ORIGINAL ARTICLE

Malaria in India: A Predictive Study

Surajit Das[1], Tapash Ranjan Saha[2], Sandeep Poddar[3], Sabyasachi Das[4]

[1] IMS Business School, Roypur Ghutiary Sharif Road, P.O: Champahati, P.S.-Sonarpur South 24 Pgs, Kolkata, West Bengal 700150, India
[2] IMS Group, 93 Mukundapur Main Road, Eastern Metropolitan Bypass, near Devi Shetty Hospital, Mukundapur, Kolkata, West Bengal 700099, India
[3] Lincoln University College, No. 12-18, Off Jalan perbandaran, SS6/12, 47301 Petaling Jaya, Selangor, Malaysia
[4] Department of Physiology, Faculty of Medicine, Lincoln University College, No. 2, Jalan Stadium, SS 7/15, Kelana Jaya,47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

ABSTRACT

Introduction: Malaria is devastating infectious disease not only India but also throughout the globe due to its high morbidity and mortality factor for last few centuries. From 19th and early 20th centuries, almost a quarter of the Indian populations were severely suffering from malaria. The economic loss due to increased mortality in malaria was estimated 10 million rupees per year in 1935. According to the World Malaria Report of 2017, malaria incidence accounted for 58% of cases in India. The objective of this study is to prediction of "annual" malaria incidences in India, depending on the basis of last 22 years national malaria epidemiology data. **Methods:** This study uses data from the official website of the National Program for the Control of Vector borne Diseases (NVBDCP) (http://nvbdcp.gov.in/) from 1995 to 2016. For creating a forecasting tool on Malaria surveillance in India, Econometric forecasting model (ARIMA Model ((0,1,1) (1,0,0) 12)) was used. **Results:** ARIMA statistical model ((0,1,1) (1,0,0) 12) found to be highly effective and significant ($P < 0.05$) in prediction of future epidemiological surveillance of malaria in India. ARIMA statistical model could be successfully use in prediction of annual malaria incidences in India after adjusting different highly contributing environmental and geographical factors, such as climate change, temperature, rainfall, and relative humidity. **Conclusion:** The historical forecast of the occurrence of malaria in India will allow the government to improve planning, control and prevention through public health interventions. In addition, the pharmaceutical industry will assist medical members in pre-treatment and drug interventions to respond to the increased or decreased occurrence of malaria.

Keywords: Malaria, India, ARIMA Model, Econometric forecasting model

Corresponding Author:
Sabyasachi Das, PhD
Email: sabyasachi@lincoln.edu.my
Tel: +60 11-61908761

INTRODUCTION

Malaria is a devastating parasitic disease that accounted for its high mortality and morbidity effect in India for past few centuries. At the end of the nineteenth century and the beginning of the twentieth century, nearly a quarter of the Indian populations (including both urban and rural) were suffering from malaria, especially in northeast India, Ganges delta, and the lowland and forest ranges of Easternghat as well as Westernghat. Following annual national malaria epidemiological report of 2014, 275 million people in India resides at high malaria transmission (> 1 case per 1,000) areas (1). However, 0.88 million confirmed cases of malarial infection were reported in 2013, with a prevalence of *Plasmodium falciparum* (53%) infections (1,2). According to global malaria report 2017, WHO had warned, 50% of global populations (698 million) were residing in the malaria risk zone (3). The same global malaria report also suggested that India had contributed 6 % of global malaria incidences and more importantly 51% of global *P. vivax* infection with an estimation of total 1.31 million (95%CI, 0.94 to 1.83 million) confirmed cases along with 23990 deaths (3). Mortality due to malaria infection in India were prevalent in the states of Orissa, followed by Chhattisgarh, Mizoram, Meghalaya, Gujarat, Maharashtra, Rajasthan, Karnataka, Madhya Pradesh, Jharkhand and Goa, (4). The primary and principal cause of mortality was the emergence and spreading of resistant parasite (5). *P. vivax* and *P. falciparum*, both the parasite had developed multiple gene mutations, which resulted the emergence of drug resistant parasite (6, 7). Increase spreading of parasite resistant

to commonly used antimalarial chloroquine (CQ) and sulfadoxine-pyrimethamine (SP) were a serious threat in the direction of the national malaria control (8-13). In India, government had introduced artemisinin based combination therapy against uncomplicated malaria since 2009, after replacing single use of chloroquine and sulfadoxine-pyrimethamine (14).

The relationships between different parasite species (*P. falciparum* and *P. vivax*) differ among diverse geographical locations. *P. falciparum* infections were responsible for forty to ninety percent of infections in ethnic forest areas; however 10-20% of malaria incidence in most lowland and mountainous areas of northern, northwestern India and south of Tamil Nadu caused due to *P. falciparum*. Recent epidemiology based studies of malaria proved that the alteration of different climatic conditions like temperature rainfall, and relative humidity perhaps increase the disease burden of malaria. It was proved with precession that in most of places malarial transmission was seasonal, having one prominent peak at monsoon season and another just before the onset of winter (15). Malaria reduced the economic growth of different malaria endemic countries by greater than 1% point per year. Unrestricted urbanization, drought, labor migration and neglected control measures were the most selective contributing factors for the revival of malarial infection in India, and the underlying problem of malaria would worse in the upcoming years.

Early prediction of any epidemic or endemic infections is always extremely helpful for the government to take potential preventive measures prior to the onset of infection. Different countries like Thailand, Sri Lanka and Ethiopia were successfully used ARIMA statistical model for the early prediction of dengue or malaria infections in respective countries (16). These early predictions put enormous advantage for the management and control of the disease burden. Therefore the objective of our study was to evaluate the effectiveness of autoregressive integrated moving average (ARIMA) statistical model for the prediction of annual malaria incidences in India.

MATERIALS AND METHODS

Sampling Data

We used the secondary data, derived from the annual national malaria epidemiological report during 1995 to 2016 for our predictive research analysis. We collected the year wise and state wise annual malaria incidence dataset from national malaria epidemiological report database (WWW.nvbdcp.gov.in/.)

Statistical procedure

We used ARIMA statistical model version (0,1,1) (1,0,0) 12), 2019; through E-views 10 for extensive analysis of the chronological data set of malarial incidence in India after adjusting different environmental, geographical, seasonal contributing factors, which ultimately results more effective prediction power compared to existing other predictive model. Therefore different countries of the globe were successfully used similar statistical model to forecast various communicable and non-communicable diseases for last few decades (16). Furthermore relative simplicity and stability of this statistical tool provide a greater advantage in forecasting of malaria incidences over complex statistical model, where more detailed relatively complex data set were required for the successful execution of the model.

Model Building Steps

Model identification: The first thing to consider in the forecast analysis was to determine whether series belonged to stationary and to check if the mean and variance were stable. The stationarity of the used data set were critically verified through Augmented Dickey Fuller (ADF) test. If the series was not stationary, the transformation can be performed to achieve stationarity.

Estimation of the model: Once the final model was identified, the parameters of the model must be estimated. The parameters were estimated using the maximum likelihood estimation (MLE) which represented the values of the parameters of the orders p and q as well as the constant term and the residues that were obtained. After estimating the model parameters, a residue analysis was performed to determine how well the model fits the data. It assumed that the model fits well if there was no visible pattern in the residue graph, ie there was any autocorrelation in model residuals.

RESULTS

Malaria Incidence in India

Malaria infection was prevalent in last two to three decades in throughout India. Figure I represented the malaria situation across the country, as evidenced by the monitoring data for the period 1995-2016. Incidences of malaria infection in India were extensively high in 1990s. After initial decrease in 2013, the incidence of malaria infection was begun to increase since 2014 (Figure 1).

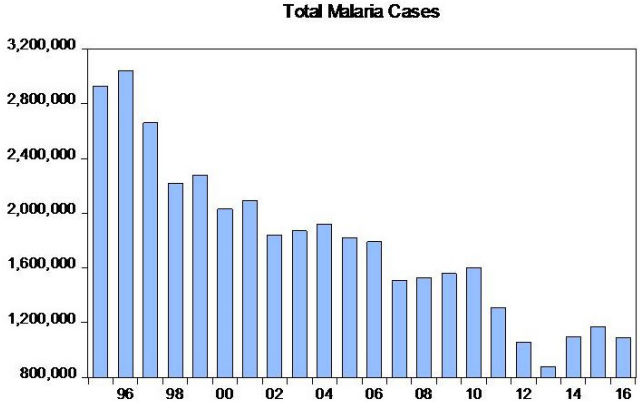


Figure 1: Histogram of Malaria Incidence in India from 1995 to 2016

Epidemiological Indicators for Malaria

For easy and better handling of the sample size we have converted the total integer of malaria positive incidence into to log of malarial incidence. We have analyzed and represented the annual data of malaria cases in India from 1995 to 2016 and converted the data into log 10. (Table I). This log 10 value is very important to analyze the ARIMA model. Highest log-10 value of 6.482874 was observed in 1996 where as the lowest value of 5.944483 were found in 2013.

Table I: Epidemiological Indicators for Malaria in India (1995-2016)

Year	Total Malaria Cases	LOG
1995	2930000	6.466868
1996	3040000	6.482874
1997	2660000	6.424882
1998	2220000	6.346353
1999	2280000	6.357935
2000	2030000	6.307496
2001	2090000	6.320146
2002	1840000	6.264818
2003	1870000	6.271842
2004	1920000	6.283301
2005	1820000	6.260071
2006	1790000	6.252853
2007	1510000	6.178977
2008	1530000	6.184691
2009	1560000	6.193125
2010	1600000	6.20412
2011	1310000	6.117271
2012	1060000	6.025306
2013	880000	5.944483
2014	1100000	6.041393
2015	1170000	6.068186
2016	1090000	6.037426

Model Development

We evaluated the stationarity of the data series consisting of malaria cases in India. In univariate forecasting approach the order of difference for stationarity of the data may be different but it should be 1 (1 , or 0.)

The stationarity of the data series and the proper order of amalgamation among the data series were precisely verified through augmented dickey test (ADF). The ADF-tests were done through estimation of regression equation which was presented in Table II. It demonstrated that the calculated data was stationary at 5% significance level because of the $p < 0.05$, so the null-hypothesis was being rejected at 5% level.

ARIMA model assessment

ARIMA model had three values to be determined i.e.

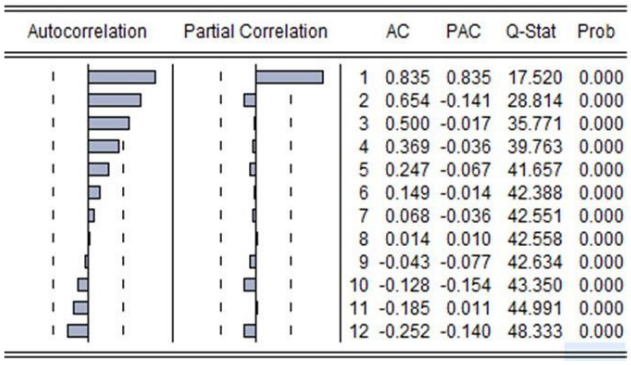
Table II: Augmented Dickey- Fuller unit root test

Null Hypothesis: LOG01 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, max lag = 1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.733847	0.0435
Test critical values:	1% level	-4.498307	
	5% level	-3.658446	
	10% level	-3.268973	

'd' value representing the differential order of the trend after being stabilized, 'p' value that represents the partial autocorrelation between the data (AR term) and 'q' value that represents the autocorrelation between the data (MA term), on the basis of which the forecasting is done. To identify the lag of p & q correlogram has been used in Figure 2. Figure 2 showed that there was a significant pick of ACF and PACF. PACF was used to identify the order of AR term and the significant pick is 1 on the other hand ACF is used to identify the MA term and the number of significant pick is 3.so the probable model is (1, 3).

Date: 09/03/19 Time: 13:41
Sample: 1995 2016
Included observations: 22



	AC	PAC	Q-Stat	Prob
1	0.835	0.835	17.520	0.000
2	0.654	-0.141	28.814	0.000
3	0.500	-0.017	35.771	0.000
4	0.369	-0.036	39.763	0.000
5	0.247	-0.067	41.657	0.000
6	0.149	-0.014	42.388	0.000
7	0.068	-0.036	42.551	0.000
8	0.014	0.010	42.558	0.000
9	-0.043	-0.077	42.634	0.000
10	-0.128	-0.154	43.350	0.000
11	-0.185	0.011	44.991	0.000
12	-0.252	-0.140	48.333	0.000

Figure 2: Assessment of ACF and PACF

Assessment of ARIMA regression

Following ARIMA regression, it was found that the present values significantly depend upon the past values on the third lag value with significance of $p < 0.05$. This was done to prepare the data for forecasting so that the future predicted values will depend upon the trend of the past values of the incidence cases of malaria in India. The p-value of the ARIMA regression was significant at the level of 0.05 and the forecasting of malaria through ARIMA ((0,1,1) (1,0,0) 12)) was very reliable in India (Table III).

In our case, we proposed the model as

$$yt = 0.89yt\text{-}1 + \omega t + (\text{-})0.699\omega t\ t\text{-}3$$

Here, yt represented absolute malaria case in current time; yt-1 represented the value of just previous period whereas yt-3 is the d value 3 time lag, and ωt represented the error term.

Table III: ARMA Model

Dependent Variable: LOG01
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/03/19 Time: 13:57
Sample: 1995- 2016
Included observations: 22
Failure to improve objective (non-zero gradients) after 35 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.89675	0.001318	758.8136	0.0000
MA(3)	-0.699969	1.39E-07	-7174200.	0.0000
SIGMASQ	0.003923	0.001408	2.787079	0.0117
R-squared	0.808586	Mean dependent var		6.228837
Adjusted R-squared	0.788437	S.D. dependent var		0.146535
S.E. of regression	0.067400	Akaike info criterion		-1.728582
Sum squared reside	0.086313	Schwarz criterion		-1.579803
Log likelihood	22.01440	Hannan-Quinn criter.		-1.693534
Durbin-Watson stat	0.575950			
Inverted AR Roots	1.00			
Inverted MA Roots	1.00	-.50+.87i	-.50-.87i	

Residual Checking

We had evaluated the residual checking of AC and PAC (Figure 3). Incidences of malaria infection in India were extensively high in 1990s. After initial decrease in 2013, the incidence of malaria infection was begun to increase since 2014. We also evaluated ARCH- LM test which represented a significant findings with a Log likelihood of 83.71904 and F-statistic of 2.363474.

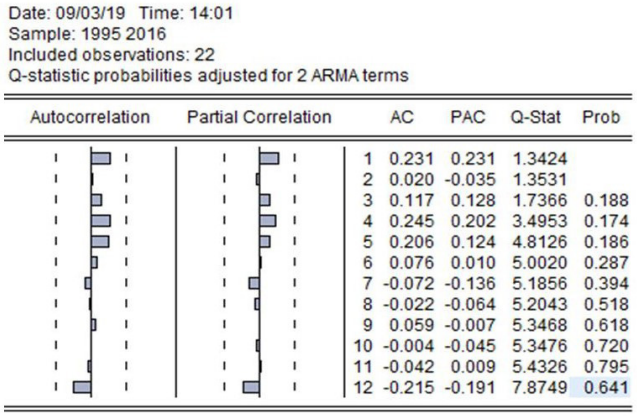


Figure 3: AC and PAC for residual checking

DISCUSSION

ARIMA statistical model version ((0,1,1) (1,0,0) 12) appeared to be highly effective and significant ($P < 0.05$) in prediction of future annual malaria incidences in India after adjusting different highly contributing environmental and geographical factors, like climate change, temperature, rainfall, and relative humidity. Different factors perhaps explain the predisposition of prevalence of infections, and its rapid spreading in different malaria endemic parts of the country such as, less parasite transmission, promoted by high-rates of parasite inbreeding, host-immunity, individual pharmacokinetics, pharmaco-dynamics, along with inadequate drug dosing with poor quality of antimalarial used (17). In addition, report suggested that Plasmodium falciparum strains in Southeastern Asia perhaps had some genetic attenuation for development of novel mutations leads to the rapid spreading of infection as well as drug-resistance (18), this rapid spreading of malarial infection results enormous obscurity towards the control and elimination of the infection. National vector borne disease control program (NVBDCP), has recommended artesunate plus SP combination (ACT) by replacing single use of CQ and SP (14). Artemisinin derivatives are the last and final order of antimalarial. Furthermore emergence and subsequent spreading of artemisinin-resistant parasite in eastern India had awfully threatened the national malaria control and elimination progress (19, 20). Furthermore, the precise genetic architecture in relation to rapid spreading of resistant parasite in northeast, southwest, eastern and central India might different because of extensive variation in socio-demographic, environmental, seasonal and parasitic factors (15, 21, 22). Therefore precise early prediction of malaria infection in India will possess enormous interest in disease prognosis, treatment and management.

On this circumstances ARIMA (0,1,1) (1,0,0) 12 was found as a functional statistical software for the successful estimation of annual malaria incidences in the future after adjusting different environmental and geographical factors. The environmental factors are key contributor of this model as vector (anopheles mosquito) transmission and outbreak were solely depended on environmental condition such as temperature, rainfall etc (15, 23). Previous report suggested that ARIMA statistical model was successfully predicted the malaria incidence in Bhutan. Furthermore, this statistical model was widely used in countries like Sri Lanka and Ethiopia for successful forecasting malaria. Recently, a synchronous active research on ARIMA (0,1,1) (0,1,0) was conducted in NIMR, New, Delhi, for probable prediction of malaria (16). ARIMA statistical model was also successfully predicted the dengue incidences in Thailand (24). Therefore, depending on our results, we could say that ARIMA model had enormous prospect for future prediction of malaria in India.

CONCLUSION

Chronological prediction of the onset of malaria in India offers the government the opportunity to improve planning, control and prevention through public health interventions. In addition, the pharmaceutical industry will help health professionals before treatment and medications, depending on whether the incidence of malaria is increasing or decreasing. As a result, the prediction allows better application of control measures and strategies to control the spread of malaria.

ACKNOWLEDGEMENTS

We are gratefulness to IMS Business School, Kolkata

for active support and offering the essential facilities to execute the work. We are especially grateful to Lincoln University College, Malaysia for providing the support to execute the work. We are also highly thankful to the National Vector-borne Diseases Control Program (NVBDCP) for their online year wise and state wise data set for malaria incidences in India which we uses as the sources of our primary data.

REFERENCES

1. NVBDCP annual report 2014-2015, nvbdcp.gov.in/Doc/Annual-report-NVBDCP-2014-15.pdf
2. Guidelines for Diagnosis and Treatment of Malaria in India 2014. National Institute of Malaria Research, New Delhi, Government of India. 2014
3. World malaria report 2017 by World Health Organisation (WHO); https://www.who.int/malaria/publications/world-malaria-report-2017/en/
4. Annual report 2014-2015, National Vector Borne Disease Control Programme; https://nvbdcp.gov.in/Doc/Annual-report-NVBDCP-2014-15.pdf
5. World Health Organization (WHO). Assessment and monitoring of antimalarial drug efficacy for the treatment of uncomplicated falciparum malaria. Geneva. 2003. http:// www.cdc.gov/malaria/resources/pdf/drug.../WHO2003_monitoring.pdf
6. Das S, Tripathy S, Dash SK, Chattopadhayaya S, Mandal D, Hati AK et al. Progressive Increase in Point Mutations Associates Chloroquine Resistance: Even after Withdrawal of Chloroquine use in India. Int Journal for Parasitology-Drugs and Drug Resistance. 2017; 7: 251-61
7. Das S, Banik A, Hati AK, Roy S. Low prevalence of dihydro folate reductase (dhfr) and dihydropteroate synthase (dhps) quadruple and quintuple mutant alleles associated with SP resistance in Plasmodium vivax isolates of West Bengal, India. Malaria Journal. 2016; 15: 395. DOI 10.1186/s12936-016-1445-9
8. Das S, KarMahapatra S, Tripathy S, Dash S, Chattopadhyay S, Mandal D et al. Double mutation in the pfmdr1 gene is associated with emergence of chloroquine-resistant Plasmodium falciparum malaria in Eastern India. Antimicrob Agents Chemother. 2014; 58: 5909-15
9. Sharma YD. Molecular surveillance of drug-resistant malaria in India. Current Science. 2012; 102: 696-703.
10. Das S, Chakraborty SP, Hati AK, Roy S. Association between prevalence of chloroquine resistance and unusual mutation in pfmdr-I and pfcrt gene in India. American Journal of Tropical Medicine & Hygiene, 2013; 88(5): 828-34.
11. Ahmed A, Lumb V, Das MK, Dev V, Sharma YD. Prevalence of mutations associated with higher levels of sulfadoxine-pyrimethamine resistance in Plasmodium falciparum isolates from Car Nicobar Island and Assam, India. Antimicrob. Agents. Chemother. 2006; 50: 3934-38.
12. Das S, Chakraborty SP, Tripathy S, Hati AK, Roy S. The novel quadruple mutations in dihydropteroate synthase genes of Plasmodium falciparum in West Bengal, India. Trop. Med. Int. Health. 2012; 17(11): 1329-34.
13. Das S, Chakraborty SP, Hati AK, Roy S. Malaria treatment fails with novel mutation in Plasmodium falciparum di-hidrofolate reductase (pfdhfr) gene in Kolkata, West Bengal, India. Int. J. Antimicrob. Agents. 2013; 41(5): 447-51.
14. Guidelines for diagnosis and treatment of malaria in India. Government of India. 2009.
15. Gomez-Elipe, A Otero, M van Herp, A Aguirre-Jaime. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997–2003, Malaria Journal. 2007; 6(1).
16. Sharma MK, Sharma PK, Kumar B, Mahajan S, Archana A. Testing statistical models for forecasting malaria cases in India. International Research Journal of Agricultural Economics and Statistics. 2017; 8(1): 8-14.
17. Miotto O, lmagro-Garcia J, MacInnis B, Campino S, Rocket KA et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. 2013; 45: 648-55
18. Beez D, Sanchez CP, Stein WD, Lanzer M. Genetic predisposition favors the acquisition of stable artemisinin resistance in malaria parasites. Antimicrob Agents Chemother. 2011; 55: 50-55.
19. Das S, Saha B, Hati AK, Roy S. Evidence of Artemisinin resistant Plasmodium falciparum malaria in Eastern India. N Engl J Med. 2018; 379 (20): 1962-64
20. Das S, Manna S, Saha B, Hati AK, Roy S. Novel pfkelch13 gene polymorphism associates with artemisinin resistance in eastern India. Clinical infectious Disease. 2019; 69 (7): 1144-52
21. Mishra N, Prajapati SK, Kaitholia K, Bharti RS, Srivastava B, Phookan S et al. Surveillance of Artemisinin Resistance in Plasmodium falciparum in India Using the kelch13 Molecular Marker. Antimicrob Agents Chemother. 2015; 59(5): 2548–2553.
22. Chakrabarti R, White J, Babar PH, Kumar S, Mudeppa DG, Mascarenhas A et al. Decreased in vitro artemisinin sensitivity of Plasmodium falciparum across India. Antimicrobial Agents and Chemotherapy. 2019; doi:10.1128/AAC.00101-19.
23. Wangdi K, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. Development of temporal modeling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. Malaria Journal, 2010; 9.
24. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. Dengue Bulletin. 2006; 30: 99-106.