small percent, AKI has also been found in previous studies to occur in approximately 1% of patients after non-cardiac surgery, so this number may not be all too surprising or uninformative. [16]

While it is important to take caution in interpreting clinical significance in only 0.05% of the dataset, only a small proportion of cases experience severe kidney complications after surgery, so this number may still be of interest. Within our dataset of 2477 patients, a total of only 150 patients, or 6.05% of the dataset, have $creatinineDiff$ above or equal to 1.5-fold, and it is possible that further investigation into this subset may reveal more information.

5.1.2 Model Diagnostics

We also check our model diagnostics but notice some problems, which are displayed in Figure 5.

Figure 5. Model 5.1 Diagnostic Plots



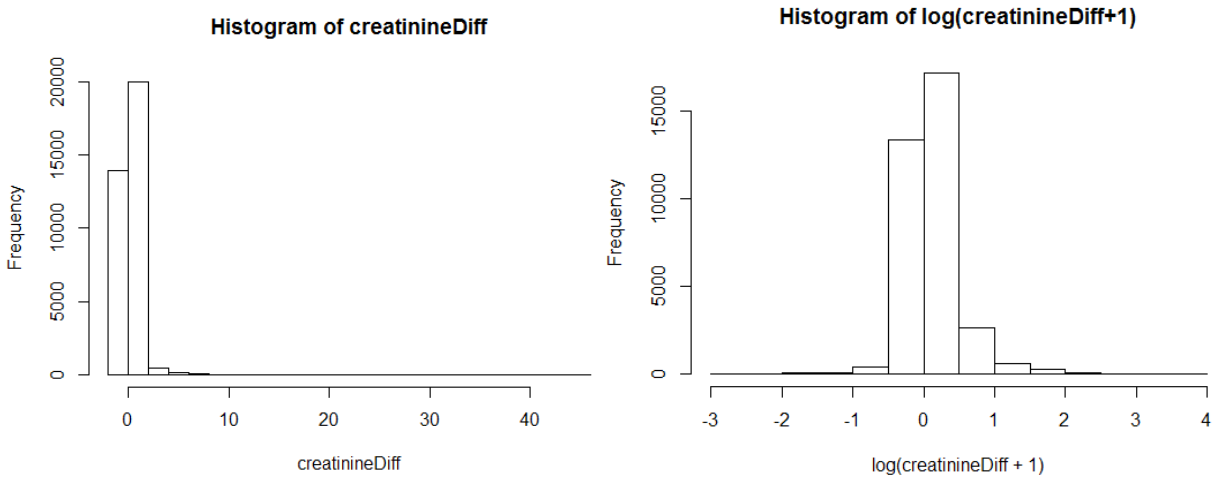
As we can see from both the QQ plot, which does not follow the $y = x$ trend line and suggests right skew, as well as the skewed histogram of residuals, the model does not uphold the assumption of normality. The residual plot is not randomly distributed about $y = 0$ and the response vs. fitted values do not seem to closely follow the $y = x$ line.

While this model does seem to suggest some interesting results for a small subset of values following a range of $gARVOutsideRange$, the assumptions of normality do not appear to be upheld. We should take even greater caution when interpreting these results.

5.2 Using a Logarithmic Transformation for *creatinineDiff*

As we have seen in the model diagnostic plots in Figure 5, assumptions that errors are normally distributed are not fully upheld. To try improving this by removing skew, we add 1 to $creatinineDiff$ to ensure that values are greater than 0 and then take the log of $(creatinineDiff + 1)$. We can see the results of this transformation in Figure 6.

Figure 6. Comparison of Distributions of $creatinineDiff$ and $log(creatinineDiff + 1)$



With this transformation, we try to fit the generalized additive model

$$\log(creatinineDiff + 1) = \beta_0 + s(meanBP) + s(gARVOutsideRange) + s(relTime) + s(BMI) + s(age) + gender + insulin + s(A1c90day) + \varepsilon$$

5.2.1 Model Results and Analysis

The results of the GAM are displayed in Table 4 and the nonparametric component smoothing functions on the scale of the linear predictor are displayed in Figure 7.

Table 4. Parametric and Nonparametric Effects of Model 5.2

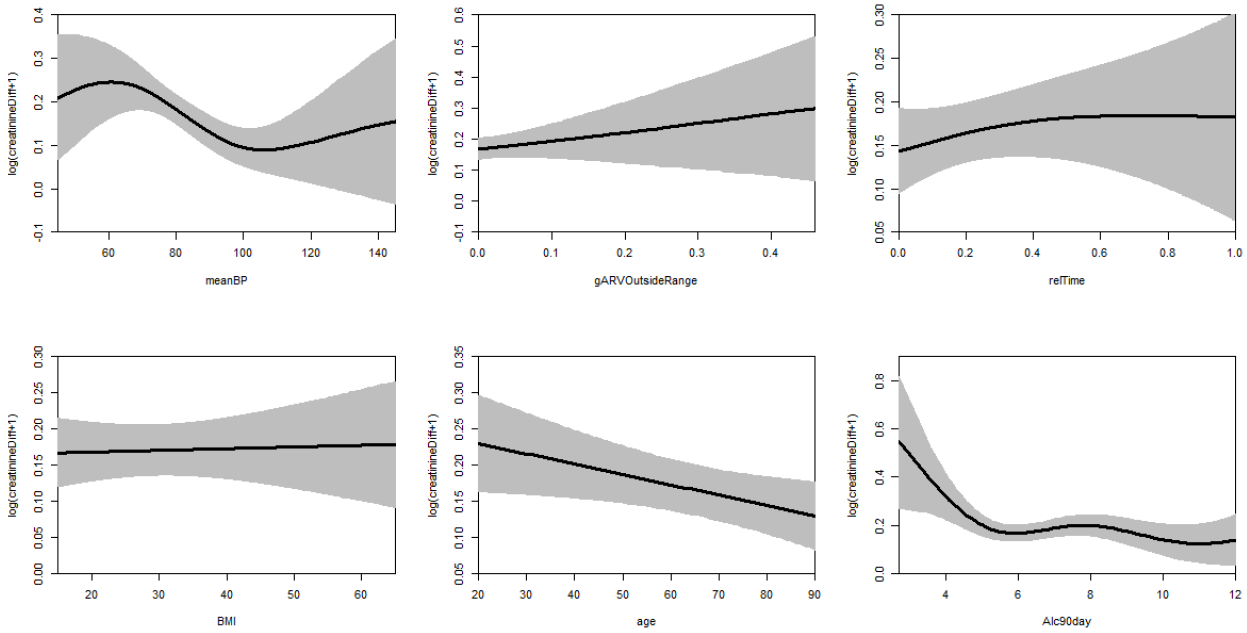
Sample size = 2477

Parametric Effects	Estimate	*t* value
Intercept	$1.8 * 10^{-1}$	$14.40 ***$
Gender	$1.5 * 10^{-2}$	0.84
Insulin	$1.3 * 10^{-1}$	$4.98 ***$

Nonparametric Effects	F statistic
s($meanBP$)	$4.49 ***$
s($gARVOutsideRange$)	1.13
s($relTime$)	0.50
s(BMI)	0.04
s(age)	$4.85 *$
s($Alc90day$)	1.89

$*$ p-value < 0.05, $**$ p-value < 0.01, $***$ p-value < 0.001

Figure 7. Plots of the Component Smooth Functions of Model 5.2



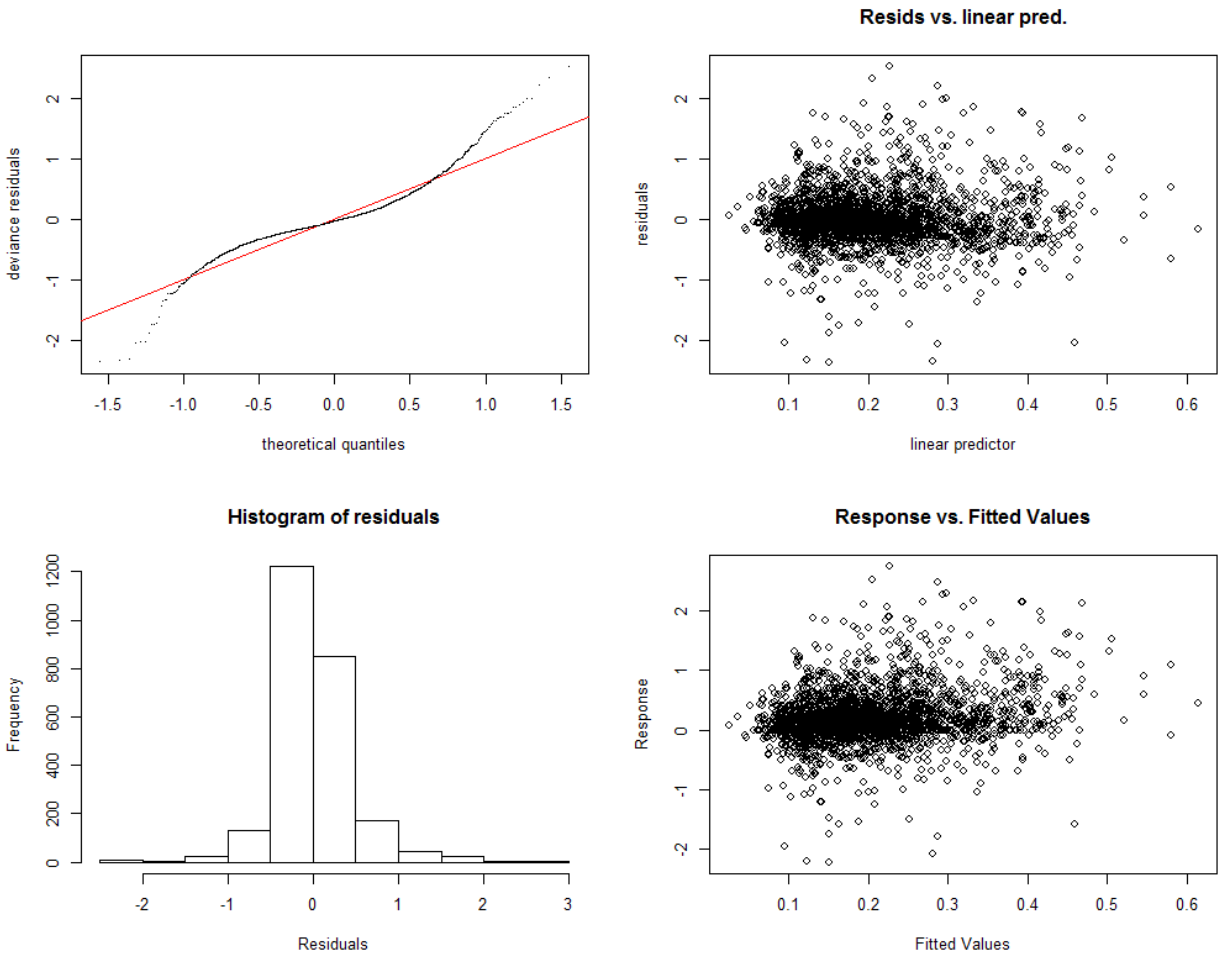
We see a similar but stronger relationship from our previous model for lower $creatinineDiff$ values when $meanBP$ is near common or average values. The $gARVOutsideRange$ component smooth function appears to be positive, but no significant nonparametric effects were found. This model was also rerun with $gARVOutsideRange$ inputted as only a parametric effect; however, again no significance was found.

Additionally, we can see that $A1c90day$ outside of common ranges seems to result in higher $creatinineDiff$'s, although this component smooth function too was not found to have a significant nonparametric effect. Age appears to have significant nonparametric effects, where an increase in age from the 1st quartile (age = 53) to the 3rd quartile (age = 71) results in a decrease in $log(creatinineDiff + 1)$ from 0.19 to 0.17, or equivalently a decrease in $creatinineDiff$ of 2 percentage points from 0.21 to 0.19, which is not clinically significant.

5.2.2 Model Diagnostics

We also examine this model to determine if our transformations have improved the issues with Model 5.1 by examining model diagnostics in Figure 8.

Figure 8. Model 5.2 Diagnostic Plots



As we can observe from the QQ and residual plots, the distribution now appears to uphold normality better than before, although the distribution is heavy tailed. Additionally, the residual

plot appears to be more randomly distributed about $y = 0$ and the response vs. fitted values scatterplot appears to adhere closer to the $y = x$ line with relatively equal variance.

This model overall upholds assumptions much better; however, it no longer captures the signal that the previous model was able to find and does not provide us with clinically significant effects in percent change of creatinine.

6. Future Directions

The methods discussed thus far were not fully able to capture the data appropriately. This section considers some alternatives still to be explored.

6.1 Quantile Regression

Generalized additive models attempt to estimate $E(Y_i)$, but approaches like quantile regression allow us to estimate the median or other quantiles of Y_i. Quantile regression is more robust to outliers and considering the skew in our data, it may be more flexible and appropriate in capturing variation in a dataset that is more heterogeneous than we initially assumed. By looking at quantiles, we can gain a broader picture of the relationship between our predictors and the response.

Quantile regression considers the relationship between the conditional quantile of the response and the covariates. The τ^{th} quantile of y can be written as

$$Q_\tau(y|x) = F^{-1}(\tau|x)$$

where q is the quantile, y is the dependent variable, x is the independent variable, and $F^{-1}(\tau|x)$ is the conditional inverse CDF function.

A linear quantile regression model can be written as

$$Q_\tau(y|x) = \boldsymbol{x}^\tau \boldsymbol{\beta}(\tau)$$

where $\beta(\tau) = \Big(\beta_0(\tau), \dots, \beta_p(\tau) \Big)^T$ is the quantile coefficient that denotes the marginal change in the τ^{th} quantile of y resulting from a marginal change in x.

However, in order to continue evaluating nonparametric components, we can also consider nonparametric quantile regression, which can be written as

$$Q_\tau(y|x) = g_\tau(x)$$

where $g_\tau(x)$ is a smoothing function which could be approximated by a linear combination of spline basis functions. A formula interface for nonparametric quantile functions can also be found in the ′rqss′ R package.

6.2 Robust Regression

We can consider robust regression to help capture errors that are not fully normally distributed, which is an issue that we have seen in our model. Replacing the assumption of normal distribution on the errors with a heavier-tailed t-distribution is one way of performing robust regression that could potentially address our issues with heavy tailed errors. Nonparametric Bayesian approaches can also be considered, for instance using a mixture of normal distributions to more flexibly model the error. Alternatively, a contaminated normal distribution where residuals are assumed to follow a mixture of two normal distributions is a simpler alternative.

6.3 Extrema Weighted Features

In all our current models, we focus on using some measure of generalized average realized variability to model blood pressure variability. However, as we have seen, utilizing $gARV$ in the approaches that we have explored have not resulted in highly promising results that are clinically significant in models that uphold assumptions.

Exploring other features that could replace $gARV$ as a more robust measure of blood pressure variability in time series data with emphasis on deviations outside of healthy ranges can be explored. We are looking at extrema weighted features as a novel way of measuring variability in functions like blood pressure that allows for adaptive weighting of any predefined local measures of variability in the extremes or tails.

7. Conclusion

Duke electronic health records provide a rich set of data, much of which is currently underutilized. Using a combination of different datasets, we explored confounders and extracted generalized average variability of mean arterial pressure curves to determine the relationship of blood pressure variability and renal outcomes. Specifically, we chose to consider $gARV$ for ranges that are outside what is considered healthy and utilize the percent change in creatinine as a measure for renal outcomes. We utilized generalized additive models to determine marginal relationships between covariates and the response, and found an increasing, clinically significant smoothing function for $gARV$ outside the healthy range in estimating the percent difference in creatinine, with the strongest relationship in an area of few data points. However, this signal was lost in the log-transformed model that adhered more closely to model assumptions, and it is difficult to confirm whether signal does exist in the $gARVOutsideRange$ features that would help to predict an increase in $creatinineDiff$.

Because of this, $gARVOutsideRange$ is potentially not a sufficient measure of blood pressure variability, which is complex time series data with potentially greater effects in extreme ranges, and we may need to explore more advanced methods to gain enough evidence that blood pressure variability intraoperatively is or is not an important effect in renal outcomes.

References

[1] Sega et al. Blood pressure variability and organ damage in a general population. *Hypertension* 2002;39:710-4

[2] Mancia et al. Blood pressure variability. *Pathophysiology of Hypertension* 1997; 117–169.

[3] Parati G. Blood pressure variability: its measurement and significance in hypertension. *J Hypertens* 2005; 23 (Suppl 1):S19–S25.

[4] Sun et al. Association of intraoperative hypotension with acute kidney injury after elective Noncardiac Surgery. *Anesthesiology* 2015;213:515-23

[5] Buck, Carl J. 2014 ICD-9-CM for Hospitals, Volumes 1, 2 and 3. Saunders, 2013.

[5] Kurella M, Lo JC, Chertow GM 2005 Metabolic syndrome and the risk for chronic kidney disease among nondiabetic adults. *J Am Soc Nephrol* 16:2134–2140.

[7] Parati et al. Relationship of 24-hour blood pressure mean and variability to severity of target-organ damage in hypertension. *J Hypertens* 1987; 5:93-8

[8] Reich et al. Intraoperative tachycardia and hypertension are independently associated with adverse outcome in noncardiac surgery of long duration. *Anesth Analg 2002*; 95:273-7

[9] Monk et al. Anesthetic management and one-year mortality after noncardiac surgery. *Anesth Analg* 2005; 100:4-10

[10] Hansen et al. Pressure in relation to cardiovascular outcomes investigators: prognostic value of reading-to-reading blood pressure variability over 24 hours in 8938 subjects from 11 populations. *Hypertension* 2010; 55:1049–57

[11] Aronson et al. Anesthesiology research endeavors: intraoperative systolic blood pressure variability predicts 30-day mortality in aortocoronary bypass surgery patients. *Anesthesiology* 2010; 113:305–12

[12] Aronson et al. Does perioperative systolic blood pressure variability predict mortality after cardiac surgery? An exploratory analysis of the ECLIPSE trials. *Anesth Analg* 2011; 113:19–30

[13] Palmer, B. Disturbances in renal autoregulation and the susceptibility to hypertension-induced chronic kidney disease. *American Journal of the Medical Sciences* 2004;330-343

[14] Hastie T, Tibshirani R: Generalized additive models. *Stat Sci* 1986; 1:297- 318.

[15] Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society* 2011: 73(1):3-36

[16] Brienza N, Giglio MT, Marucci M. Preventing acute kidney injury after noncardiac surgery. *Curr Opin Crit Care* 2010;16:353–8

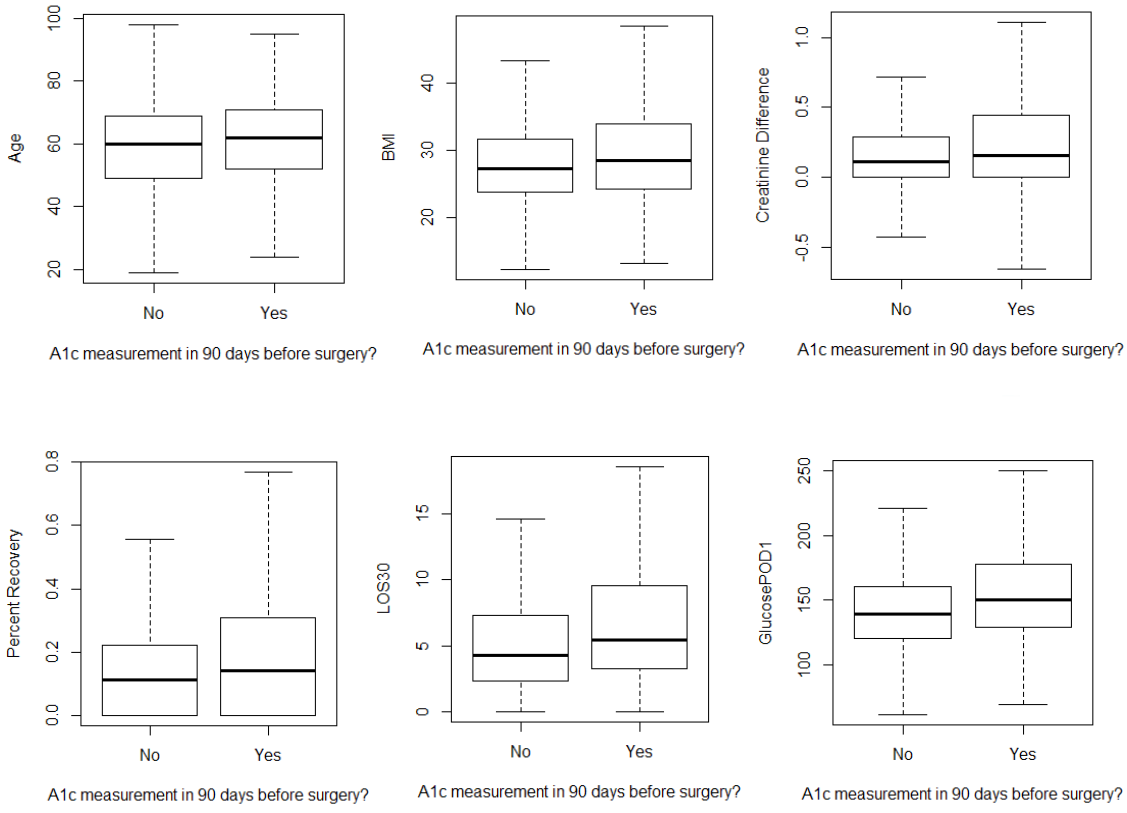
[17] Ricci, Z. *et al.* Classification and staging of acute kidney injury: beyond the RIFLE and AKIN criteria. *Nat. Rev. Nephrol.* 2011

Appendix A. Comparison of Patients with and without A1c Measurements

We found that approximately 44.78% of cases with A1c90day measurements were female and 51.01% of cases without Alc90day measurements were female.

We found statistically significant differences between each of the comparisons listed in Figure A.1 using one-sided t-tests.

Figure A.1 Comparison of Cases with and without $A1c90day$ Measurements



Appendix B. Variability Above/Below Healthy Ranges on Patients with A1c Measurements

Below is a model containing features that measure variability above the healthy range separately from below the healthy range. Since we now utilize $relTimeAbove$ and $relTimeBelow$, we no longer include $relTime$ as in Section 5 (since $relTime$ would simply become a linear combination of $relTimeAbove$ and $relTimeBelow$).

We fit the following GAM with confounders

$$creatinineDiff = \beta_0 + s(meanBPAbove) + s(meanBPBelow) + s(gARVAboveRange) + s(gARVBelowRange) + s(relTimeAbove) + s(relTimeBelow) + s(BMI) + s(age) + gender + insulin + s(A1c90day) + \varepsilon$$

The results of the GAM are displayed in Table B.1 and the nonparametric component smoothing functions on the scale of the linear predictor are displayed in Figure B.1.

Table B.1 Parametric and Nonparametric Effects of Model B

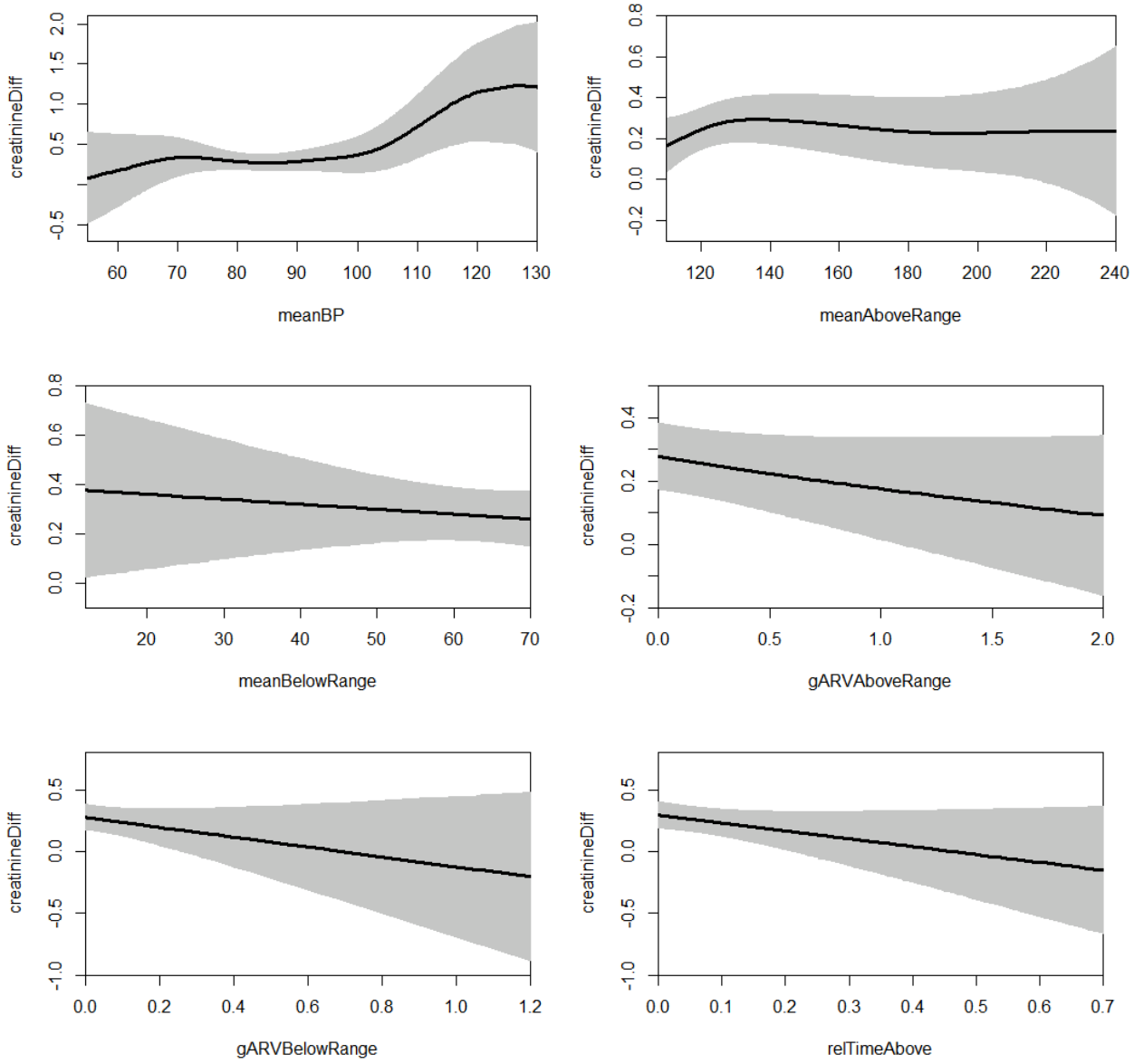
Sample size = 1781

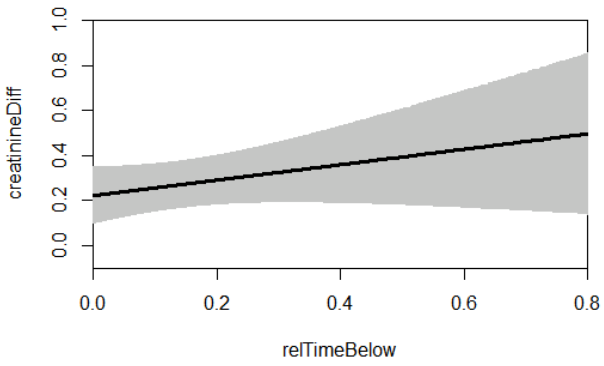
Parametric Effects	Estimate	*t* value
Intercept	$3.1 * 10^{-1}$	10.90 ***
Gender	$2.3 * 10^{-2}$	0.57
Insulin	$2.1 * 10^{-1}$	3.68 ***

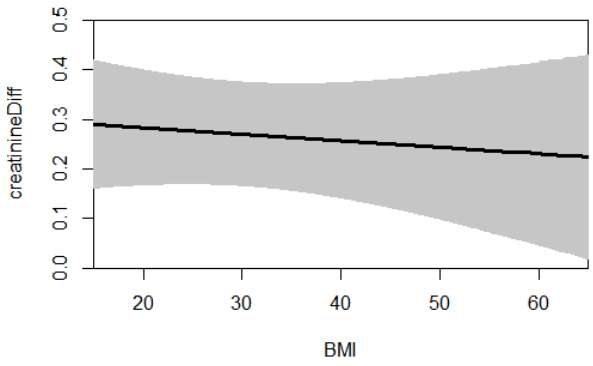
Nonparametric Effects	F
s(***meanBP***)	1.75
***s*(*meanAboveRange*)**	0.83
***s*(*meanBelowRange*)**	0.36
***s*(*gARVAboveRange*)**	1.68
***s*(*gARVBelowRange*)**	1.88
s(***relTimeAbove***)	2.59
s(***relTimeBelow***)	1.70
s(***BMI***)	0.26
s(***age***)	3.90 ***
s(***A1c90day***)	1.01

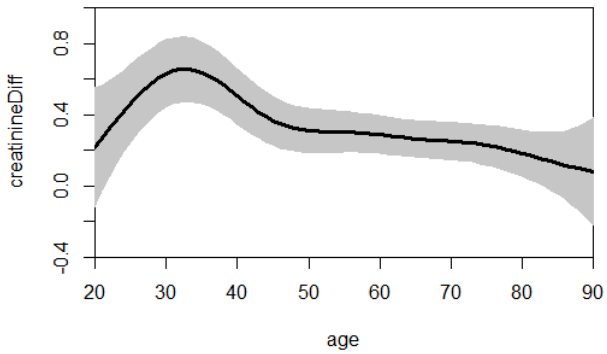
* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001

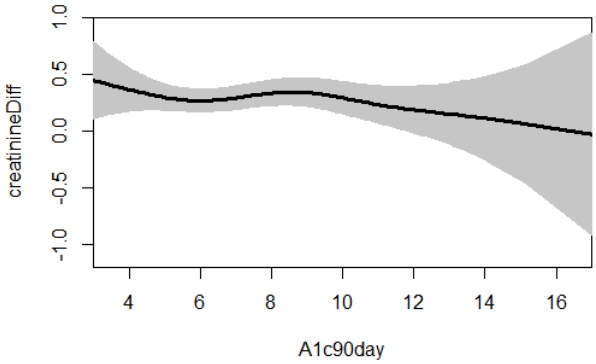
Figure B.1 Plots of the Component Smooth Functions of Model B




creatinineDiff
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
relTimeBelow


creatinineDiff
0.0
0.1
0.2
0.3
0.4
0.5
20
30
40
50
60
BMI


creatinineDiff
-0.4
0.0
0.4
0.8
20
30
40
50
60
70
80
90
age


creatinineDiff
-1.0
-0.5
0.0
0.5
1.0
4
6
8
10
12
14
16
A1c90day

Appendix C. Example Blood Pressure Curves of Cases with *gARVOutsideRange* > 0.5

Figure C.1 Blood Pressure Curves of Cases with $gARVOutsideRange > 0.5$