YaleNUSCollege

SpO_2/FiO_2 Ratio (SF Ratio) As a Predictor of Mortality in ICU Patients: Retrospective Study Using MIMIC III Database.

Ahmed Gobba

Capstone Final Report for BSc (Honours) in

Mathematical, Computational and Statistical Sciences

Supervised by: Willem van den Boom

AY 2019/2020

YaleNUSCollege

Contents

Chapter 1

Clinical Background

1.1 Importance of Biomarkers in ICU Studies

Allocation of resources to patients to minimize mortality is a constant priority for healthcare professionals. This is especially important in the area of healthcare we have chosen to focus on in this paper: critical care, where resources such as equipment and attention of specialists are scarce. A critical care specialist focuses on the most vulnerable and urgent patients who are placed in an Intensive Care Unit (ICU). In this setting, the specialist is often faced with a difficult decision of which patients to allocate resources to. An unfortunate reminder that has recently put such a predicament under the spotlight is the COVID-19 global pandemic. A recent study on ICU capacity in Wuhan, the disease's epicentre in China, states that at a point during the current crisis the number of COVID-19 patients who need ICU resources is 1120. However, only 600 ICU beds existed. As a result, only 25% of the patients who had died by the time of the study received the intubation and mechanical ventilation that they required (Wu and McGoogan, 2020). In Lombardy, the disease's epicentre in Italy, a study states that under pre-crisis conditions, the city's total

ICU capacity of 720 already operates at 85% - 90% occupancy during winter months. To make things worse, during the first two weeks after the city's first confirmed COVID-19 case, the number of COVID-19 related ICU admissions rose exponentially to 556. Moreover, the exponential estimate for the total number of COVID-19 ICU admissions in the following two weeks was 14,452 - an even more drastic disparity between supply and demand of ICU resources (Grasselli, Pesenti, and Cecconi, 2020). Hence, ICU capacities are stressed under normal conditions and even further during times of crises. Faced with such a perpetual dilemma of resource allocation, a critical care specialist uses various biomarkers (physiological indicators) in order to try and predict severe outcomes in the patients cohort. As such, a persistent priority in medical research is the discovery and analysis of connections between various biomarkers and unfavourable patient outcomes (Ware, 2017).

1.2 Problem: PF Ratio is an Important but Challenging Biomarker

One such biomarker that is tracked in ICU settings is the PaO_2/FiO_2 ratio (PF ratio). The numerator, PaO_2 refers to the partial pressure of oxygen in arterial blood. It is measured in mmHg via drawing a sample of blood from an artery in the wrist or groin, and analysing it in the laboratory. The denominator, FiO_2 refers to the initial fraction of inspired oxygen and is approximately 21% in breathable atmospheric air. The PF ratio is used to monitor the patient's pulmonary functions. In an ICU, it can be controlled by providing the patient with oxygen concentrations above

21% using devices such as a mechanical ventilator. It is most notably used in the diagnosis of fatal illnesses such as Acute Respiratory Distress Syndrome (ARDS) (Bernard et al., 1994). Moreover, it has been shown to be a significant identifier of mortality risk in the general ICU population (Villar et al., 2011) as well as in specific subsets of patients such as newborns with Meconium Aspiration Syndrome (MAS) (Narayanan et al., 2019) and post-operation cardiac surgery patients (Esteve et al., 2014).

Despite the various merits of the PF ratio in diagnosing patients and maintaining their stability, there is a major challenge associated with its use - the measurement of its numerator, PaO_2 . The procedure is invasive and delayed, making it not feasible to track at frequent intervals for all patients.

1.3 Motivation: Can SF Ratio be an Alternative to PF Ratio and for who?

A biomarker that is more convenient to measure than PaO_2 is SpO_2 or peripheral capillary oxygen saturation. It is defined as the ratio of oxygenated haemoglobin to the total amount of haemoglobin in the blood. It is measured using a pulse oximeter, which uses that oxygenated and deoxygenated haemoglobin absorb and radiate particular wavelengths of light to different extents (Jubran, 1999) . The oximeter illuminates light at specific wavelengths through the skin (usually at the fingertips) and almost instantaneously calculates the ratio of absorption of these wavelengths to extrapolate the proportion of oxygenated haemoglobin in the

blood, or SpO_2 (Jubran, 2015). Therefore, unlike PaO_2 , SpO_2 can be measured in a non-invasive and instantaneous manner.

Using SpO_2 instead of PaO_2 in the calculation of the PF ratio gives a different biomarker ratio - SpO_2/FiO_2 ratio (SF ratio). Although the SF ratio might seem as an intuitive replacement to the PaO_2/FiO_2 ratio , a critical difference between PaO_2 and SpO_2 is that the former is generally a more accurate measure of a patient's oxygenation level; it is measured directly from a main artery while the latter is measured at the end of capillaries.

Nevertheless, several studies have linked SpO_2 to mortality in patients with certain conditions. For instance, the Tromsø study in 2015 concluded that an $SpO_2 \leq 95\%$ is associated with all-cause mortality and mortality caused by pulmonary diseases after adjusting for sex, age, history of smoking, self-reported diseases and respiratory symptoms, BMI, and CRP concentration (Vold et al., 2015). Another study established SpO_2 as a predictor of mortality in patients with systemic sclerosis (Swigris et al., 2009).

Similarly, SF ratio, which includes SpO_2 , is linked to mortality in certain patient populations. For instance, SpO_2/FiO_2 Time-at-Risk (SF-TAR), defined as the total time spent with severe hypoxemia (SF ratio $\leq$ 145), is significantly correlated with hospital mortality for mechanically ventilated patients, and is as good or a better predictor of it than arterial gas-derived measurements of the PF ratio. Moreover, SF ratio is a non-invasive surrogate for the PF ratio to diagnose certain patient populations such as children with ALI or ARDS (Rice et al., 2007) and children with smoke inhalation injury (Cambiaso-Daniel et al., 2017). Hence, both

SF ratio and SpO_2 have been shown to be significant predictors of mortality in certain subpopulation.

However, not all patient subpopulation exhibit a significant link between SpO_2 or SF ratio and mortality. For example, a prospectively planned meta-analysis study using participant data from 5 randomized clinical trials (conducted from 2005-2014) of infants born before 28 weeks' gestation period found no significant difference between a lower SpO_2 target range (85%-89%) and a higher SpO_2 target range (91%-95%) on mortality or major disability at a corrected age of 18 to 24 months (Askie et al., 2018). Therefore, it seems that the use of SpO_2 or SF ratio as a predictor of mortality might not be applicable to all patient phenotypes, with potential for further sub-phenotyping.

1.4 Research Statement

The main goal of this capstone can be summarized as follows:
Investigate whether SpO_2/FiO_2 ratio is a predictor of mortality in general ICU patient population, over what range, and for which subsets of the population using a retrospective analysis of ICU patient records.

By focusing on the SF ratio instead of only SpO_2 , we also investigate whether the former is a more helpful predictor as it theoretically allows us to account for the different levels of mechanical ventilation that an ICU patient receives. In essence, we believe it allows us to account for the patient's ability to convert inspired oxygen to peripheral oxygen saturation at the tissue level.

Chapter 2

Data

2.1 Data Overview

For this capstone we used **MIMIC III**, an openly available relational database developed by the MIT Lab for Computational Physiology. It contains de-identified data of 61,532 intensive care unit stays: 53,432 stays for adult patients and 8,100 for neonatal patients at the Beth Israel Deaconess Medical Center over the June 2001 - October 2012. It includes demographics, vital signs, laboratory tests, medications, mortality, etc. The database is divided into different tables of data that contain information about a patient's stay and are linked to each via identifiers such as a unique hospital admission ID and a unique patient ID.

For specific details on the data extraction process for our analysis, refer to Appendix B.

Chapter 3

Methods: How to model effect of Biomarkers on ICU mortality?

3.1 Analysis Goal

To reiterate our goal, we first want to test for a correlation between SF ratio and patient outcome for the general ICU population and subsets of it. Next and perhaps more importantly, we want to examine over what ranges of SF ratio does significance hold if present. Therefore, we need to find a method of modelling that allows us to examine for more than mere significance of the biomarker.

3.2 The Problem with Generalized Linear Models: They're Linear

When we think of binary response variables such as mortality, we intuitively think of a linear logistic regression model. It belongs to a family of models called Generalized Linear Models (GLM). The word 'Linear' in

the name does not stand for the relationship between the response variable and the predictor being a straight line, but rather to the fact that the predictor or a function of is modelled by a linear combination of the covariates. The general form of a GLM with m covariates is:

$$g[\mathrm{E}(Y)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m$$

where g is called the link function linking the expected value of response variable, Y with a linear combination of the covariates, $\beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m$ (Wood, 2017). That is, the model only allows for the response variable to be connected to the covariates in a linear manner. Since our goal is to explore an unknown relationship between mortality and SF ratio, we cannot introduce bias into our modelling by assuming this linear relationship.

3.3 Generalized Additive Models

3.3.1 Overview

Another family of models that allows us to explore a potentially non-linear relationship is the family of Generalized Additive Models (GAM). A GAM is an extension of a generalized linear model; it also involves a linear combination, but allows for smoothing functions applied to the covariates (Hastie, 2017). A general structure for a GAM can be:

$$g\left(\mu_i\right) = \mathbf{A}_i \boldsymbol{\theta} + f_1\left(x_{1i}\right) + f_2\left(x_{2i}\right) + f_3\left(x_{3i}, x_{4i}\right) + \ldots$$

where $\mu_i = \mathbb{E}(Y_i)$ and Y_i is a response variable that belongs to an exponential family distribution with mean μ_i. $\mathbf{A}_i$ is a row of the model

matrix for any model components that are strictly parametric, and $\boldsymbol{\theta}$ is the parameter vector of those components. Functions $f_1, f_2, f_3, \ldots$ are smoothing functions for covariates $x_1, x_2, x_3, \ldots$ (Wood, 2017).

3.3.2 Additional Requirements for Using GAM

The use of smoothing functions comes with the need to specify two additional model properties - how to represent the smooth functions and how to control the smoothing.

The representation of the smoothing function can be done by choosing a basis that defines the space of functions that our smoothing functions belongs to. In general, a function $f(x)$ can be represented as the summation of a basis functions $b_j(x)$ as follows:

$$f(x) = \sum_{j=1}^{k} b_j(x)\beta_j$$

where β_j are unknown parameters (Wood, 2017). There are various choices for the basis of a smoothing function and the, the choice defines how the smoothing takes place. A more in depth discussion on the choice of bases and their shortcomings can be found in *Generalized additive models: an introduction with R*.

The second requirement we need to specify is the degree of smoothing. We don't want to overfit or underfit the data, but adequately represent the true underlying relationship between the response variable and covariates. To do this, we add a penalty term to the to the least squares fitting function. That is, instead of minimizing

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2,$$

where $\mathbf{y}$ is our response variable vector, $\mathbf{X}$ is our covariate vector and $\boldsymbol{\beta}$ is our vector of coefficients, we minimize,

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \int f''(x)^2 dx$$

where λ is the smoothing parameter chosen by us and the second derivative of the smoothing function expresses the "wiggliness" of the plotted line. If the the data is overfit and the line has a lot of curves, the second derivative will be higher and the penalty higher and the opposite is also true.

The final step that we need to decide on is the choice of λ. If lambda is too high then the line will be over-smoothed (straighter) and if it is too low the line will be undersmoothed (more curved).

The general method to select λ is through cross-validation. The principle of cross-validation arises from that we cannot choose a model based on its prediction performance of data it was fitted from. Instead, in cross-validation, we fit the model to a section of the data and test its prediction performance on the rest of the data.

One example of cross-validation is called *ordindary cross-validation* and involves fitting the model to all the data points in the response but one. Let y_i be the left out data point and $\hat{f}^{[-i]}$ be the model fitted to all the data points except for y_i. Next, we calculate the squared difference between the left out data point, y_i and its prediction by fitted model $\hat{f}^{[-i]}(x_i)$. This

step is done for all the data points and the squared differences are averaged out to get the ordinary cross validation score (Wood, 2017),

$$\mathcal{V}_o = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_i^{[-i]} - y_i\right)^2.$$

We choose λ to minimize the average prediction error $\mathcal{V}_o$. More cross-validation methods can be found in *Generalized additive models: an introduction with R*.

Chapter 4

Covariates

4.1 Covariates to Control for Differences in Population

To analyze the relationship between the covariate we are concerned with and the response variable, we need to control for differences within the diverse patient cohort we have chosen for the study. Hence, in any models fitted, we control for age, gender, Body Mass Index (BMI) and maximum sequential organ failure assessment score (SOFA score). SOFA score is a score assigned to the patient to determine the extent of organ function and possibility of organ failure based on six different scores for the respiratory, cardiovascular, hepatic, coagulation, renal and neurological systems. The SOFA score allows us to control for the difference in the level of sickness between patients which affects their final outcome. More specifically, we use the maximum SOFA score of a patient as it is a good quantifier of organ dysfunction present at ICU admission (Moreno et al., 1999).

We also note that for any controlling variable that is non-binary, there is not reason to assume that it's relationship with the response variable is

linear. Therefore, for all controlling covariates except for gender (binary) we add a smoothing function in our GAM.

4.2 SpO_2 vs SpO_2/FiO_2: to Ratio or not to Ratio

One aspect of our research question is to investigate whether the SF ratio is a better predictor of mortality than SpO_2 . Our hypothesis is that the SF ratio captures a a critical difference between patients - the level of oxygenation denoted by FiO_2 .

The best way to examine this is to plot the relationships between each of SpO_2 , FiO_2 and SF ratio and the probability of patient mortality as predicted by each of them. This would allow us to compare whether the trends captured by each model are the same or whether one models captures more or less of a trend. However, this direct comparison of plotted trends is not directly possible as all three metrics are on different scales, and thereby cannot be plotted on the same axis. To overcome this problem, we choose to instead plot the percentile of each of the metrics within the patient cohort vs the probability of hospital mortality and its 95% confidence interval.

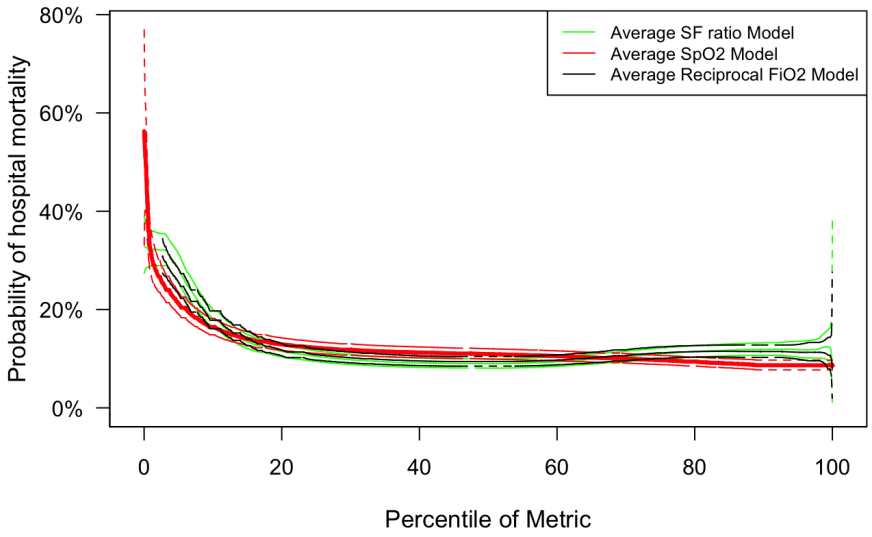

FIGURE 4.1: Model Comparison: Average SF ratio vs Average SpO_2 vs Average FiO_2

All three metrics are significantly correlated to patient outcome. However, as we can see from figure 4.2 above, the SF ratio model captures both the initial down trend of the SpO2 model and some of the upward trend towards the end FiO2 model. Therefore, we conclude that the use of the SF ratio is better as it captures part of the unique trends of both the SpO_2 and FiO_2 .

4.3 Does Transformation help?

In using the SF ratio however, there are two concerns. First of all, the normal range of FiO_2 is much wider than that of SpO_2 . FiO_2 ranges from 21% to 100% based on the level of oxygenation provided to the patient, while SpO_2 only ranges from 92% - 100% (Lapum et al., 2018). In taking a ratio of these two, we are concerned that the SF ratio will be more

representative of FiO_2 than SpO_2 . The second concern we have is the effect that extreme values in both SpO_2 and FiO_2 might have in increasing outlier values in the SF ratio.

To check whether these concerns are indeed an issue, we consider two different transformations of the SF ratio that tackle the two concerns respectively. We test if using them shows a difference in the trend that may not have been captured by the untransformed SF ratio. The two transformations are:

1. **Linear rescaling**: We rescale both the SpO_2 (numerator) and FiO_2 (denominator) to the same scale from 1 to 2 before taking the ratio and multiplying by a 100. We shall call this the Linear Rescaled SF ratio.

2. **Removing Extremes**: We remove the first and last percentile of both the SpO_2 and FiO_2 before taking the ratio we shall call this No-Extreme SF ratio.

We again plot the percentile of each metric to the probability of mortality and its 95% confidence interval as predicted by each model.

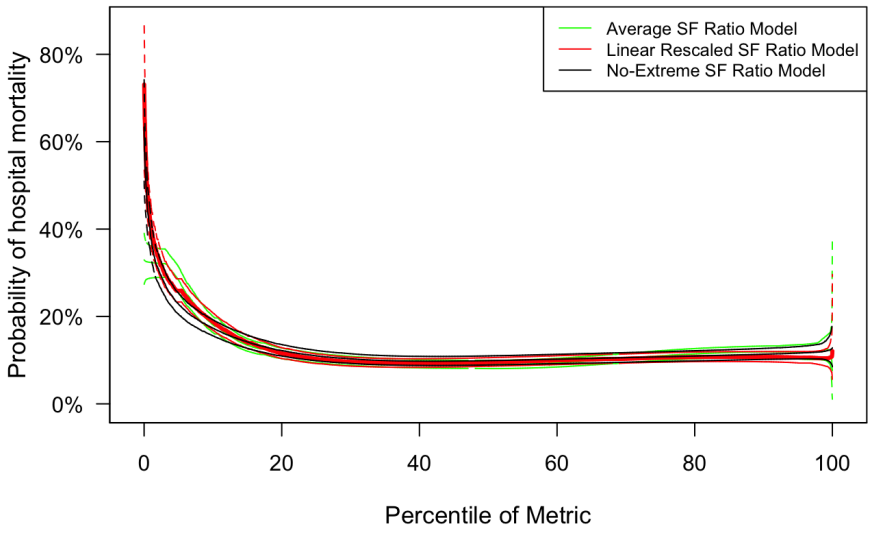

FIGURE 4.2: Model Comparison: Average SF ratio vs Linear Rescaled SF ratio vs No-Extreme SF ratio

As the plot above suggests, all three models almost exactly coincide in predicting the confidence interval of the response variable. Therefore, it does not seem that the two suggested transformations of the SF ratio add benefit. Consequently, we will continue using the SF ratio in our models.

4.4 Timeframe of SF Ratio aggregation

Chapter 5

Analysis and Findings

Bibliography

Askie, Lisa M et al. (2018). "Association between oxygen saturation targeting and death or disability in extremely preterm infants in the neonatal oxygenation prospective meta-analysis collaboration". In: *Jama* 319.21, pp. 2190–2201.

Bernard, Gordon R et al. (1994). "The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination." In: *American journal of respiratory and critical care medicine* 149.3, pp. 818–824.

Cambiaso-Daniel, Janos et al. (2017). "Correlation Between PaO2/FIO2 and Peripheral Capillary Oxygenation/FIO2 in Burned Children With Smoke Inhalation Injury." In: *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 18.10, e472–e476.

Esteve, Francisco et al. (2014). "Evaluation of the PaO2/FiO2 ratio after cardiac surgery as a predictor of outcome during hospital stay". In: *BMC anesthesiology* 14.1, p. 83.

Grasselli, Giacomo, Antonio Pesenti, and Maurizio Cecconi (2020). "Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response". In: *JAMA*.

Hastie, Trevor J (2017). "Generalized additive models". In: *Statistical models in S*. Routledge, pp. 249–307.

Jubran, Amal (1999). "Pulse oximetry". In: *Critical care* 3.2, R11.

— (2015). "Pulse oximetry". In: *Critical Care* 19.1, p. 272.

Lapum, Jennifer L et al. (2018). "Vital Sign Measurement Across the Lifespan-1st Canadian edition". In:

Moreno, Raúl et al. (1999). "The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study". In: *Intensive care medicine* 25.7, pp. 686–696.

Narayanan, Anand et al. (2019). "PaO2/FiO2 Ratio as Predictor of Mortality in Neonates with Meconium Aspiration Syndrome". In: *American journal of perinatology* 36.06, pp. 609–614.

Rice, Todd W et al. (2007). "Comparison of the SpO2/FIO2 ratio and the PaO2/FIO2 ratio in patients with acute lung injury or ARDS". In: *Chest* 132.2, pp. 410–417.

Swigris, JJ et al. (2009). "Exercise peripheral oxygen saturation (SpO2) accurately reflects arterial oxygen saturation (SaO2) and predicts mortality in systemic sclerosis". In: *Thorax* 64.7, pp. 626–630.

Villar, Jesús et al. (2011). "A risk tertiles model for predicting mortality in patients with acute respiratory distress syndrome: age, plateau pressure, and PaO2/FiO2 at ARDS onset can predict mortality". In: *Respiratory care* 56.4, pp. 420–428.

Vold, Monica Linea et al. (2015). "Low oxygen saturation and mortality in an adult cohort: the Tromsø study". In: *BMC pulmonary medicine* 15.1, p. 9.

Ware, Lorraine B (2017). *Biomarkers in critical illness: new insights and challenges for the future*.

Wood, Simon N (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

Wu, Z and JM McGoogan (2020). "Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention." In: *JAMA*.

Appendix A

Other Figures

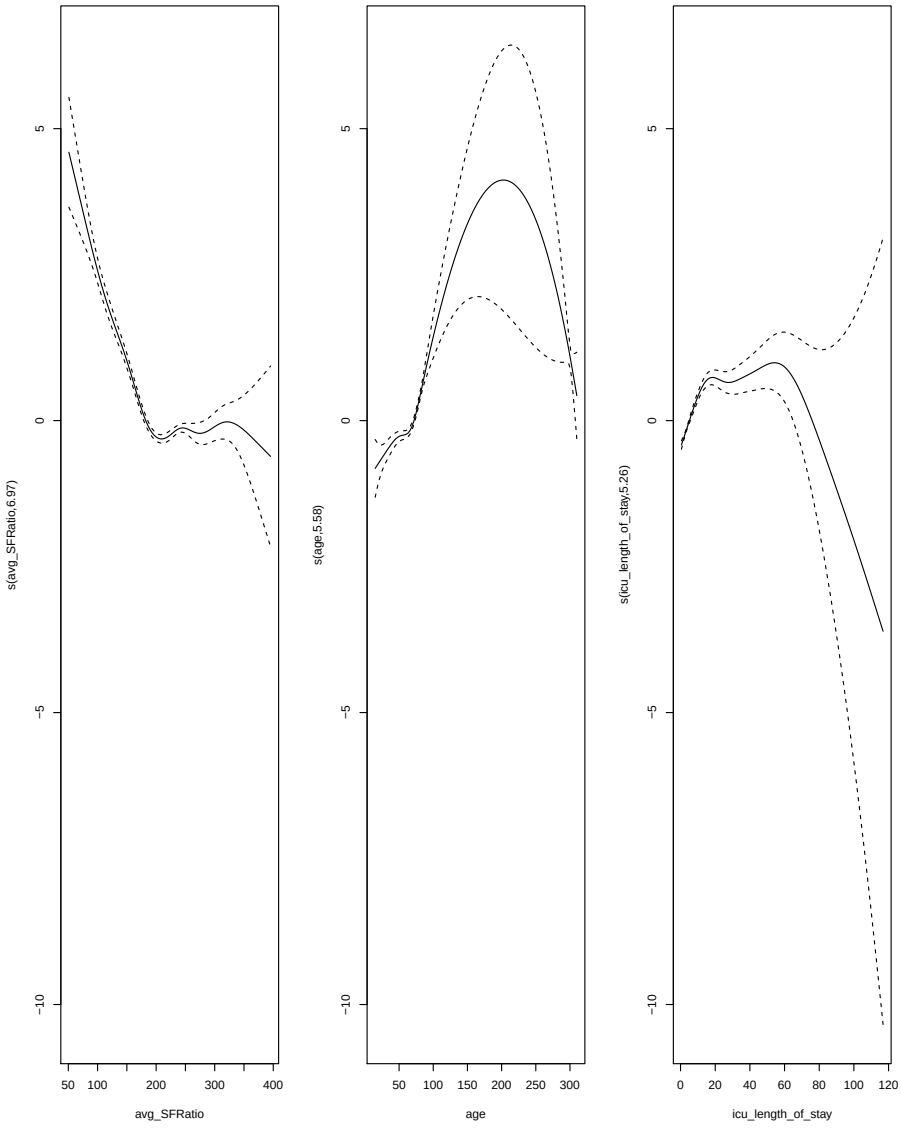

FIGURE A.1: Results of GAM model

Appendix B

Data Extraction

For the patient and stay identifiers we extract SUBJECT_ID, HADM_ID from `ADMISSIONS` table and ICUSTAY_ID from `ICUSTAYS` table. From the `ADMISSIONS` table we extract the time of death of the patient if applicable and if it lies between the ICU admission time and ICU discharge time in `ICUSTAYS` table we indicate ICU mortality. Similarly if the time of death is between admission time and discharge time in the `ADMISSIONS` table, we indicate Hospital mortality. From the `Patients` table we extract the gender of the patients' and calculate their age.

For all patients and ICU stays we extract the FiO_2 values and their chart times from the `CHARTEVENTS` table. Keeping in mind that at normal atmospheric conditions, FiO_2 is around 21%, we apply the following transformations. For the values between 0 and 1, we convert them to percentages by multiplying by a 100 and only keep those between 21% and 100%. Next, if the reading is recorded as greater than 1 but lower than 21, the value is likely to be erroneous and we discard it. Next, if the value is between 21 and 100, the value is likely to already be a percentage and we take it as such. Finally, we discard all the values above a 100 that are remaining. From the same `CHARTEVENTS` table we extract the patients'

height and weight.

From the `CHARTEVENTS` table, we also extract SpO_2 values and chart times but we only keep those which indicate 0 for ERROR which stands for error in measurement. Moreover, we filter the values and we discard those below 10 and above a 100 since they are either physiologically impossible or unlikely.

At the end of this stage, the current dataset accounts for 46,476 Patients, 61,532 ICU Stays with 12,713,362 observations of either SpO_2 or FiO_2 or both. On further examination of the data, we find that for every FiO_2 measurement for a given ICU stay, for a given unique patient, there is a corresponding SpO_2 measurement at the same chart time but not vice versa. Accordingly, we restrict my data to only those chart times with both SpO_2 and FiO_2 measurements. This further subsets the number of observations into 703,201 observations.