Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2021;13(2):178-84

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2020-80264

COVID-19 Mortality Prediction: An Algorithm by Bayesian Autoregressive Model

COVID-19 Mortalite Tahmini: Bayesian Otoregresif Modele Göre Bir Algoritma

Gajendra K. VISHWAKARMA[a], Atanu BHATTACHARJEE[b,c], Sharvari SHUKLA[d]

[a]Department of Mathematics and Computing, Indian Institute of Technology Dhanbad, INDIA
[b]Department of Section of Biostatistics, Centre for Cancer Epidemiology, Tata Memorial Centre, Mumbai, INDIA
[c]Homi Bhabha National Institute, Mumbai, INDIA
[d]Symbiosis Statistical Institute, Symbiosis International University, Pune, INDIA

ABSTRACT Objective: This pandemic of COVID-19 is tedious to control. The only lockdown is the way to stop the spread of this infection. Conventional health care is facing a real challenge to operate. Primarily the challenge is to provide health care support for COVID-19 patients with limited resources and continue the health care services like earlier. Perhaps, this challenge is the same but the magnitude differing depending on geographical locations around the globe. **Material and Methods:**-In this article, we presented a Bayesian algorithm with the Code to predict cases, death and recovery numbers death due to COVID-19 in Indian population. This code is possible to run at different time points and different geographical locations around the world. This code will help us to get the best strategy to optimize the implementation of lockdown in different regions**. Results:** We applied the Bayesian auto-regressive model. The autoregressive model is separated into linear and non-linear models. The posterior estimates of the regression coefficients are obtained by the linear autoregressive model. The model is performed for confirmed, recovered and died cases separately and corresponding estimates are generated. **Conclusion:** The model would provide physicians with an objective tool for counselling and decision making at different hotspots and small areas to implement.

Keywords: COVID-19; population; decision; Bayesian, autoregressive model

ÖZET Amaç: COVID-19 pandemisinin kontrolü yorucu ve uğraştırıcıdır. Bu enfeksiyonun yayılmasını durdurmanın tek yolu karantinadır. Konvansiyonel sağlık bakımı gerçek bir zorlukla karşı karşıyadır. Öncelikle zorluk, sınırlı kaynaklara sahip COVID-19 hastaları için sağlık bakımı desteği sağlamak ve sağlık hizmetlerini daha önce olduğu gibi sürdürmektir. Belki de bu güçlük aynı ancak büyüklüğü dünyadaki coğrafi konumlara bağlı olarak farklılık gösterir. **Gereç ve Yöntemler**: Bu yazıda, Hindistan'da vakaları, COVID-19 nedeniyle olan ölümleri ve iyileşen hasta sayılarını tahmin etmek için ölüm Kodlu bir Bayes algoritması sunduk. Bu kodun farklı zaman noktalarında ve dünyadaki farklı coğrafi bölgelerde işlemesi mümkündür. **Bulgular**: Bayesian oto-regresif model uyguladık. Oto-regresif model linear ve non-linear modellere ayrılır. Regresyon katsayılarının posterior tahminleri linear oto-regresif modelden elde edildi. Model doğrulanmış, iyileşen ve ölen hastalar için ayrı ayrı uygulandı ve ilgili hesaplamalar oluşturuldu. **Sonuç:** Bu model doktorlara farklı sıcak noktalarda ve küçük alanlarda danışmanlık ve karar verme aracı sağlayacaktır.

Anahtar kelimeler: COVID-19; popülasyon; karar; Bayesian, otoregresif model

In this current coronavirus disease-2019 (COVID-19) pandemic, we are waking up to the limitations of their analog health care system. We need an immediate digital solution, pursued on several fronts, to address this crisis. It is crucial to implement predicting maps during this outbreak and attain all possible ways to ex-

Correspondence: GAJENDRA K VISHWAKARMA
Department of Mathematics and Computing, Indian Institute of Technology Dhanbad, INDIA
E-mail: vishwagk@rediffmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 27 Nov 2020 ***Received in revised form:*** 19 Feb 2021 ***Accepted:*** 24 Feb 2021 ***Available online:*** 15 Mar 2021

plore an estimation of transmission rate who infected whom, where, when and what the rate? But it is difficult to understand the transmission because it is not known always. We can find out when a particular host individual contacted the infection, and thereafter transmitted to others. It is required to uncover the transmission process case to cases to create dynamic models.

Lockdown is the only solution to stop the COVID-19 pandemic. We preferred to perform the Bayesian autoregressive modeling. A total of 28 days of India's COVID-19 confirmed and mortality data is considered. The autoregressive model is used to predict the next 50 days' projections on mortality and confirmed cases for the general population.

The univariate Bayesian approach is an already established method. It is possible to obtain parameter estimation by iterative generalized least square.[1,2] The most common approach is the generalize the conditional autoregressive distribution.[3] The multivariate approach is already explored.[4-7] The autoregressive approach is valid to predict the COVID-19 disease. We apply the autoregressive model in the COVID-19 India model.

The objective of this work is to predict the appearance of COVID-19 related death in Indian population. The data visualization technique is utilized to represent the gravity of the problem. The linear and non-linear autoregressive modeling is performed. The National level data available of COVID-19 confirmed and death counts are considered to predict the COVID-19 case count and death for the next 50 days. Similarly, the same model is utilized with prior information on national data to predict Indian population death and confirmed cases for the next 50 days. We provide the code that will help us to get the best strategy and useful to predict as time to time basis. The model would provide policy makers with an objective tool for decision making at different hotspots and small areas to implement. The model is easy to run in OpenBUG software to generate the prediction. The OpenBUG software is open source. It is possible to generate prediction with convergence with updated real time data.

STATISTICAL MODEL

LINEAR AUTOREGRESSIVE REGRESSION

A linear model is defined as

$$Y(t) = \beta_0 + \beta_1 X_1(t) + \beta_2 X_2(t) = \cdots + \beta_m X_m(t) + Z(t) \ (1)$$

It is defined as $Y(t)$ as the observation of the dependent variable at time t. Now $X_i(t)$ is presented as the i-th independent variable at time t, and finally

$Z(t)$ is the error term at time t. The errors $Z(t), t = 1,2 \dots .n$ term is assumed with mean zero, have a constant variance, and autocorrelated. Now objective is to compute the Bayesian inference for the m with regression coefficient β_i.The unknown parameters of the error process. It is assumed that $Z(t)$ to follow AR (1) with autocorrelation θ.

NON-LINEAR AUTOREGRESSIVE MODEL

However, the number of cases count is unknown in the future. It may progress by non-linear trends. The non-linear autoregressive modeling is defined as

$$Y(t) = \exp(\beta_0 + \beta_1 t) + Z(t) \ (2)$$

Similarly, the intercept is presented as β_0 and the regression parameter is β_1 are not known. The corresponding error for this model is estimated at $Z(t)$. The autocorrelation θ with AR (1) process is assumed.

The OpenBUG code is presented below. Suppose the observations are assumed as t. The Bayesian computation is performed with 28 observations i.e. 4 weeks count data.

It helped to estimates of β_0, β_1 , and σ^2 respectively.

A total of 30,000 iterations for simulation and 5,000 burn-in-was performed. The observation with outcome Y is assumed with a normal distribution. The variance-covariance matrix of AR (1) with correlation $\boldsymbol{\theta}$ is defined.

The posterior estimate is computed through simulation.

RESULTS

Indian confirmed, recovered and death count data are presented (Figure 1) for the last 28 days. It shows the real and overwhelming success of lockdown to retain the case count with limited numbers. Perhaps, death is contributed to a marginal proportion. The available data on https://www.mohfw.gov.in/ website is presented. It presents a comparison between the proportion of death and recovery among confirmed cases (Figure 2A). Similarly, a 10% higher rate of death in Indian population with similar recovery rates is presented (Figure 2B).

Further, we applied the Bayesian auto-regressive model. The autoregressive model is separated into linear and non-linear models. The linear model is defined in equation (1). Similarly, the non-linear model is presented in equation (2).

The posterior estimates of the regression coefficients are obtained by the linear autoregressive model. The model is performed for confirmed, recovered and died cases separately. The result of posterior estimates is presented in Table 1. A total of four parameters are estimated. The β_0 and β_1 represents the intercept and regression coefficients. The variance of the model is presented by V. the amount of autocorrelation is estimated by θ. A total of 30,000 iterations are performed to obtain convergence and estimates (Figure 3). Now the linear autoregressive model is defined as Model 1 (for confirmed cases), Model 2 (for died case) and Model 3 (recovered cases). The posterior regression estimates of all the parameters can be used for prediction of the case and death count for the next 50 days (Figure 4). The linear model simulation is presented in Figure 5. It shows to obtain the convergence for the parameters. The posterior estimates of the non-linear model are presented in Table 2. The OpenBUG code is presented below can be used to obtain the prediction.

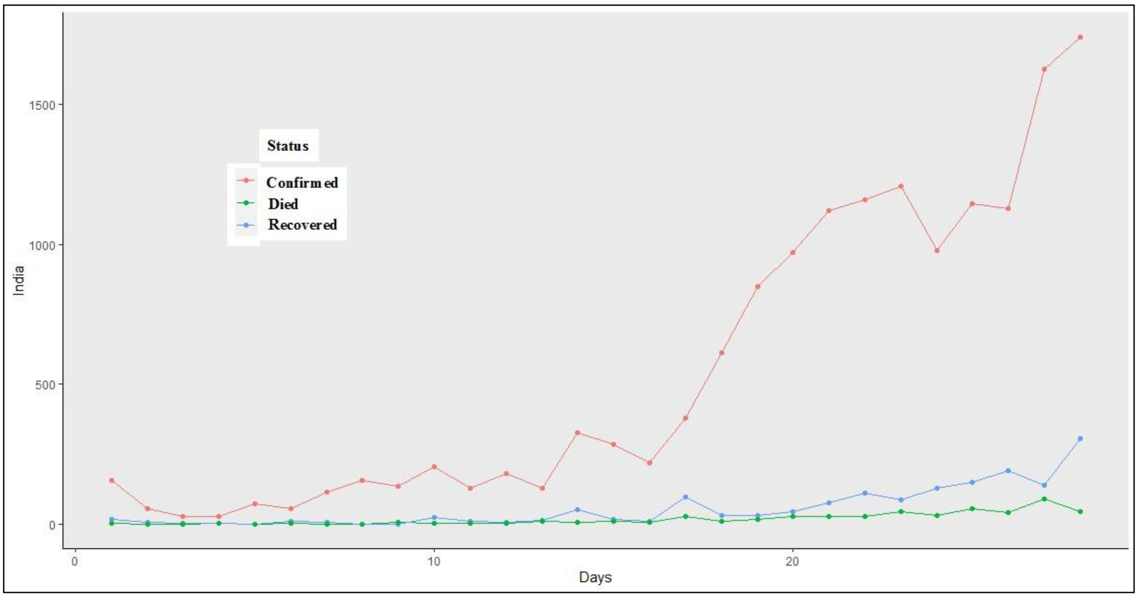

FIGURE 1: The comparison between the official COVID-19 pandemic data of confirmed, died and recovered cases in India.

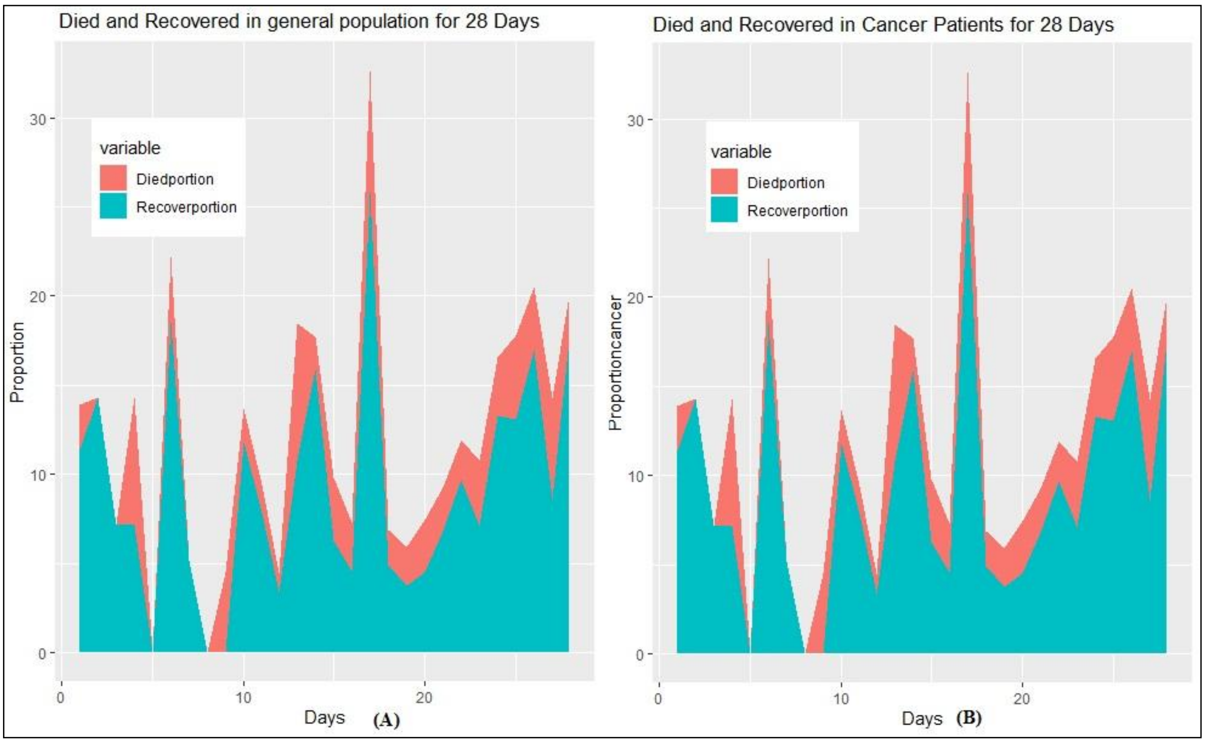

FIGURE 2: The comparison between died and recovered proportion in general population.

TABLE 1: Linear autoregressive posterior estimates of confirmed, deceased and recovered cases on prediction.

Parameter	Mean	SD	95% (HPD)
Indian states confirmed cases			
β_0	4.38	0.44	(3.56,5.11)
β_1	0.11	0.01	(0.08,0.14)
θ	0.92	0.04	(0.82,0.97)
v	3527	1164	(1873,6366)
Indian states deceased cases			
β_0	0.38	0.52	(-0.72,1.51)
β_1	0.13	0.02	(0.09,0.17)
θ	0.71	0.17	(0.34,0.95)
v	46.29	27.33	(10.29,110.9)
Indian states recovered cases			
β_0	-0.76	0.98	(-0.71,1.41)
β_1	0.22	0.03	(0.14,0.30)
θ	0.74	0.17	(0.36,0.96)
v	307.3	191.5	(77.96,767)

SD: Standard deviation, HPD: Highest posterior density.

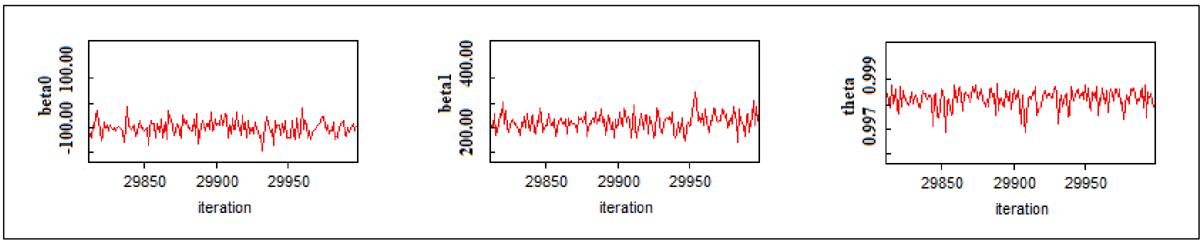

FIGURE 3: Trace plot history in Model 1.

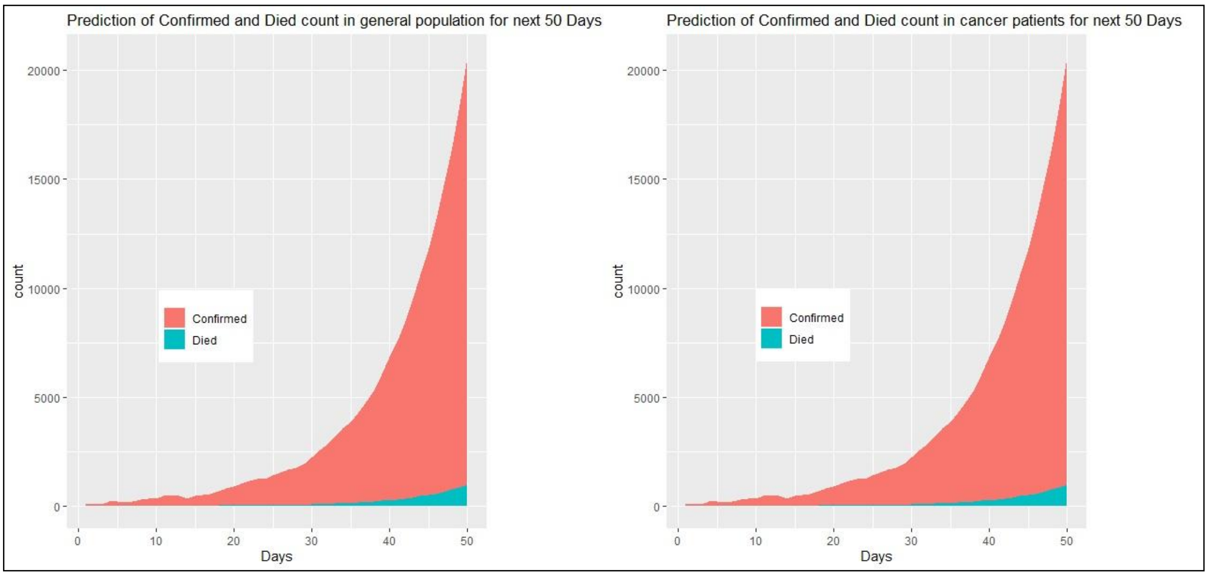

FIGURE 4: The prediction of case count and death count for next 100 days in general population.

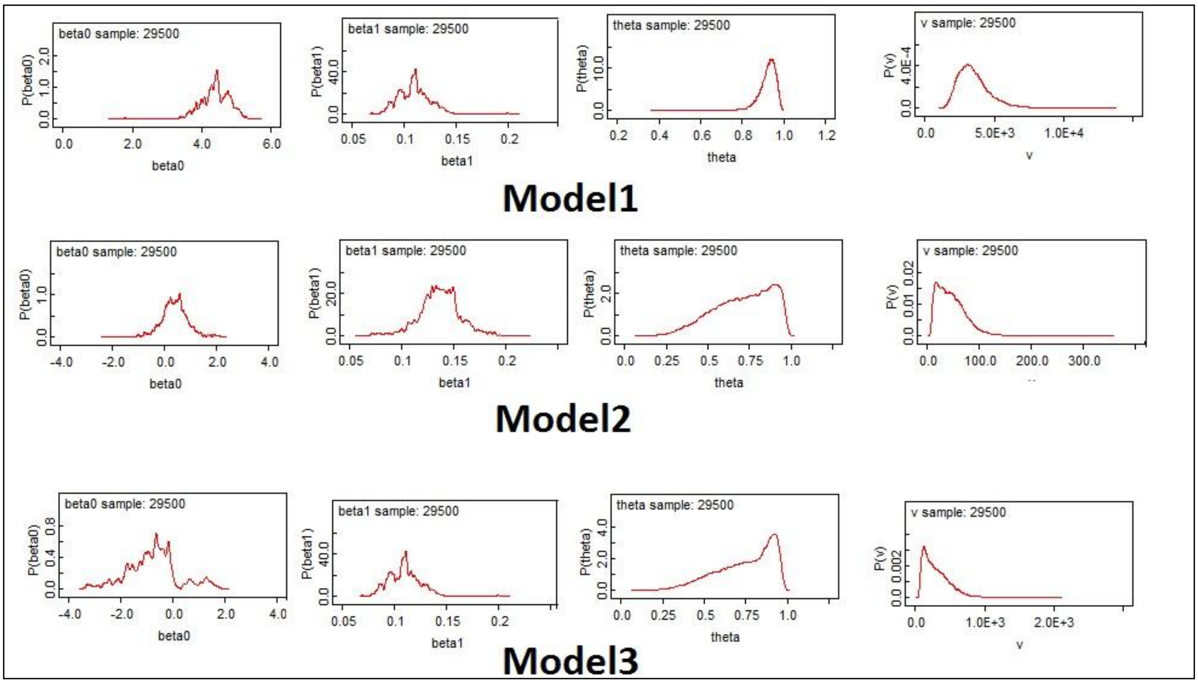

FIGURE 5: Iteration convergence of different parameters.

TABLE 2: Non-linear autoregressive posterior estimates of confirmed, deceased and recovered cases on prediction.

Parameter	Mean	SD	95% (HPD)
Indian states Confirmed cases			
β_0	10.91	237.3	(-59.08,72.1)
β_1	15.84	35.08	(-44.67,78.47)
θ	1.61	.000	(1.46,1. 74)
v	0.00	0.00	(0.00,0.00)
Indian states Deceased cases			
β_0	4.92	57.52	(-58.92,72.97)
β_1	25.5	24.12	(-21.68,71.08)
θ	1.96	0.13	(1.70,2.22)
v	0.000	.000	(0.00,0.00)
Indian states Recovered cases			
β_0	-5.87	29.42	(-60.07,52.92)
β_1	26.2	12.13	(0.69,49.91)
θ	1.31	0.13	(1.04,1.62)
v	0.00	0.00	(0.00,0.00)

HPD: Highest posterior density.

CONCLUSION

This document aimed to provide a practical scenario in presence of COVID-19 infection spread among Indian population. COVID-19 cases, death counts are available for every day. In terms of data availability, especially in the public domain, it is a unique scenario. There are several scopes to develop the model and get better prediction. In the last ten months government administration to the common public, everyone put all efforts to stop the spread of this virus. This behaviour is entirely new, and getting an error-free prediction is beyond the scope. Perhaps, all effort to predict cases and deaths were to make alert the administration.

Similarly, every effort was taken by the Indian government to prove the predictions were wrong, and that was the ultimate purpose of this prediction. After having, several challenges like developing economy, vast and dense population, India could retain the total fatality under 1.58 lakhs until now. This work served to the government of India to make the decisions by making lockdown criteria time to time relax or stringent until now. The right autoregressive prediction will obviously make it simple to plan the lockdown criteria.

Source of Finance

Authors are thankful to the Science & Technology, Government of India, for providing necessary support to carry out the present research work through file number MSC/2020/000063.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

All authors contributed equally while this study preparing.

REFERENCES

1. Jin X, Banerjee S, Carlin BP. Order-free co-regionalized areal data models with application to multiple-disease mapping. J R Stat Soc Series B Stat Methodol. 2007;1;69(5):817-38. [Crossref] [PubMed] [PMC]
2. Hoon Kim, Dongchu Sun, Tsutakawa RK. A bivariate bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. Journal of the American Statistical Association. 2001;96(456):1506-21. [Crossref]
3. Julian Besag. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological). 1974;36(2):192-236. [Crossref]
4. Langford IH, Leyland AH, Rasbash J, Goldstein H. Multilevel modelling of the geographical distributions of diseases. Journal of the Royal Statistical Society. 1999;48(2):253-68. [Crossref]
5. Leyland AH, Langford IH, Rasbash J, Goldstein H. Multivariate spatial models for event data. Statistics in Medicine. 2000;19(17-18):2469-78. [Link]
6. Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics. 2003;4(1):11-25. [Crossref] [PubMed]
7. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57(1):97-109. [Crossref]