Yale-NUS College

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

Yale NUS College

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 even more 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 difficult decisions of which patients to allocate resources to. An unfortunate reminder that has recently put such decisions under the spotlight is the current 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 while 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 (**wu2020**). 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% 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, estimates for the total number of COVID-19 ICU admissions in the following two weeks suggest an even more drastic shortage with linear estimates at 869 and exponential estimates at 14,542 (**Grasselli2020**). 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 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.

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). It is used to monitor the patient's pulmonary functions. 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 testing it in the laboratory. The denominator, FiO_2 refers to the initial fraction of inspired oxygen and is approximately 21% in breathable atmospheric air. In an ICU, it can be controlled by providing the patient with a form of oxygen therapy through to the use of devices such as mechanical ventilators. The PF ratio is most notably used in the diagnosis of extreme illnesses such as Acute Respiratory Distress Syndrome (ARDS) (**bernard1994american**). Moreover, it has been shown to be a significant identifier of mortality risk the general ICU population (**villar2011risk**) and as a predictor of mortality in specific subsets of patients such as newborns with Meconium Aspiration Syndrome (MAS) (**narayanan2019pao2**) and post-operation cardiac surgery patients (**esteve2014evaluation**).

However, there is a major challenge associated with the use of PaO_2 - its measurement. The procedure is invasive and delayed, making it not feasible to track PaO_2 at frequent intervals measure for all patients.

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

A more convenient biomarker to measure, however, is SpO_2 or peripheral capillary oxygen saturation. It is defined as the ratio of oxygenated haemoglobin by the total amount of haemoglobin in the blood. It is measured using pulse oximetry which uses the principle that oxygenated and deoxygenated haemoglobin have different absorption spectra at particular wavelengths of light (**jubran1999pulse**) . 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 (**jubran2015**). Therefore, unlike PaO_2 , SpO_2 can be measured in a non-invasive and instantaneous manner.

However, a critical difference between, PaO_2 and SpO_2 is that the former a generally more accurate measure of a patient's oxygenation level since it is measured directly from a main artery while the latter is measured at the end of capillaries. Nonetheless, recent studies have shown that SpO_2/FiO_2 ratio (SF ratio) has been shown to be a non-invasive surrogate for PaO_2/FiO_2 ratio to diagnose certain subsets of patients such as children with ALI or ARDS (**rice2007comparison**) and children with smoke inhalation injury (**cambiaso2017correlation**). Moreover, 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 ≤ 145), is not only significantly correlated with hospital mortality for mechanically ventilated patients, but is as good or a better predictor of it than arterial gas-derived measurements of the PF ratio.

There have also been several studies that aim to link SpO_2 separately to mortality. In 2015, the Tromsø study concluded that an $SpO_2 \leq 95\%$ is associated with 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 (**vold2015low**).

However, not all studies investigating a link between SpO_2 or

SpO_2/FiO_2 ratio and mortality have yielded significant results. 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 (**askie2018association**). Therefore, it seems that the use of SpO_2 as a predictor of mortality might not be applicable to all patient phenotypes, with potential for further sub-phenotyping. Such differences between the subpopulation might also be expected for the SF ratio which includes SpO_2 in the numerator.

1.4 Research Question

The main goal of this capstone can be summarized as follows:
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 ICU patient records.

By focusing on SpO_2/FiO_2 ratio instead of only SpO_2 , we also implicitly hypothesize that the former is a more helpful predictor as it 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.

1.5 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 via identifiers such as a unique hospital admission ID and a unique patient ID.

This is the end of the chapter.

Chapter 2

Methods: How to model ICU mortality?

2.1 General Additive Models

2.2 G-Computation

2.3 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.

2.3.1 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.

Chapter 3

Covariates

3.1 Covariate Selection

Ahmed Gobba

3.1.1 To Ratio or not to Ratio

Ahmed Gobba

3.1.2 Does Transformation help?

Ahmed Gobba

3.2 Timeframe of SF Ratio aggregation

Ahmed Gobba

Chapter 4

Analysis and Findings

4.1 Logistic Regression

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

$$\ln\left(\frac{M}{1-M}\right) = \beta_0 + \beta_1(\text{Average SF Ratio}) + \beta_2(\text{Gender}) + \beta_3(\text{Age}) + \beta_4(\text{BMI}) + \beta_5(\text{Sofa}$$

where M is Hospital Mortality, BMI is the Body Mass Index of the patient and Sofa Total Score is a . The results are as follows:

Feature	**Estimate**	**Std. Error**	**z-value**	**p-value**
Intercept	-0.3427213	0.2362569	-1.451	0.147
Average SF Ratio	-0.0097838	0.0008024	-12.193	<2e-16
Gender (M)	-0.3579267	0.0585947	-6.109	1.01e-09
Age	0.0045151	0.0005280	8.551	<2e-16
BMI	-0.0316707	0.0044837	-7.063	1.62e-12
Sofa Total Score	0.1964941	0.0083214	23.613	<2e-16

TABLE 4.1: 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%.

4.2 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 (**hastie2017generalized**). The general model can be expressed as follows (**wood2017generalized**):

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

where $Y_i \sim \mathrm{EF}(\mu_i, \phi)\,, Y_i$ is the response variable, $\mu_i \equiv \mathbb{E}(Y_i)$ and $\mathrm{EF}(\mu_i, \phi)$ denotes an exponential distribution with mean μ_i and parameter, ϕ. Also, f_j are smooth functions of the covariates, x_k (**wood2017generalized**).

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**
Average SF Ratio	6.968	8.078	957	<2e-16
Age	5.582	6.665	411.2	<2e-16
Length of ICU Stay	5.262	6.268	306.8	<2e-16

TABLE 4.2: Results of GAM

The result of the model in table 4.2 shows that there is a statistically significant correlation between Hospital Mortality and Average SF Ratio. I visualize this relationship below in fig 4.1. The visualizations for all the predictors can be found in Appendix A.

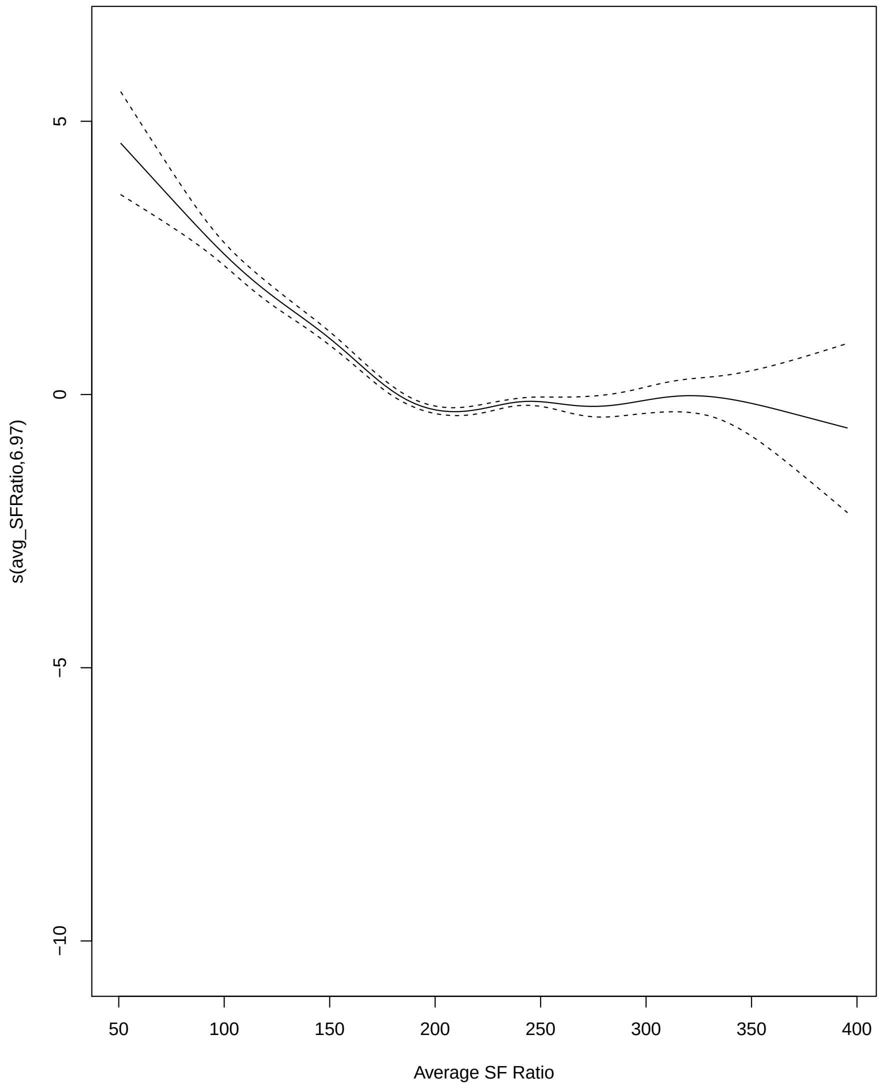

FIGURE 4.1: Results of GAM model for relationship between Average SF Ratio and log odds of Hospital Mortality

From fig 4.1 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.

Appendix A

Other Figures

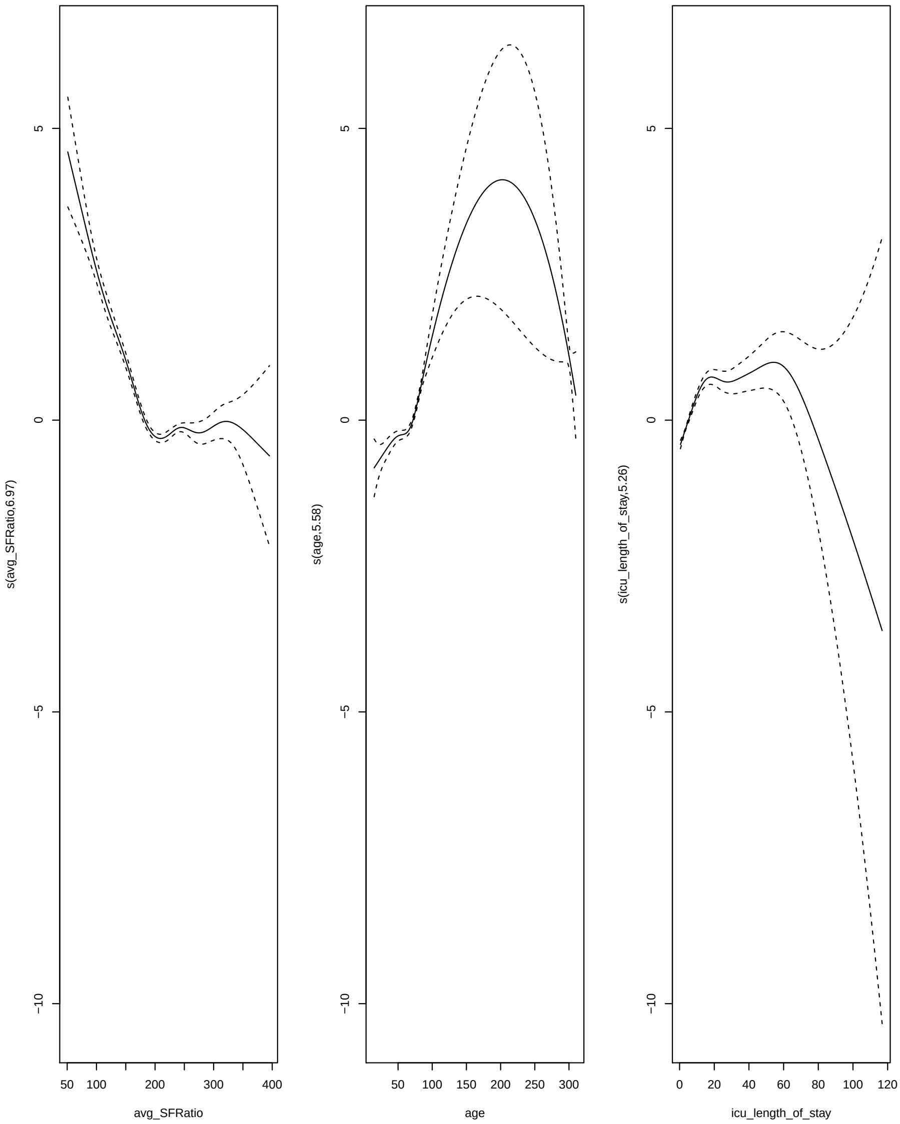

FIGURE A.1: Results of GAM model