Predictive Time Series Modeling of Malaria Incidence in India: A 20-Year Retrospective and 5-Year Forecast.

Dr Ami Soni

Abstract

This study presents a predictive time series analysis of malaria incidence in India, utilizing a two-decade dataset. Malaria, caused by Plasmodium species and transmitted by Anopheles mosquitoes, poses a significant public health challenge in the WHO South-East Asia Region. We employed linear regression and ARIMA modeling techniques to analyze historical incidence data, ensuring rigorous data cleaning for accuracy. Results indicate a statistically significant declining trend in malaria cases, with projections suggesting continued decreases in malaria cases over the next five years. These findings highlight the effectiveness of current malaria elimination strategies in India and underscore the necessity for sustained public health interventions. Ultimately, this study emphasizes the importance of data-driven approaches for understanding and forecasting malaria trends, contributing to the goal of malaria eradication by 2030.

keywords: Malaria, Predictive Time Series Analysis, ARIMA ,Linear Regression, Public health, India, Malaria incidence, Malaria eradication, Public health strategies.

Introduction

Malaria is a potentially life-threatening illness caused by several parasitic species, namely *Plasmodium vivax*, *Plasmodium falciparum*, *Plasmodium malariae*, and *Plasmodium ovale*. It is transmitted through the bite of infected female Anopheles mosquitoes.(1) In 2022, there were approximately 249 million of malaria cases globally reported across 85 endemic countries and regions, reflecting an increase of 5 million cases compared from the previous year.(2)

Within the WHO South-East Asia Region, nine countries were identified as malaria-endemic, contributing to 5.2 million cases, which represented 2% of the global malaria burden.(1) India's considerable malaria case load places it among the eleven countries with the highest burden, which collectively account for 70% of the global incidence of malaria as seen in Figure, with the remaining ten countries located in Africa. India accounted for roughly 65.7% of all malaria cases within this region, with nearly 46% of cases attributed to *P. vivax*. Together, India and Indonesia were responsible for 94% of malaria-related fatalities in the region in 2022.(4)

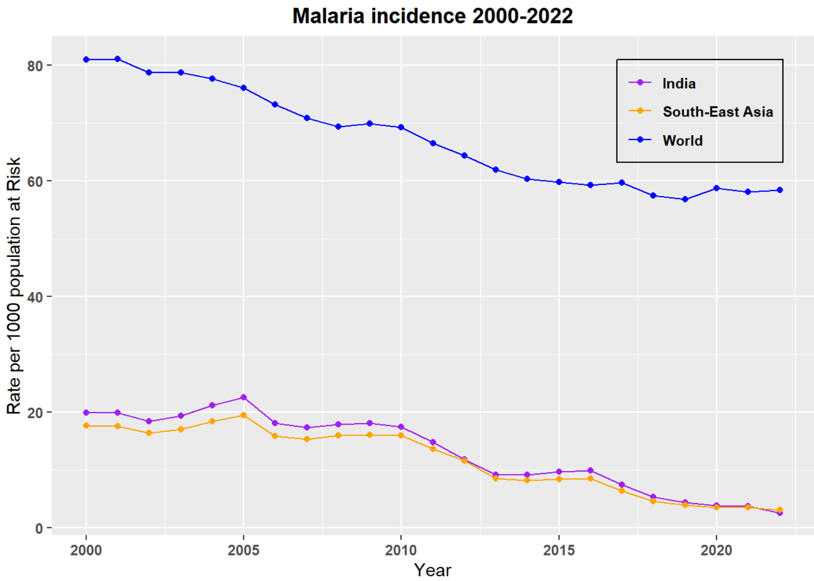

Figure 1: Malaria incidence in World, South-East Asia and India (2000-2022)

Adopted by the World Health Assembly in May 2015 and revised in 2021, the "Global Technical Strategy for Malaria 2016–2030" guides this initiative. The WHO Global Malaria Programme (GMP) is tasked with coordinating international efforts focused on controlling and ultimately eliminating malaria. Aligned with this strategy, the Government of India launched the National Framework for Malaria Elimination 2016–2030 in February 2016, followed by the National Strategic Plan for Malaria Elimination 2017–2022 in July 2017, with assistance from the WHO. India's goal is to attain malaria-free status by 2027 and achieve complete eradication by 2030. Under the NSP 2017-2022, the WHO has supported the acceleration of malaria elimination initiatives across several states. In July 2019, the High Burden to High Impact (HBHI) strategy was introduced in four highly endemic states: West Bengal, Jharkhand, Chhattisgarh, and Madhya Pradesh. Launched in 2018, the WHO's "High Burden, High Impact" initiative seeks to strengthen malaria control efforts in these high-burden nations. A significant challenge in addressing malaria in India lies in the extensive diversity of its ecosystems; approximately 80% of the malaria burden is concentrated in only ten of the 28 states, six of which are located in the northeastern part of the country.(3,4)

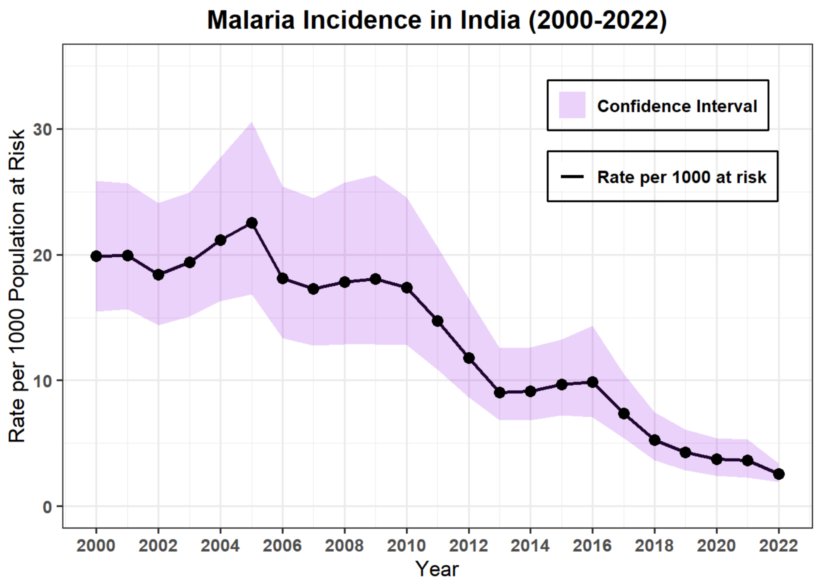

Figure 2: Malaria incidence of India with confidence internval.

Time Series Analysis in Malaria Incidence

Time series analysis is crucial for understanding and forecasting trends in malaria incidence, particularly in malaria-endemic regions like India. This study employs a 20-year dataset to analyze malaria incidences.

By examining historical data, we identify key components such as trends and irregular fluctuations. Due to the annual frequency of the data and the absence of a clear seasonal component, traditional decomposition methods like `decompose()` or `stl()` function, which generally require a larger dataset to capture seasonal components effectively were skipped. Instead, we utilized linear regression `lm()` and ARIMA `arima()` or `auto.arima()` modeling techniques, ensuring rigorous data cleaning and analysis.

By leveraging these advanced analytical methods, this study aims to provide insights into malaria trends in India, supporting data-driven public health strategies aimed at malaria elimination by 2030. Understanding these trends is essential for guiding interventions in high-burden areas and optimizing resource allocation.

Linear Regression Model:

A linear regression model was fitted to the time series to analyse the trend, using the rate per 1000 population at risk as the dependent variable and time as the independent variable. The results were plotted to visualise the time series data alongside the fitted trend line.

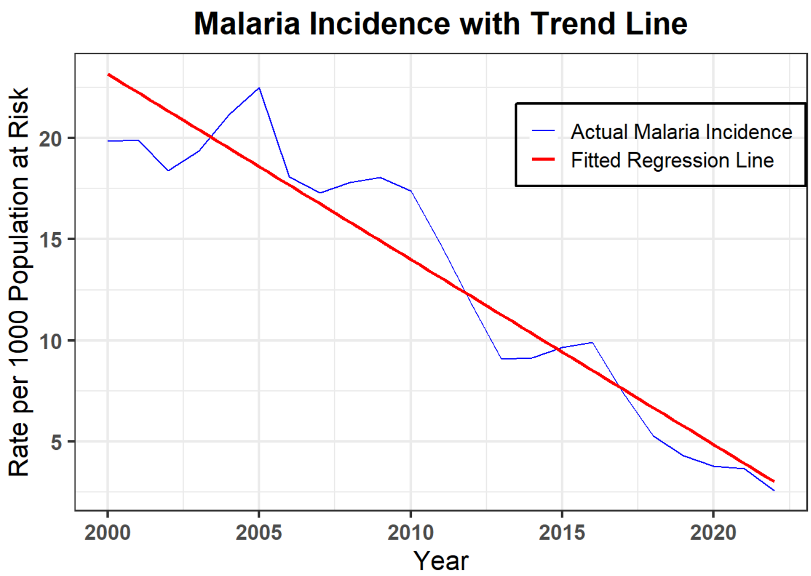

Figure 3: Linear regression model to the time series analysis

Summary on trend model:

Table 1a: Coefficients of the Trend model

Term	Estimate	Std_Error	t_value	p_value
(Intercept)	1858.76385	128.40385	14.48	2.13e-12
Time	-0.91779	0.06385	-14.37	2.44e-12

Table 1b Residuals & Summary Statistics of the Trend Model

Statistic	Value
Residuals	
Min	-3.3219
1Q	-1.3004
Median	-0.2758
3Q	1.5122

Statistic	Value
Max	3.9144
Summary Statstic	
Residual Standard Error	2.031
Multiple R-squared	0.9077
Adjusted R-squared	0.9033
F-statistic	206.6
p-value	2.441e-12

Results and Discussion:

The linear regression analysis of malaria incidence rates reveals significant findings:

Coefficients:

- **Intercept**: The estimated baseline incidence rate at the start of the time series was 1858.76 ($p < 0.001$), serving as a reference point despite not corresponding to a real value due to the data starting in 2000.
- **Time**: The coefficient of -0.91779 indicates a significant annual decrease in malaria incidence by approximately 0.92 units ($p < 0.001$), highlighting a statistically significant downward trend.

Residuals: The residuals range from -3.32 to 3.91, suggesting that while the model fits well, there are some unexplained variations.

Model Fit:

- **Residual Standard Error**: At 2.031, this value indicates a relatively good fit of the model.
- **Multiple R-squared**: The model explains about 90.77% of the variability in malaria incidence, demonstrating a strong fit.
- **Adjusted R-squared**: At 0.9033, this value confirms the model's robustness while accounting for the number of predictors.
- **F-statistic**: The value of 206.6, with a p-value of 2.441e-12, indicates that the model is statistically significant overall.

Forecasting with ARIMA

ARIMA Model and Forecast:

The `auto.arima()` function was utilised to fit an ARIMA model to the time series. The `forecast(fit, h = 5)` function was subsequently employed to project malaria incidence for the next five years, with forecasts illustrated along with confidence intervals.

Residuals Check:

The `checkresiduals(fit)` function evaluates the residuals of the ARIMA model to ensure that no patterns remain unexplained.

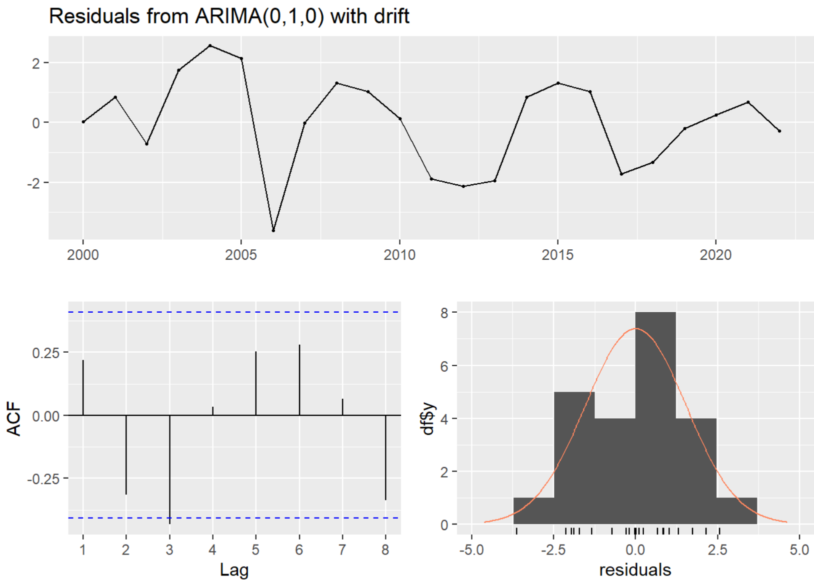

Figure 4: Residuals with Arima Model (0,1,0)

Table 2: Residuals from ARIMA (0,1,0) with drift

Statistic	Value
Q*	11.447
Degrees of Freedom	5
p-value	0.04321
Model df	0
Total lags used	5

ACF and PACF Analysis:

The ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots provide critical diagnostics for evaluating the adequacy of the fitted time series model.

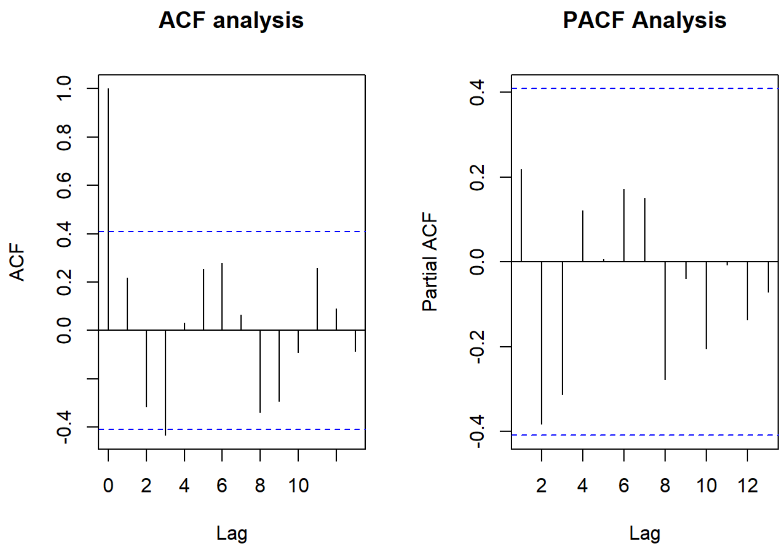

Figure 5: ACF and PACF Analysis.

ACF Residuals:

- The ACF plot indicates minimal autocorrelation among the residuals across various time lags, with most points falling within the confidence interval. This suggests that the model adequately captured the time dependence in the data. Minor spikes outside the bounds, particularly at lag 1, may indicate a slight amount of autocorrelation not captured by the model, though the overall effect is likely minimal.

PACF Residuals:

- The PACF plot similarly demonstrates that most partial autocorrelation values lie within the confidence bounds, indicating no significant correlation remains among the residuals after accounting for previous lags. A small spike at lag 1 suggests possible residual short-term autocorrelation, warranting slight adjustments to the model.

Model Fit:

Based on the ACF and PACF plots, this study concludes that the time series model has largely succeeded in capturing the underlying structure of the data. The residuals appear to behave like white noise, with negligible autocorrelation remaining, indicating a reasonable model fit.

The small spikes observed at lag 1 in both ACF and PACF suggest minor autocorrelation that the model has not fully accounted for. While this does not pose a significant issue, exploring slight model adjustments could help eliminate any dependencies.

Overall, the model performs well, with residuals predominantly behaving as random noise. The minimal remaining autocorrelation suggests the model is adequate, although further refinement could enhance its performance.

Forecast of Malaria Incidence in India for the Next Five Years

The projections as seen in Figure (@fig:Forecast), indicate a sustained decline in malaria incidence, reinforcing the positive impact of ongoing public health interventions.

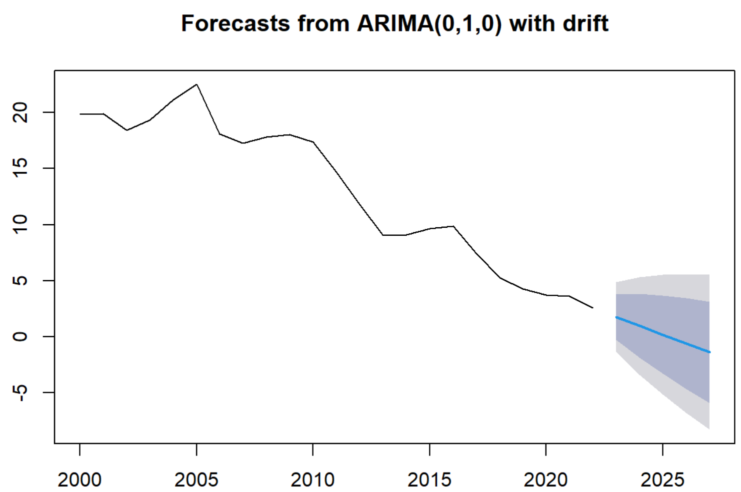

Figure 6: Forecast of Malaria Incidence in India for next 5 years

In 2023, the projected incidence is approximately **1.77 cases**, declining to **0.99 cases** in 2024. The forecast further decreases to **0.20 cases** in 2025 and shows a negative trend with **-0.59 cases** in 2026 and **-1.37 cases** in 2027, suggesting a significant reduction in malaria incidence. The **80% and 95% confidence intervals** indicate uncertainty in these predictions, with wide ranges reflecting variability in actual cases. For instance, in 2023, the 95% interval spans from **-1.31 to 4.86 cases**. Overall, the forecast suggests a positive downward trend in malaria cases.

Table 3: Forecast of Malaria Incidence in India for next five years

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2023	1.7729418	-0.2458667	3.791750	-1.314559	4.860443

Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2024	0.9863136	-1.8687128	3.841340	-3.380072	5.352700
2025	0.1996855	-3.2969935	3.696364	-5.148023	5.547394
2026	-0.5869427	-4.6245598	3.450674	-6.761945	5.588060
2027	-1.3735709	-5.8877640	3.140622	-8.277434	5.530292

Conclusion

This study provides a comprehensive analysis of malaria incidence in India over the past two decades, utilizing advanced predictive time series modeling techniques. The findings indicate a significant downward trend in malaria cases, aligning with the country's public health initiatives aimed at malaria elimination. The linear regression and ARIMA models effectively captured the historical patterns and projected a continued decline in malaria incidence over the next five years, reinforcing the positive impact of ongoing interventions.

The results emphasize the importance of data-driven approaches in understanding and forecasting malaria trends, which is crucial for policymakers and public health officials. As India strives for malaria-free status by 2027, the study underscores the necessity for sustained efforts and targeted strategies to address remaining challenges, particularly in high-burden regions. Continued monitoring and evaluation of malaria trends, coupled with robust healthcare interventions, will be essential to achieving the goal of malaria eradication by 2030. Ultimately, this research contributes to the broader understanding of malaria dynamics in India and supports the development of effective public health strategies to combat this persistent health threat.

Data Availability:

The dataset used in this study are available from World Health Oraganization website.(2)

Source Code:

The source code for this study is available at GitHub/Dr-A-Soni/Malaria_Incidence_Trends (https://github.com/Dr-A-Soni/Malaria_Incidence_Trends/tree/main).

Reference

1. World Health Organization. Malaria. [Online]. Available at: https://www.who.int/india/health-topics/malaria (https://www.who.int/india/health-topics/malaria) (Accessed: 19 September 2024).
2. World Health Organization 2024 data.who.int, Malaria incidence (per 1000 population at risk) [Indicator]. https://data.who.int/indicators/i/B868307/442CEA8 (https://data.who.int/indicators/i/B868307/442CEA8) (Accessed on 19 September 2024).
3. World Health Organization. 2023. World Malaria Report 2022. Geneva: World Health Organization. https://www.who.int/publications/i/item/9789240086173 (https://www.who.int/publications/i/item/9789240086173) (Accessed on 19 September 2024).
4. World Health Organization. "Southeast Asia Region Malaria Progress Report." https://www.who.int/docs/default-source/wrindia/malaria/sear-progress-report-ind.pdf?sfvrsn=6f754bf7_2 (https://www.who.int/docs/default-source/wrindia/malaria/sear-progress-report-ind.pdf?sfvrsn=6f754bf7_2) (Accessed September 23, 2024).