Migration Letters

Volume: 21, No: S5 (2024), pp. 701-707

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Using Bayesian Truncated Regression Model To Predict Thrombosis For COVID-19 Patients

Hindreen Abdullah Taher [*a], Nawzad Muhammed Ahmed[b]

Abstract

As it is clear that corona-virus (covid-19) has a direct danger to humanity in the world, from the beginning of appearing this virus till seven months later that medical science was unable to understand the behavior of this virus, in our study we focused on the Thrombosis for 73 patients that have covid-19 as a response variable and Age, Sodium, Blood pressure as factors. We aimed to study the effect of these factors on the thrombosis for covid-19 patients, as it is clear that the D-Dimmer below 500 means the patient does not suffer from thrombosis but above 500 does suffer, in this situation for such a response of this type Bayesian truncated regression model is an appropriate model to be used to predict thrombosis of the patients that have covid-19, our contribution in this study is using Bayesian truncated regression model to predict the thrombosis for the first time. The study has shown that if a patient's age increases by one year, thrombosis increases by 0.17, also increasing one unit of Na and systolic causes an increase in thrombosis of 0.32 and 0.70, respectively. All three variables are statistically significant because their p-value is less than 0.001.

Keywords: *Covid-19, Bayesian inference, Truncated Regression, Linear Regression,Thrombosis.*

1 Introduction

The World Health Organization (WHO) termed this illness coronavirus disease 2019 (COVID-19). The coronavirus family have been shown to enter cells through binding angiotensin-converting enzyme 2 (ACE-2), found mainly on alveolar epithelium and endothelium. Activation of endothelial cells is thought to be the primary driver for the increasingly recognised complication of thrombosis. Viral inclusion bodies have been identified in endothelial cells in a variety of organs, from the lung to the gastrointestinal tract [1]. The immune dysregulation characteristic of severe COVID-19 infection may be initiated by "pyroptosis", a particularly pro-inflammatory form of apoptosis initially described in macrophages [2], with rapid viral replication leading to massive release of inflammatory mediators. One of the most consistent findings is that of a raised D-dimer level. Although many inflammatory processes can influence D-dimer levels, it almost certainly reflects, to some extent, intravascular thrombosis in patients with COVID-19 [3, 4]. In the early studies emerging from China, an elevated D-dimer level (>1000 ng·mL−1) at admission was

[a] Department of Information Technology College of Commerce Sulaimani University, Kurdistan
[b] Department of Statistics and Informatics College of Administration and Economics Sulaimani University, Kurdistan
[*] Corresponding author: hindreen.taher@univsul.edu.iq

associated with an increased risk of in-hospital death [5]. The true prevalence of thrombosis associated with COVID-19 infection is unknown, as most studies to date do not include systematic and comprehensive investigation protocols. We want to measure the impact of Age, Sodium, and Blood pressure on Thrombosis for those patients that suffer from covid-19 [19]. The purpose behind using the Bayesian truncated regression model is to get answers to the questions that are said each age, sodium, and blood pressure affects thrombosis and how much the amount of their effect.

2. Literature Review

Some of previous work on Bayesian modeling, Thrombosis, and COVID-19 will be presented in this section. [16] The researchers conducted a literature search in PubMed using MeSH headings “disseminated intravascular coagulation”, “pulmonary embolism”, “Thromb*”, “stroke”, “myocardial infarction” and “acute lung injury”, as well as terms “COVID-19”, “SARS-CoV-2”, “2019 novel coronavirus” and “2019-nCoV”. From their study they reached that the COVID-19 disease is characterized by the interactions between hyperactive coagulation and complement systems – induced by hyper-inflammatory conditions, resulting in a pro-thrombotic state and diffuse tissue injury. There are several promising prognostic markers of disease severity, with D-dimer the most significant. The presence of thrombocytopenia appears to be a key indicator of patient deterioration. [17] The researchers conducted a scoping review using a single-engine search for studies assessing thrombosis and coagulopathy in COVID-19 patients. Additional studies were identified by the secondary review and alert services. The studies reported the occurrence of venous thromboembolism and stroke in approximately 20% and 3% of patients, respectively. A higher frequency seems to be present in severely ill patients, in particular those admitted to intensive care units. The thrombotic risk is elevated despite the use of anticoagulant prophylaxis but optimal doses of anticoagulation are not yet defined. Although an increase of biomarkers such as D-dimer has been consistently reported in severely ill COVID-19, the optimal cut-off level and prognostic value are not known. [18] They showed that The prevalence of venous thromboembolic event (VTE) and arterial thromboembolic event (ATE) thromboembolic events in patients with COVID-19 remains largely unknown. They analyzed findings from 102 studies (64503 patients). In conclusion, they discovered that the patients admitted to the ICU for severe COVID-19 had a high risk of VTE. Conversely, further studies are needed to determine the specific effects of COVID-19 on the risk of ATE or VTE in less severe forms of the disease.

3 Methodology

3.1 Bayesian Truncated regression model

A typical regression model looks like the following

$$y \sim N(f(x, c), \sigma) \tag{3.1}$$

where y is the target variable, x is the predictor variable, c denotes coefficients of the model, and f(x) is the function for describing the dependency of the mean/median of y on x. The assumption behind the model in Eq. (3.1) is that y is distributed according to a normal distribution with mean f (x, c) and standard deviation , given a value of x. In a regression setting, the goal is to estimate the parameters of the model σ (comprising c,) from observed values of x and y. One important quantity is the likelihood function L(θ), which roughly speaking is the probability of observing the data, given a fixed value of all parameters of the model θ [10,6]. One can then maximize L(θ) (or rather the log-likelihood for numerical

reasons) concerning θ to find an estimate of θ (maximum-likelihood (ML) estimation), or estimate the parameters in a Bayesian setting (Spiegelhalter and Rice 2009). Here, under the Bayesian approach, the goal is to estimate the posterior distribution of the parameters given the data, which is proportional to the likelihood times the prior distribution of the parameters

$p(\theta\ |(x, y))\ \theta\ L(\theta\ )p(\theta\ )$

(3.2)

The likelihood for a continuous variable can be calculated from its density function

$L(\theta) = pdf\ (y|f\ (x, c), \sigma)$

(3.3)

where pdf is the probability density function (typically the probability density function of a normal distribution). Equation (3.3) describes the likelihood of a single observation (x, y). Typically there are multiple observations available (e.g. indexed from 1 to N), and one can work with the joint likelihood $L_N(\theta\ )$. If the observations are independent, then the likelihood factorizes, and the joint likelihood can be written as

$L_N(\theta\ ) = \prod_{k=1}^{N} L_k\ (\theta\ )$

(3.4)

Truncation means that only instances with $L < y < U$ are observed, where L is the lower bound, and U is the upper bound. For Thrombosis data, we typically deal with lower truncated data, and we restrict ourselves to this case (i.e., $L < y$). This means that the likelihood now is a truncated distribution, which can be easily calculated as

$$L(\theta\ ) = \begin{cases} \frac{pdf\ (y|f\ (x,c),\sigma)}{1-cdf\ (L\_f\ (x,c),\sigma)} & y > L \\ 0 & otherwise \end{cases}$$

(3.5)

where CDF is the cumulative distribution function. The adjustment to the likelihood ensures that the density integrates to one. One can think of the effect of truncation as the distribution of y, conditional on $y > L$. The denominator in Eq. (3.5) is the probability that $y > L$ [7,9]. Hence, for ground-motion data Eq. (3.5) describes the likelihood of an observation, conditional on it being recorded.

$L(\theta) = p(y|\ \theta, \sigma Be, \sigma Ss)$

$= pdf\ (y|\ \mu es, \phi SS)$

$= pdf\ (y|f\ (xes, c) + \sigma Be + \sigma Ss, \phi SS)$

(3.6)

Where σB is a unique event term for each event, and σS is a unique station term for each station (thus, σ Be is the event term for the eth event). ϕSS is the standard deviation of y, conditional on σB and σS and μes is the median prediction for event e at station s[8,11].

In a truncated multi-level Bayesian model, σB and σS are considered parameters of the model that are estimated (not predicted given the other coefficients and standard deviations[12,13], as in an ML-setting for mixed-effects models). The prior distribution for event terms σB is a normal distribution with mean zero and standard deviation τ; for station terms σS , the prior distribution is a normal distribution with mean zero and standard deviation ϕS2S. Typically, event terms and station terms are assumed to be independent, so the prior distribution for the event terms is $p(\sigma B) = \Pi NE\ i=1\ p(\sigma Bi\_0, \tau)$, and the prior distribution for station terms is $p(\tau S) = \Pi NS_{\ j=1}\ p(\sigma Sj\_0, \phi S2S)$, where NE is the number of observations[14], and NS is the number of stations. The other parameters to be estimated are the coefficients c and the standard deviations τ, ϕSS , and ϕS2S . We denote their prior distribution as $p(p(c, \tau, \phi S2S, \phi SS)$. Thus, combining the prior distributions with the factorized likelihood, we can write the posterior as

$p(\theta\ |D) \propto p(c, \tau, \phi S2S, \phi SS)\ p(\sigma B)\ p(\sigma S)L_N(\theta)$

$$\propto p(c, \tau, \phi S2S, \phi SS) \prod_{i=1}^{NE} p(\sigma Bi|0,\tau) \prod_{i=1}^{Ns} p(\sigma Sj|0,\phi S2S)) \prod_{k=1}^{N} L_k(\theta) \quad (3.7)$$

where D is the observed data set (target and predictor variables), N is the number of observations; and p(.) is a probability density function. Substituting the likelihood $L_k(\theta)$ with the explicit conditional distribution of the observation [15].

4 Application

4.1 Data Description

The data were collected through interviewing and monitoring 73 patients by the researchers from Shahid Aso Hospital in Sulemani governorate, the dataset contains variables Thrombosis as the response variable and Age, Sodium, and Blood pressure as explanatory variables.

4.2 Results

Table (1) represents the Bayesian estimated coefficient of the explanatory variables.

Variables	Coef.	Std.Err	z	P>|z|	[95% Conf. Interval]	
Age	0.170042	0.0477611	4.17	0.000	0.105415	0.2926349
Na	0.321296	0.0472668	7.26	0.000	0.2506546	0.4359371
systolic	0.701227	0.045942	15.92	0.000	0.6411909	0.8212801

The three explanatory variables have a positive impact on thrombosis for patients that have COVID-19, and they are all statistically significant because their p-value is less than 0.001. The impact is demonstrated by the fact that if a patient's age increases by one year, thrombosis increases by 0.17, and increasing one unit of Na and systolic causes an increase in thrombosis of 0.32 and 0.70, respectively.

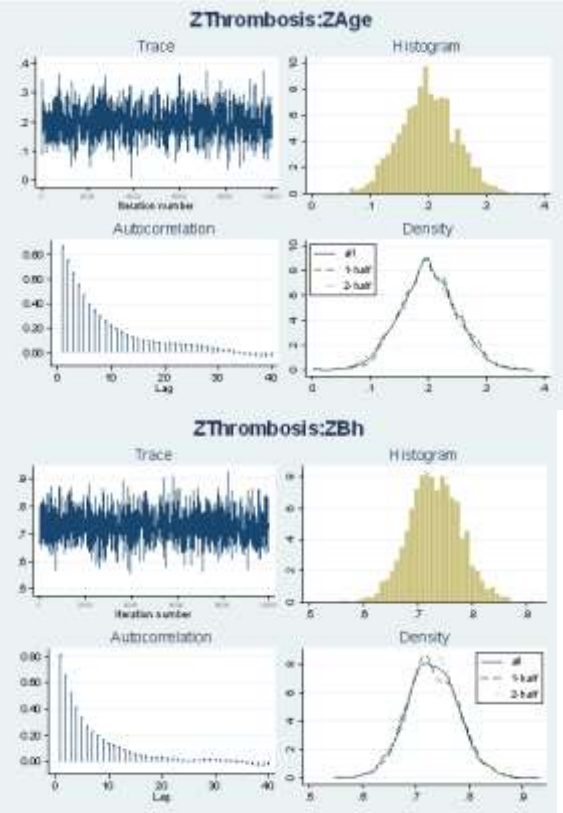
Table (2) shows the simulated sample information.

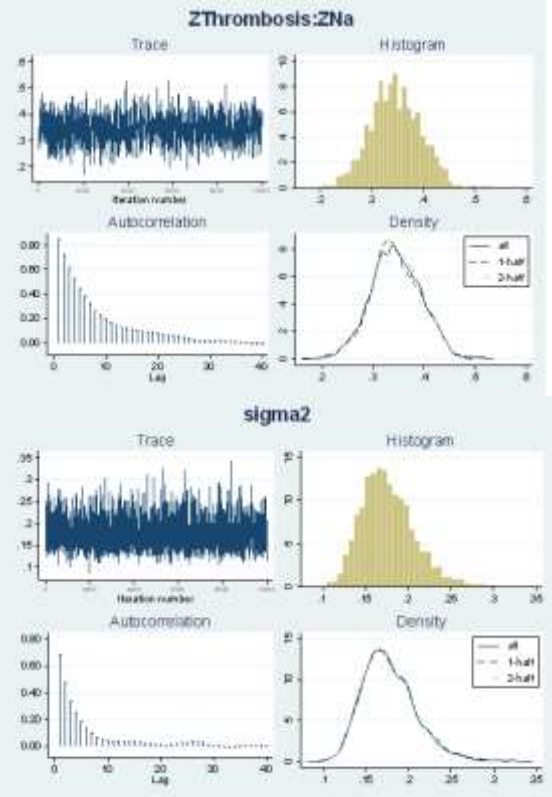
MCMC sample size =	10,000
Number of obs. =	73
Acceptance rate =	0.65400
Efficiency: min =	0.07487
Avg =	0.31970

Table (3) Demonstrate the credible interval of each simulated explanatory variables.

	Mean	Std.Dev.	MCSE	Median	[95% Cred. Interval]	
Age	0.199	0.049	0.016	0.187	0.101	0.295
Na	0.343	0.048	0.024	0.349	0.245	0.440
Systolic	0.831	0.044	0.028	0.842	0.825	0.970
sigma2	0.177	0.031	0.001	0.173	0.126	0.249

Sum to up the table the mean and the median are too close in their values which means the simulated samples are rather symmetric, as a piece of evidence the below figure (1) improves it.





5. Conclusions

In this study, we focused on Thrombosis for 73 patients that have covid-19 as a response variable and Age, Sodium, and Blood pressure as factors. The D-Dimmer below 500 means the patient does not suffer from thrombosis but above 500 does suffer. All three variables are statistically significant because their p-value is less than 0.001. The impact can be shown in a way that if the patient's age increased by one year the thrombosis will also increase by 0.17, and increasing one unit of Na and systolic leads to increasing in thrombosis by (0.32 and 0.70) respectively.

References:

[1] Varga, Z., A. Flammer, J. P. Steiger, M. Haberecker, R. Andermatt, A. S. Zinkernagel, and F. R. M. Holger. "Correspondence Endothelial cell infection an d endotheliitis in." The Lancet 6736, no. 20 (2020): 19-20.

[2] Cookson, B. T. "Brennan (2001)." Pro-inflammatory programmed cell death." Trends in Microbiology 9, no. 3: 113-114.

[3] Tirumani, Sree Harsha, Ata A. Rahnemai-Azar, Jonathan D. Pierce, Keval D. Parikh, Sooyoung S. Martin, Robert Gilkeson, and Nikhil H. Ramaiya. "Are asymptomatic gastrointestinal findings on imaging more common in COVID-19 infection? Study to determine frequency of abdominal findings of COVID-19 infection in patients with and without abdominal symptoms and in patients with chest-only CT scans." Abdominal Radiology 46, no. 6 (2021): 2407-2414.

[4] Malas, Mahmoud B., Isaac N. Naazie, Nadin Elsayed, Asma Mathlouthi, Rebecca Marmor, and Bryan Clary. "Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: a systematic review and meta-analysis." EClinicalMedicine 29 (2020): 100639.

[5] Milloy, M. J., and Evan Wood. "Withdrawal from methadone in US prisons: cruel and unusual?." The Lancet 386, no. 9991 (2015): 316-318.

[6] Ahmed, Dr Nawzad M. "Limited Dependent Variable Modelling (Truncated and censored Regression models) with Application." The Scientific Journal of Cihan University–Sulaimanyia PP 82: 96.
[7] Ahmed, Nawzad M., and Handren A. Taher. "Multi-response Regression Modeling for an Agricultural Experiment." Journal of University of Human Development 4, no. 2 (2018): 46-52.
[8] Ahmad, Nawzad M. "Bayesian Classifier for a Gaussian Distribution, Decision Surface Equation, with Application." Iraqi Journal of Statistical Science 18 (2010): 35-58.
[9] Rahim, Sham Azad, Shahla Othman Salih, Ayad Othman Hamdin, and Hindreen Abdullah Taher. "Predictions the GDP of Iraq by using Grey–Linear Regression Combined Model." The Scientific Journal of Cihan University–Sulaimaniya 4, no. 2 (2020): 130-139.
[10] Kuehn, Nicolas Martin, Tadahiro Kishida, Mohammad AlHamaydeh, Grigorios Lavrentiadis, and Yousef Bozorgnia. "A Bayesian model for truncated regression for the estimation of empirical ground-motion models." Bulletin of Earthquake Engineering 18, no. 14 (2020): 6149-6179.
[11] Taher, H. A., & Ahmed, N. M. (2023). Using Bayesian Regression Neural Networks Model to Predict Thrombosis for Covid-19 Patients. resmilitaris, 13(1), 2077-2087. [12] Fung, Tsz Chai, Andrei L. Badescu, and X. Sheldon Lin. "Fitting censored and truncated regression data using the mixture of experts models." North American Actuarial Journal (2021): 1-25.
[13] Puza, Borek D., Helen L. Johnson, Terence J. O'Neill, and Simon C. Barry. "Bayesian truncated Poisson regression with application to Dutch illegal immigrant data." Communications in Statistics—Simulation and Computation® 37, no. 8 (2008): 1565-1577.
[14] Puza, Borek D., Helen L. Johnson, Terence J. O'Neill, and Simon C. Barry. "Bayesian truncated Poisson regression with application to Dutch illegal immigrant data." Communications in Statistics—Simulation and Computation® 37, no. 8 (2008): 1565-1577.
[15] Marsh, Thomas L., and Ronald C. Mittelhammer. Truncated Regression in Empirical Estimation. No. 1841-2016-152301. 2000.
[16] Page, Eden M., and Robert AS Ariëns. "Mechanisms of thrombosis and cardiovascular complications in COVID-19." Thrombosis research 200 (2021): 1-8.
[17] Al-Ani, Fatimah, Samer Chehade, and Alejandro Lazo-Langner. "Thrombosis risk associated with COVID-19 infection. A scoping review." Thrombosis research 192 (2020): 152-160.
[18] Tan, Boun Kim, Sabine Mainbourg, Arnaud Friggeri, Laurent Bertoletti, Marion Douplat, Yesim Dargaud, Claire Grange, Hervé Lobbes, Steeve Provencher, and Jean-Christophe Lega. "Arterial and venous thromboembolism in COVID-19: a study-level meta-analysis." Thorax 76, no. 10 (2021): 970-979.
[19] Taher, H. A., & Ahmed, N. M. (2023). Using Bayesian Regression Neural Networks Model to Predict Thrombosis for Covid-19 Patients. resmilitaris, 13(1), 2077-2087.