YaleNUSCollege

# $SpO_2/FiO_2$ ratio (SF ratio) as a predictor of mortality in ICU patients: Retrospective study using MIMIC 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

# Background Information

## 1.1 Introduction

Allocation of resources to critical care patients to minimize mortality is a priority for healthcare professionals. As such, the search for connections between various physiological indicators and unfavourable outcomes such as extreme illnesses or even mortality is a constant priority of researchers. One such pulmonary physiological indicator is the $PaO_2/FiO_2$ ratio (PF ratio). $PaO_2$ refers to the partial pressure of oxygen in arterial blood and is measured in mmHg via drawing a sample of blood from an artery in the wrist, that then is tested in a laboratory. $FiO_2$ refers to the initial fraction of inspired oxygen and is approximately 21% in breathable atmosphere and can be controlled with the use of mechanical ventilation. The PF ratio is importantly used in the diagnosis of extreme illnesses such as Acute Respiratory Distress Syndrome (ARDS) and has been shown to be predictor of mortality in specific subsets of patients such as newborns with Meconium aspiration syndrome (MAS) (Narayanan et al., 2019).

However, there are challenges associated with measurement of $PaO_2$

specifically. Most importantly, the procedure is invasive and is therefore not easy to measure for all patients and to track at frequent intervals. A more convenient biomarker to measure however is $SpO_2$ or peripheral capillary oxygen saturation, an estimate of the amount of oxygen in the blood. It is measured using pulse oximetry, a noninvasive method for monitoring a person's oxygen saturation. Moreover, $SpO_2/FiO_2$ ratio (SF ratio) has been shown to be a non-invasive surrogate for $PaO_2/FiO_2$ ratio to diagnose subsets of patients such as children with ALI or ARDS (Rice et al., 2007) and children with smoke inhalation injury (Cambiaso-Daniel et al., 2017).

A retrospective study found that the $SpO_2/FiO_2$ Time-at-Risk (SF-TAR), defined as the total time spent with severe hypoxemia (SF ratio $\leq$ 145), is not only significantly correlated with hospital mortality for mechanically ventilated patients, but is as well or a better predictor of it than arterial gas-derived measurements of the PF ratio.

Moreover, there have been several studies that aim to link $SpO_2$ to mortality. In 2015, the Tromsø study concluded that an $SpO_2 \leq 95\%$ is associated by all-cause mortality and mortality caused by pulmonary diseases (over a 10-year follow-up period) after adjusting for sex, age, history of smoking, self-reported diseases and respiratory symptoms, BMI, and CRP concentration. When Forced Expiratory Volume (FEV1) was included as a covariate, the correlation remained significant for mortality due to pulmonary diseases but no longer significant for all-cause mortality (Vold et al., 2015).

However, a prospectively planned meta-analysis 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 $SpO_2$ as a predictor of mortality might not be applicable to all patient phenotypes, with possible analysis of further sub-phenotyping. Such differences between the subpopulation might also be expected for the SF ratio which includes $SpO_2$ .

## 1.2 Aim and Objectives

The main goal of this Capstone can be summarized in the following statement:

*Investigate whether $SpO_2/FiO_2$ ratio is a statistically significant predictor of mortality in general ICU patient population or subsets thereof using a retrospective analysis of data.*

The use of $SpO_2/FiO_2$ ratio instead of solely $SpO_2$ allows us to account for the different levels of mechanical ventilation that an ICU patient receives. In essence, it allows us to account for the patient's ability to convert inspired oxygen to peripheral oxygen saturation at the tissue level.

## 1.3 Data

For this capstone I will be using **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 by identifiers such `HADM_ID` which refers to a unique hospital admission and `SUBJECT_ID` which refers to a unique patient.

# Chapter 2

# Methods

## 2.1 Pre-Analysis Data Preparation

The pre-analysis data preparation involved extraction of data from different tables, combining them and calculating the SF ratio. The following subsections describe these processes.

### 2.1.1 Data Extraction

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

For all patients and icu stays I 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%, I apply the following transformations. For the values between 0 and 1, I 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 I discard it. Next, if the value is between 21 and 100, the value is likely to already be a percentage and I take it as such. Finally, I discard all the values above a 100 that are remaining. From the same CHARTEVENTS table I extract the patients' height and weight.

From the CHARTEVENTS table, I also extract $SpO_2$ values and chart times but I only keep those which indicate 0 for ERROR which stands for error in measurement. Moreover, I filter the values and I 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, I 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 I restrict my data to only those chart times with both $SpO_2$ and $FiO_2$ measurements. This further subsets the number of observations further into 703,201 observations.

### 2.1.2 Dealing with duplicate $SpO_2$ values

On examination of the data I find that for 63,563 of the 703,201 observations (approx. 9%) there exists another $SpO_2$ measurement with the same

ICUSTAY_ID and chart time. To deal with this, I decided to remove any of these observations in which the two $SpO_2$ values differed by 5 or more (percentage scale) and took the average of the two values for the remainder of the observations. At the end of this stage, I have data for 16,113 Patients, 17,737 ICU Stays with 636,203 observations with both $SpO_2$ and $FiO_2$ values. I then calculate the SF ratio as $SpO_2 \ / FiO_2 \ \times 100$.

# Chapter 3

# Initial Analysis and Results

## 3.1 Logistic Regression

I first compute the average SF ratio per ICU stay. Next, I fit the following logistic regression :

$$\ln\left(\frac{M}{1-M}\right) = \beta_0 + \beta_1(\text{AverageSFRatio}) + \beta_2(\text{Gender}) + \beta_3(\text{Age})$$

where $M$ is Hospital Mortality. The results are as follows:

| **Feature** | **Estimate Std.** | **Error** | **z value** | **Pr(>\|z\|)** |
|---|---|---|---|---|
| (Intercept) | 0.5738878 | 0.1150207 | 4.989 | 6.06E-07 |
| avg_SFRatio | -0.0120058 | 0.0005277 | -22.752 | <2.00E-16 |
| genderM | -0.2428841 | 0.0421135 | -5.767 | 8.05E-09 |
| age | 0.0054761 | 0.0003447 | 15.889 | <2.00E-16 |

TABLE 3.1: Results of logistic regression

Hence, the Average SF ratio is significantly correlated with Hospital Mortality and a unit increase in SF ratio decreases odds of hospital mortality by 1.19%.

## 3.2 Generalized Additive Model

A logistic regression assumes a linear relationship between mortality and the different features which may not necessarily be true. To explore a potential non-linear relationship I use a generalised additive model. A generalized additive model is an extension of a generalized linear model with a linear predictor involving a sum of smooth functions of covariates (Hastie, 2017). The general model can be expressed as follows (Wood, 2017):

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

where $Y_i \sim \text{EF}(\mu_i, \phi), Y_i$ is the response variable, $\mu_i \equiv \mathbb{E}(Y_i)$ and $\text{EF}(\mu_i, \phi)$ denotes an exponential distribution with mean $\mu_i$ and parameter, $\phi$. Also, $f_j$ are smooth functions of the covariates, $x_k$ (Wood, 2017).

Now fitting a GAM model to our data with Hospital Mortality as response variable and a smoothening function applied to predictors Average SF Ratio, Age and Length of ICU Stay and Gender taken as a linear predictor. We obtain the following results:

| **Feature** | **edf** | **Ref.df** | **Chi.sq** | **p-value** |
|---|---|---|---|---|
| s(avg_SFRatio) | 6.968 | 8.078 | 957 | <2e-16 |
| s(age) | 5.582 | 6.665 | 411.2 | <2e-16 |
| s(icu_length_of_stay) | 5.262 | 6.268 | 306.8 | <2e-16 |

TABLE 3.2: Results of GAM

The result of the model in table 3.2 shows that there is a statistically significant correlation between Hospital Mortality and Average SF Ratio.

I visualize this relationship below in fig 3.1. The visualizations for all the predictors can be found in Appendix A.

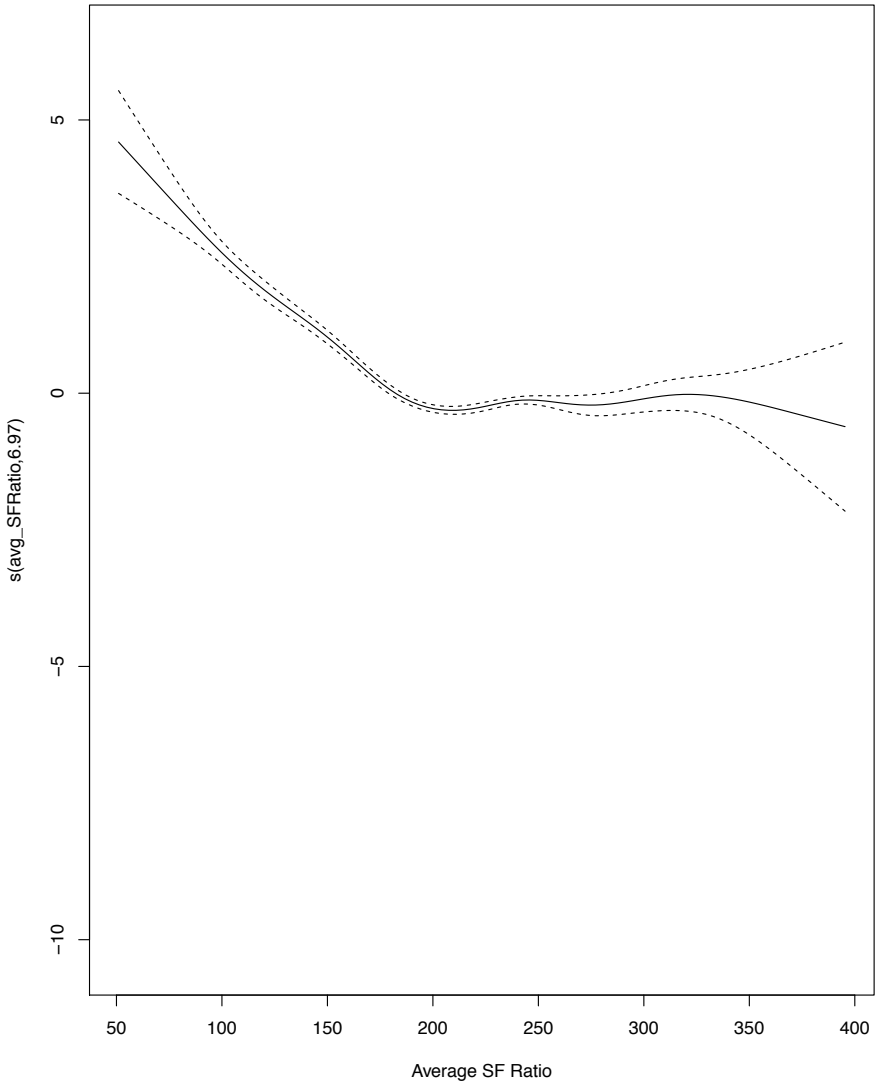


FIGURE 3.1: Results of GAM model for relationship between Average SF Ratio and log odds of Hospital Mortality

From fig 3.1 we see that a increase in Average SF Ratio from around 100 to around 200 causes a reduction in the log odds of Hospital Mortality with minimal standard error.

# 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.

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.

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

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.

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.

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

# Appendix A

# Other Figures

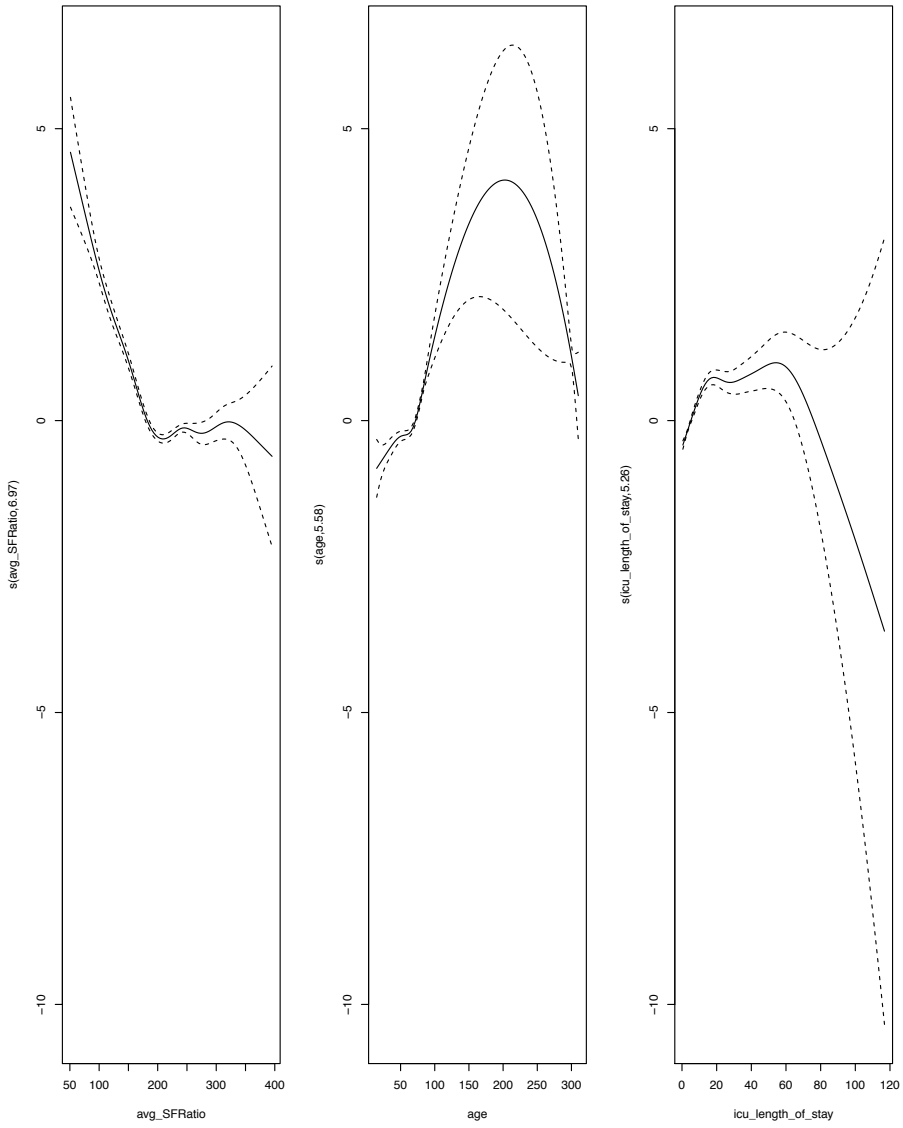


FIGURE A.1: Results of GAM model