World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Environmental determinants of infectious disease incidence: A discrete time series analysis in Bangladesh

Khondoker Rukyath Alam * and Mohammad Ahsan Uddin

Department of Statistics, Faculty of Science, University of Dhaka, Dhaka, Bangladesh.

World Journal of Advanced Research and Reviews, 2024, 24(03), 2590-2602

Publication history: Received on 26 November 2024; revised on 22 December 2024; accepted on 25 December 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.3.3697

Abstract

Weather is an important determinant for human health. There is an apparent increase in outbreak of many infectious diseases which reflects the combined impacts of rapid demographic, environmental, social, technological and other changes in our way-of-living. This study aimed to analyze the effects of environmental factors on infectious disease occurrences in Bangladesh using time series data from 2013 to 2022. Environmental data (monthly average temperature, monthly average relative humidity and monthly average rainfall) were collected from the Bangladesh Agricultural Research Council (BARC) and disease data (Tetanus, HIV and Chickenpox counts) were obtained from the Infectious Disease Hospital in Dhaka. Utilizing Poisson and Negative Binomial (NB) Generalized Linear Models (GLMs), the impact of weather variability on disease transmission was examined. The best-fitted models yielded AIC values of 682.7 for Tetanus, 605.96 for HIV and 608.01 for Chickenpox, with corresponding RMSE values of 3.33, 0.22 and 1.67. The study found that monthly average rainfall negatively affects Tetanus counts while temperature and humidity had significant positive and negative effects on Chickenpox counts respectively. No significant effect of environmental factors on HIV counts was observed. These findings can inform preventative measures and strategies to reduce disease impact.

Keywords: Infectious disease; HIV; Temperature; Humidity; Count regression model; Dhaka

1. Introduction

Global warming is an undeniable reality with climate change exerting multifaceted impacts on human health. Climate change is expected to displace millions of people over the coming century [1]. The World Health Organization (WHO) estimates that unhealthy environments, particularly pollution and climate change contribute to approximately 12.6 million deaths annually. Diseases caused by microorganisms such as bacteria, viruses, fungi, and parasites are called infectious diseases, remain a significant global health challenge. In 2016 alone, infectious diseases accounted for roughly 10 million deaths, representing one-fifth of all global fatalities [2]. Among these, lower respiratory tract infections have the highest mortality rate, followed by enteric infections causing diarrhea, tuberculosis, AIDS caused by HIV, and malaria [2].

Weather factors such as temperature, relative humidity and rainfall significantly influence the occurrence and spread of infectious diseases. Variations in these climatic conditions affect the survival, reproduction, and transmission of pathogens and their vectors. Higher temperatures can enhance viral replication rates and increase vector activity, while humidity impacts the viability of airborne pathogens [3]. Rainfall can create suitable breeding sites for vectors like mosquitoes, leading to higher incidences of diseases such as malaria and dengue [4]. Understanding these climatic influences is essential for predicting disease outbreaks and implementing effective public health measures.

---

* Corresponding author: Kh Rukyath Alam

Chickenpox, also known as Varicella, is a highly contagious disease caused by the varicella-zoster virus (VZV). Although it is usually mild, varicella can cause serious complications, especially in infants, adults and those with weakened immune systems. The World Health Organization (WHO) estimates that millions of cases occur annually worldwide, with significant morbidity and mortality in regions with low vaccination coverage [5]. In Bangladesh, UNICEF reportes a sharp rise in varicella cases in January 2019, with incidents nearly doubling from 2,710 to 5,376 within a single week [6]. Despite the availability of vaccines, outbreaks persist due to insufficient immunization effort [7]. Tetanus is a non-communicable infectious disease caused by Clostridium tetani spores. Due to the ubiquity of this bacterium, eradication is not feasible. While all age groups are susceptible, tetanus is particularly severe in newborns. It causes over 200,000 deaths annually, primarily affecting neonates and young children, with figures likely underestimated due to infrequent reporting [8]. Most cases are in middle-aged individuals, with 25.5% in the 30-45 age group and 23.5% in the 45-60 age group furthermore, 90.6% of patients are from rural areas [8]. In developing countries, tetanus is more common among the young due to inadequate immunization programs and improper injury treatment. Despite being life-threatening, tetanus is preventable through vaccination. Human immunodeficiency virus (HIV) targets and weakens the immune system, leading to the most advanced stage of the disease, acquired immunodeficiency syndrome (AIDS). The World Health Organization (WHO) reports that HIV has resulted in approximately 40.4 million deaths globally, with ongoing transmission in all countries. As of the end of 2022, an estimated 39.0 million people were living with HIV, and 630,000 had died from HIV-related causes that year. In Bangladesh, there were 947 new HIV cases and 232 deaths reported in 2022, bringing the total to 10,984 diagnosed cases and 2,086 deaths since 1989 [9]. WHO, along with the Global Fund and UNAIDS, has strategies in place aligned with the Sustainable Development Goal (SDG) target 3.3 to end the HIV epidemic by 2030 [10].

While there have been significant strides in understanding the epidemiology of infectious diseases, a comprehensive analysis of the interaction between environmental factors and the temporal dynamics of disease occurrences in Bangladesh remains insufficiently explored. Existing studies typically focus on individual diseases or specific environmental variables, often neglecting the combined impact of multiple environmental determinants on various infectious diseases. Furthermore, the application of time series models has generally been limited to assessing disease trends without integrating both lagged disease counts and environmental factors within a unified analytical framework.

This study aimed to fill this gap by investigating how monthly average temperature, relative humidity and rainfall influence the counts of tetanus, HIV and chickenpox. Utilizing Poisson and Negative Binomial (NB) Generalized Linear Models (GLMs), this study would incorporate lagged values of disease counts and environmental factors to provide a detailed understanding of the temporal and environmental drivers of these diseases.

2. Material and methods

2.1. The Study Area

Bangladesh, a northeastern part of South Asia, is bordered by the majestic Himalayas to the north and the Bay of Bengal to the south. To the west, it shares a border with West Bengal, India, and to the east, it is bordered by the hilly and forested regions of Tripura, Mizoram (India) and Myanmar. It is the eighth-most populous country in the world and among the most densely populated with a population of 170 million in an area of 148,460 square kilometers (57,320 sq mi) based on the most recent published census (2022). In this study, Bangladesh was chosen as the main research site because of its high population density with a poor quality of life as well as poor educational status. Therefore, it is of significant public health implication in the control and prevention of Infectious diseases in Bangladesh.

2.2. Data Collection

Data for this study, including hospital records and climate data, had been extracted from secondary sources. Weather data comprised monthly records of temperature, total precipitation and relative humidity from 2013 to 2022 across 35 stations in Bangladesh was obtained from the Bangladesh Agricultural Research Council (BARC). Data were averaged across the 35 stations to derive the monthly average temperature (in degrees centigrade), monthly average relative humidity and monthly average rainfall (in millimeters). Medical records of cases of Tetanus, Chickenpox and HIV were obtained from the Infectious Disease Hospital, Mohakhali, Dhaka, Bangladesh, where patients came to receive health care from all over the country. All the cases for Tetanus and HIV in the medicine ward from 2013 to 2022 was enrolled from the hospital archive. But cases for Chickenpox were obtained for the year 2015–2022. Checklist data collection procedure was adapted in this study. In the hospital, Tetanus and Chickenpox were diagnosed in men, women and children. Whereas males, females and transgender were all diagnosed with HIV. The reason for using monthly parameters was that the diseases incidences being too low at the weekly or daily scale coefficient.

2.3. Statistical Analysis

Consider Y_t, t=1, 2, ..., N to be a time series of counts taking non negative integer values, Y_t is the response process. Under the GLM framework, the probability density function of Y_t, given the past, is expressed as [11]:

$$f((y_t; \mu_t \mid F_{t-1})) = \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!}, t = 1, 2, ..., N \quad .... (2.3.1)$$

where, F_{t-1} denotes the information available to the observer up to time t. In this model, the conditional expectation of the response is equal to its conditional variance.

$$E(Y_t \mid F_{t-1}) = Var(Y_t \mid F_{t-1}) = \mu_t, t = 1,2,...,N \quad .... (2.3.2)$$

While the Poisson distribution is commonly used for modeling count data, it assumes that the mean and variance are equal. This assumption often does not hold when there is over-dispersion, where the variance exceeds the mean. To address this, we used the Negative Binomial (NB) regression model, which introduces an additional dispersion parameter to accommodate the extra variability.

Consider Y_t, t = 1, 2, ..., N, representing a time series of count data, where Y_t is the response variable. In the NB model, the conditional distribution of Y_t, given past information F_{t-1}, follows a Negative Binomial distribution with a mean μ_t and a variance of $\mu_t + \frac{\mu_t^2}{\phi}$, where ϕ is the dispersion parameter.

The probability density function of Y_t, given the past, is expressed as [12]:

$$f(y_t; \mu_t, \phi|\mathcal{F}_{t-1}) = \frac{\Gamma(y_t + \phi^{-1})}{\Gamma(y_t+1)\Gamma(\phi^{-1})}\left(\frac{\phi^{-1}}{\phi^{-1}+\mu_t}\right)^{\phi^{-1}}\left(\frac{\mu_t}{\mu_t+\phi^{-1}}\right)^{y_t} \quad .... \quad (2.3.3)$$

In this model, the expected value of Y_t, given past information, is μ_t, and the variance is:

$$E(Y_t|F_{t-1}) = \mu_t, \;\; Var(Y_t|F_{t-1}) = \mu_t + \frac{\mu_t^2}{\phi}, t= 1, 2, ..., N ....(2.3.4)$$

Decomposition of Tetanus, HIV and Chickenpox data from 2013 to 2022 (2015-2022, for Chickenpox) and analyzation of the four components (observed, trend, seasonal and random error) were conducted in order to identify the presence of trend and seasonal components in data. From the Augmented Dickey Fuller (ADF) test, Tetanus, HIV and Chickenpox were found to be stationary at 10% level of significance with p-values 0.07, 0.08 and 0.01 respectively. To determine the potential lag time of serial correlation of the series through data analysis, tools such as the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were utilized.

Several discrete time series regression models were constructed based on ACF and PACF plots of diseases where the first six models were with the short range serial dependence by first and second order autoregressive term (Y_{t-1} and Y_{t-2}) and seasonality captured by regressing the moving average components (λ_{t-11}, λ_{t-12} and λ_{t-13}) for different models. Harmonic regression is a type of linear regression model where the predictor variables are trigonometric functions of a time-related variable. Harmonic seasonal factors, such as sinusoidal terms (sin (2 * month/12) and cos (2 * month/12)), were utilized to account for seasonality in comparisons. General form of the formulated model is as follows:

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i \, Y_{t-i} + \sum_{j=11}^{q} \beta_j \, \lambda_{t-j} + \sum_{k=1}^{r}\left(\alpha_k \cos\frac{2\pi \acute{t}}{S_k} + \gamma_k \sin\frac{2\pi \acute{t}}{S_k}\right) + \epsilon_t$$

Where, Y_t denotes disease occurrences at time t, Y_{t-i} disease occurrences at lag i, i= 1, 2, 3, ..., λ_{t-j}=moving average components at lag j, j= 11, 12, 13, ..., $\cos\frac{2\pi \acute{t}}{S_k}$ and $\sin\frac{2\pi \acute{t}}{S_k}$ are the fourier terms (k=6 (Semi-annual period), 12(Annual period)). β's, α_k and γ_k are the coefficients to be estimated. A number of models were constructed incorporating the weather factors (monthly average temperature (T_t), monthly average relative humidity (H_t) and monthly average rain fall (R_t)) as covariates. Here, t= 1, 2, ..., 120.

From the fitted models, the best models were selected based on Akaike Information Criterion (AIC) and diagnostic checks. For each of the three diseases, two models with the lowest AIC values were chosen—one that excluded weather factors and one that included them. The goodness- of- fit of the models was assessed by using time series tools (to check autocorrelation functions of residuals). Estimates of the regression coefficients for the optimal model would be used to make conclusion on the weather effects on Tetanus, HIV and Chickenpox occurrences. Variant statistical techniques and statistical packages like R language version 4.1.2 and Microsoft Excel were applied to obtain the relative results of this study. For discrete time series analysis, "tscount" package (function "tsglm") was used with other packages in RStudio.

Finally, the models developed were verified by dividing the data file into two data sets. For Tetanus and HIV, the data spans from January 2013 to June 2021, serving as our training dataset. For Chickenpox, the training dataset covers the period from January 2015 to June 2021. To assess the impact of environmental factors on disease occurrences, the models' performance was compared using a testing dataset from July 2021 to December 2022. Additional validation methods, such as using 2- or 3-year subsets of data to test the model, were applied and produced similar outcomes.

The predictive accuracy of the models was also evaluated using the AIC, which assesses how well the model fits the data. AIC is calculated as AIC=n* ln (residual sum of squares/ n) + 2k), where "n" is the number of observations and k is the number of estimated parameters. Additionally, the root mean square error (RMSE) was used to assess model performance, with smaller RMSE values indicating better forecasting ability. The RMSE, which is the square root of the mean squared differences between actual and predicted values, serves as a reliable measure of model quality, including for generalized linear models (GLMs). These findings were further validated by plotting 'Observed vs Forecasted' values for both models.

3. Results

3.1. Description and Univariate Analysis

The mean of monthly Chicken-pox disease counts was the highest 15.688 with a standard deviation (SD) of 18.185 among other diseases. For Chicken-pox, the maximum and minimum illness counts were 91 and 0 respectively. On the other hand, HIV had the lowest mean monthly disease counts 5.558 with a maximum of 23 and a minimum of 0. The Tetanus counts varied between 3 to 33 with mean and SD of 12.442 and 5.135 respectively. In the first graph (Fig 1), the Tetanus counts primarily displayed cyclic behavior with significant fluctuations around 2016 but lacked a clear seasonal pattern and long term trend. The HIV counts in the second graph behaved similarly to the Tetanus counts. In the last one, there appeared a peak in Chicken pox cases around 2019 and a notable decrease in cases was observed around 2020. Another increase followed post-2020, with a peak nearing 2022. In each year, seasonal variation was quite noticeable. At the beginning of the year (February to May), the chicken pox counts remained really high. The graph reflected the cyclical nature of chicken pox incidence, which might be influenced by various factors such as vaccination rates, public health measures and natural disease cycles.

3.2. Development of Time Series Poisson and NB Regression Models

3.2.1. Analysis of ACF and PACF Plots

Decomposition of disease counts revealed that the observed data of Tetanus, HIV and Chicken Pox cases were affected by a mix of long-term trends, pronounced seasonal patterns and random variations. An important characteristic of infectious diseases is their tendency for past cases to influence current case numbers [13]. If a serial relationship exists between past and current cases, autoregression is included in the model. To determine the potential lag time of this serial correlation, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of monthly disease counts were constructed.

The ACF plots (Fig 2) for Tetanus, HIV and Chicken Pox revealed significant autocorrelations at the initial lags, indicating a strong immediate correlation with previous values. The ACF plot for HIV suggested underlying patterns that might be captured by AR or MA components. For Chicken Pox, the presence of higher autocorrelations at seasonal lags (multiples of the seasonal frequency) compared to other lags suggested potential periodic patterns in the data. To deal with these patterns, a dynamic regression with fourier terms was often preferable to the models. The autocorrelations at small lags were typically large and positive for each of the series indicating presence of trend in the dataset. The PACF plot for Tetanus suggested that an AR (1) model was appropriate, as only the first lag was significant. For HIV, the PACF plot showed significant autocorrelations at the first lag and smaller spikes at higher lags, indicating that an AR (1) model might be a good starting point, but additional terms may need to be considered. The PACF plot for Chicken Pox implied that an AR (2) model is suitable, with the first two lags being significant and capturing the autocorrelation structure effectively. To aid the longer-term trends and patterns by averaging past observations and smooths the noise yielding a more stable and reliable predictor, MA (3) would be considered for each of the series.

3.2.2. Selection of Optimal Models

Thirteen discrete time series regression models were formulated based on ACF and PACF plots of diseases where the first six models have been constructed taking into consideration the short range serial dependence by first and second order autoregressive term (Y_{t-1} and Y_{t-2}) and seasonality captured by regressing the moving average components (λ_{t-11}, λ_{t-12} & λ_{t-13}) for different models. Then seven additional regression models were constructed incorporating environmental factors (monthly average temperature.

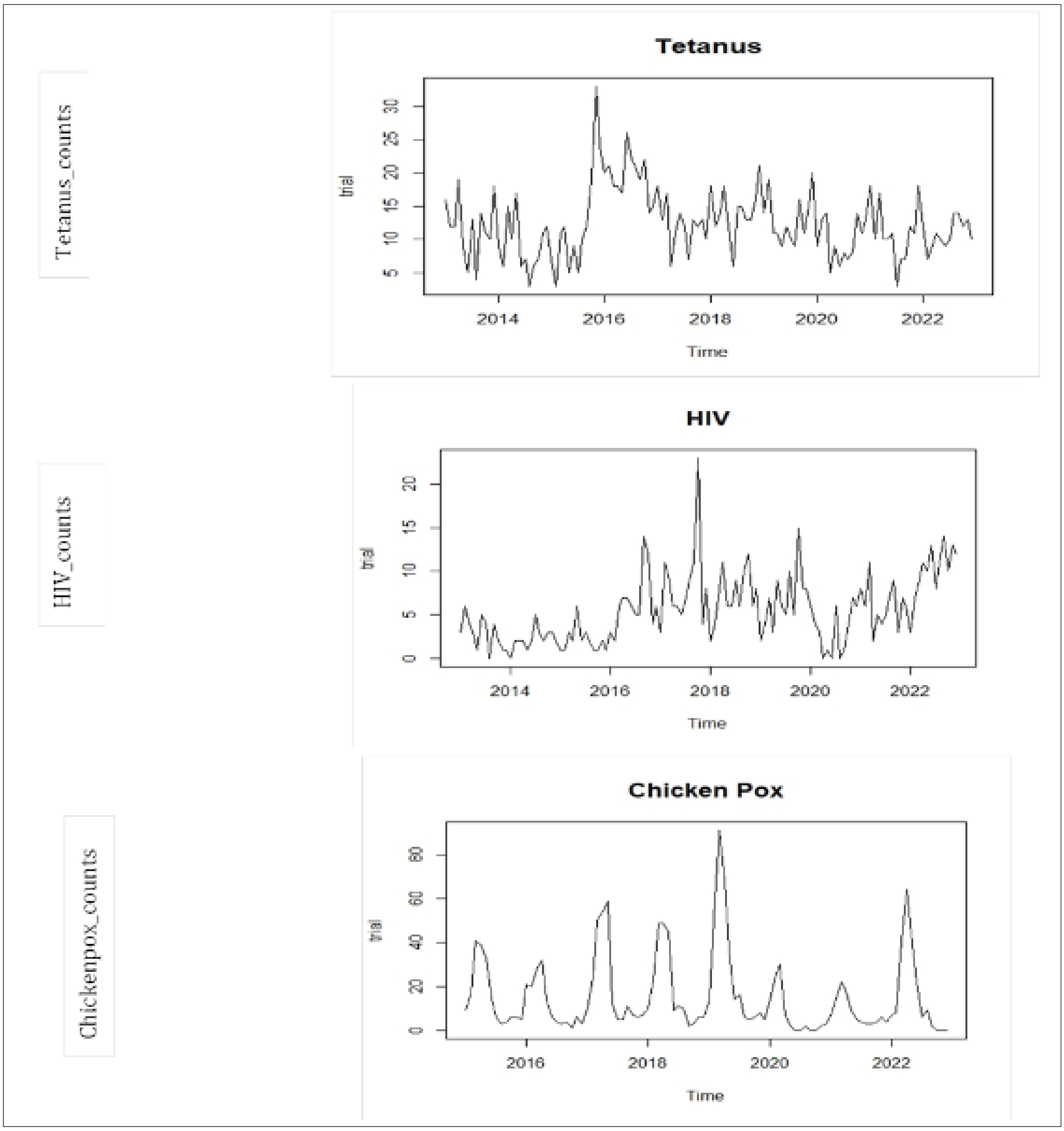

Figure 1 Time series plots of monthly disease counts for Chicken Pox, Tetanus, and HIV

Monthly average relative humidity and monthly average rain fall) as covariates based on the ACF and PACF plots for diseases.

All models were applied to each disease time series. The ten models with the lowest AIC values for each series were enlisted in the table.

The AIC values serve as a criterion for model selection, with lower values indicating a better balance between model complexity and goodness of fit. For Tetanus, the models with the lowest AIC values are M5 (697.64), M7 (697.93), and M13 (682.70), suggesting that M13 provided the best fit. For HIV, models M4 (609.61), M6 (609.01), M9 (605.96), M11 (603.57), and M13 (597.17) showed lower AIC values, with M13.

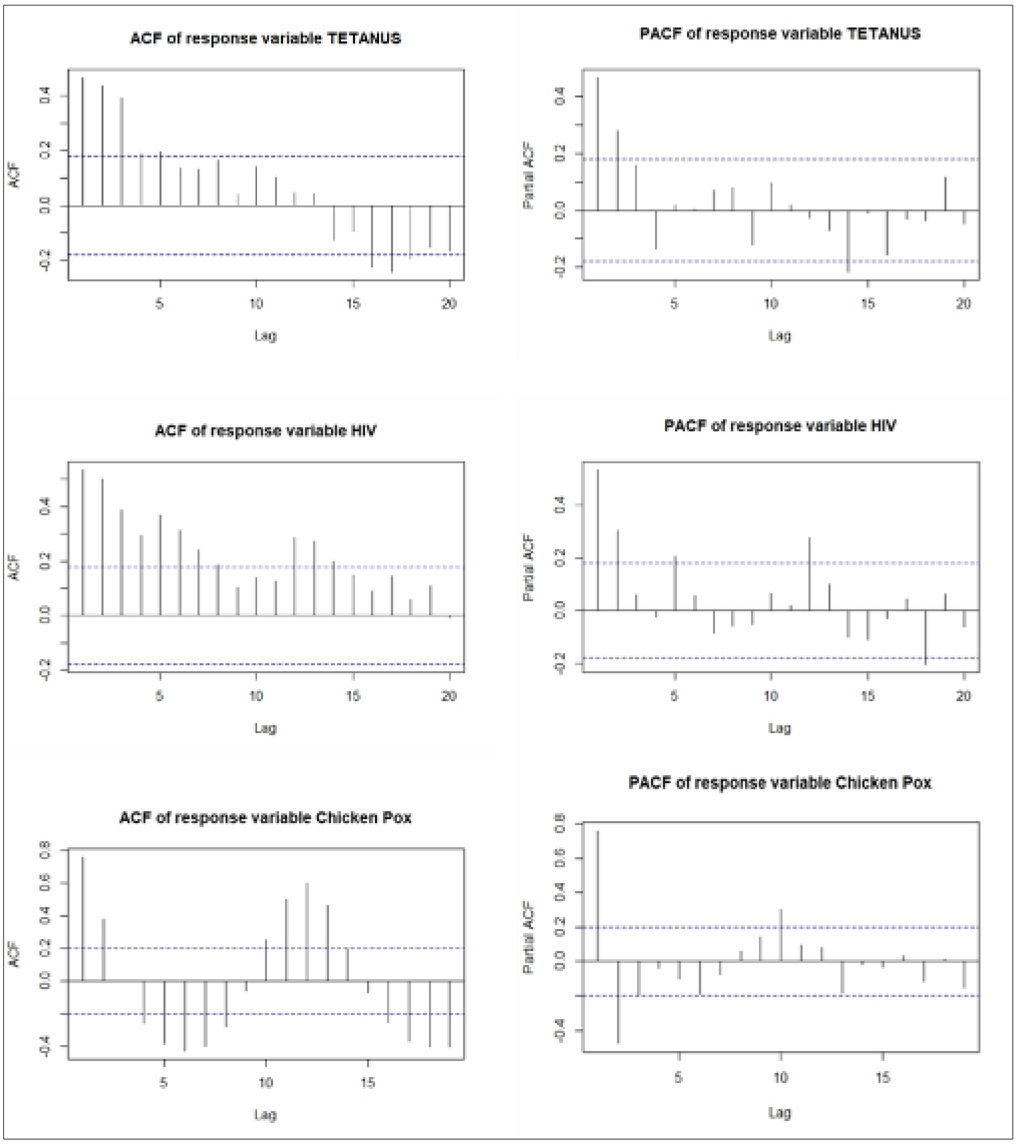

Figure 2 ACF and PACF plots for Diseases (e.g. Tetanus, HIV and Chicken pox)

Being the optimal model. For the analysis of Chicken pox counts, both Poisson and Negative Binomial (NB) regression models were fitted as overdispersion (mean and variance are not equal) was doubted to be there in the series from the descriptive part (mean= 15.688 and variance= 330.69). The results indicated that the AIC values for the NB models are smaller compared to those of the Poisson models. Additionally, the standard errors of the estimates in the NB models were lower than those in the Poisson models (Appendix). So, NB time series regression models were continued in which M6 (577.19), M9 (626.45), and M13 (608.01) models (for Poisson M13 model, AIC value=713.08) had the lowest AIC values, and M13 is the best fitting model. Overall, Model M13 consistently showed the lowest AIC values across all three diseases, indicating it was the most robust model among those evaluated.

Table 1 Comparison of the candidate models based on AIC values for three diseases

Models	**Equation with Coefficients**	**AIC Values**		
		Tetanus	**HIV**	**Chicken pox**
M1	$\beta_0 + \beta_1 Y_{t-1}$	707.12	612.96	639.88
M2	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \lambda_{t-13}$	716.60	613.79	638.06
M3	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \lambda_{t-12} + \beta_3 \lambda_{t-13}$	718.49	612.31	634.98
M4	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \lambda_{t-11} + \beta_3 \lambda_{t-12} + \beta_4 \lambda_{t-13}$	719.45	609.61	619.75
M5	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \cos\frac{2\pi t'}{12} + \beta_3 \sin\frac{2\pi t'}{12}$	697.64	-	-
M6	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 \cos\frac{2\pi t'}{12} + \beta_4 \sin\frac{2\pi t'}{12} + \beta_5 \cos\frac{2\pi t'}{6} + \beta_6 \sin\frac{2\pi t'}{6}$	-	609.01	577.19
M7	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 T_t + \beta_4 H_t + \beta_5 R_t + \beta_6 \cos\frac{2\pi t'}{12} + \beta_7 \sin\frac{2\pi t'}{12} + \beta_8 \cos\frac{2\pi t'}{6} + \beta_9 \sin\frac{2\pi t'}{6}$	697.93	-	-
M8	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 T_t + \beta_3 H_t + \beta_4 R_t + \beta_5 \cos\frac{2\pi t'}{12} + \beta_6 \sin\frac{2\pi t'}{12} + \beta_7 \cos\frac{2\pi t'}{6} + \beta_8 \sin\frac{2\pi t'}{6}$	-	614.73	610.61
M9	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 T_t + \beta_3 H_t + \beta_4 R_t$	704.63	605.96	626.45
M10	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \lambda_{t-13} + \beta_3 T_t + \beta_4 H_t + \beta_5 R_t$	706.64	-	-
M11	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 \lambda_{t-13} + \beta_5 T_t + \beta_6 H_t + \beta_7 R_t$	-	603.57	638.21
M12	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 \lambda_{t-12} + \beta_3 \lambda_{t-13} + \beta_4 T_t + \beta_5 H_t + \beta_6 R_t$	708.25	613.76	639.13
M13	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 \lambda_{t-11} + \beta_4 \lambda_{t-12} + \beta_5 \lambda_{t-13} + \beta_6 T_t + \beta_7 H_t + \beta_8 R_t$	682.70	597.17	608.01

T_t: average monthly temperature, H_t: average monthly humidity, R_t: average monthly rainfall, t=1, 2, ..., 120, t′= time of January, 2015= 25 (For Tetanus and HIV) and 1 (For Chicken pox)

3.2.3. Diagnostic Checking and Selection of the Best-Fitting Model

Two optimal models were selected for each disease series based on the AIC and ACF plot of the residuals: one model that excluded environmental factors and another that incorporated them. For Tetanus (Fig 2), while all models effectively captured the time-dependent structure of the data, the M5 and M13 models most significantly captured the autocorrelation structure. Graphical analysis of the residuals from these two models indicated that they fluctuated randomly around zero, showing no discernible trend in their variation. As M13 (AIC value= 682.70) had lower AIC value compared to M5 (AIC value= 697.64), it was the best fitting for Tetanus counts. It was an evidence (Table 2) that Model M13 included several significant Predictors as the confidence interval (95%) doesn't include zero. Notably autoregressive terms (Y_{t-1} and Y_{t-2}) and rainfall were the predictor variables (β_1, β_2 and β_8) which show significant effects on Tetanus occurrences. Although average monthly temperature and average monthly relative humidity did not significantly affect the occurrences, rainfall exhibited a small but significant negative effect ($\widehat{\beta_8} = -0.001$). Therefore, for an increase in rainfall there would be decrease in number of Tetanus occurrences. Standard errors and confidence intervals were (level = 95%) obtained by parametric bootstrapping with 500 replications.

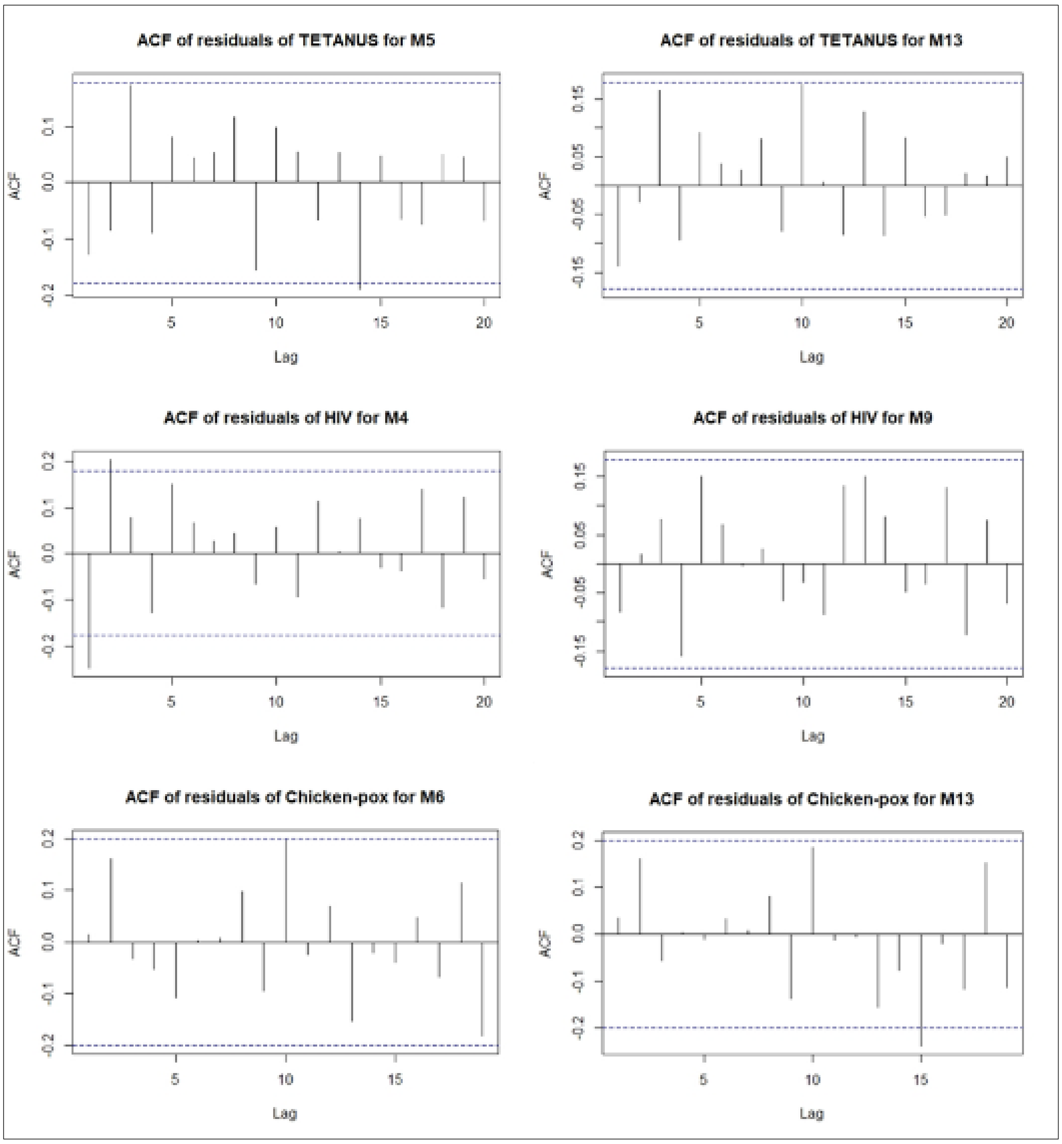

Figure 3 ACF plots of residuals excluding environmental factors (M5, M4 and M6) and including environmental factors (M13, M9 and M13) for diseases (e.g. Tetanus, HIV and Chicken pox)

For HIV, several models successfully captured the temporal dynamics of HIV disease data. Notably, models M4 and M9 demonstrated robust performance in capturing the autocorrelation structure. Based on the AIC values and analysis of the ACF plot of residuals (Fig 3) depicted that the model incorporating weather variables demonstrated superior fit compared to the model without these variables, as indicated by smaller AIC value 605.96. The selected model, M9, was subsequently fitted to the data, and its parameter estimates and other relevant statistical outputs were detailed (Table 3). Environment factors (e.g. average monthly temperature, average relative humidity, average monthly rainfall) had no significant effects on the occurrences of HIV disease at 5% level of significance. The first autoregressive term(Y_{t-1}) and linear trend had very small but significant effect on HIV incidences as their confidence intervals (95%) didn't include zero. Indeed, it's crucial to acknowledge that there might be other significant variables influencing HIV.

Table 2 Estimates of the Regression Coefficients for Model M13 for Tetanus

Coefficient of	Estimate	Std. Error	CI (Lower)	CI (Upper)
Intercept	1.412	0.752	0.823	3.73
Y_{t-1}*	0.369	0.098	0.058	0.441
Y_{t-2}*	0.219	0.099	0.010	0.373

λ_{t-11}	0.503	0.299	-0.422	0.694
λ_{t-12}	-0.689	0.387	-0.814	0.520
λ_{t-13}	0.029	0.259	-0.673	0.358
T	0.004	0.013	-0.022	0.0273
H	0.0002	0.006	-0.015	0.009
R*	-0.0012	0.001	-0.0011	-0.0001
Linear Trend	-0.011	0.015	-0.044	0.014

*: significant, T: temperature, H: humidity, R: rainfall

Table 3 Estimates of the Regression Coefficients for Model M9 for HIV

Coefficient of	Estimate	Std. Error	CI (Lower)	CI (Upper)
Intercept	-0.278	1.124	-2.712	1.767
Y_{t-1}*	0.346	0.107	0.091	0.497
Y_{t-2}	0.244	0.099	-0.003	0.379
T	0.027	0.016	-0.003	0.059
H	-0.002	0.013	-0.021	0.026
R	0.001	0.000	-0.001	0.001
Linear Trend	0.005	0.022	0.028	0.109

Table 4 Estimates of the Regression Coefficients for Model M13 for Chicken pox

Coefficient of	Estimate	Std. Error	CI (Lower)	CI (Upper)
Intercept	0.233	1.92	-0.360	6.908
Y_{t-1}*	0.998	0.139	0.532	1.000
Y_{t-2}	-0.274	0.134	-0.501	0.025
λ_{t-11}	0.315	0.175	-0.172	0.510
λ_{t-12}	-0.195	0.186	-0.453	0.301
λ_{t-13}	0.044	0.156	-0.316	0.295
T*	0.054	0.003	0.017	0.142
H*	-0.018	0.002	-0.098	-0.002
R	-0.001	0.000	-0.002	0.001
Linear Trend	-0.017	0.006	-0.151	0.097
Dispersion*	0.247	5.135	0.156	3.320

Occurrences that were not included in this particular research. It was noticeable that the models with moving average terms and fourier terms (M6 and M13) captured the autocorrelation structure better than the other models considered without environmental factors. Models M6 and M13 were the two models based on AIC values (577.19 and 608.01 respectively) and ACF plots of residual (Fig 3), those randomly fluctuate around zero with no discernible trend in the variation of residuals as incidence values increase. Based on the AIC values M6 was the best fitted model. To evaluate the effect of environmental factors, M13 was chosen and then applied to the data. From the resulting estimates for its parameters and other pertinent statistical findings (Table 4), the environment factors (e.g. average monthly temperature, average relative humidity) had significant effects on the occurrences of Chicken pox at 5% level of significance. The estimated value of regression coefficients for monthly average temperature was ($\widehat{\beta_6}$=0 .054); it had positive effect with Chicken pox. Therefore, for a unit increase in temperature there would be increased number of Chicken pox occurrences. Besides there was a significant negative effect ($\widehat{\beta_7}$= -0.001) of relative humidity on the disease occurrences, that was for an unit increase in relative humidity there would be a decrease number of disease occurrences. The first autoregressive term(Y_{t-1}) had significant large effect on Chicken pox incidences as its confidence intervals (95%) didn't include zero. The estimated value of dispersion parameter of the Negative Binomial distribution was 0.247 which was though greater than zero but a small value. A small value like it indicated that there was still some ignorable overdispersion which wasn't captured in our model.

3.2.4. Evaluation of Model Accuracy

For Tetanus and HIV, the data spanned from January 2013 to June 2021, serving as training dataset. For Chicken Pox, the training dataset covered the period from January 2015 to June 2021. To assess the impact of environmental factors on disease occurrences, the models' performance was compared using a test dataset from July 2021 to December 2022.

Table 5 Forecasting Accuracy for all Three Diseases

Diseases	Models	RMSE	MAPE
Tetanus	M5	3.83	54.23%
	M13	3.33	47.16%
HIV	M4	0.5	6.94%
	M9	0.22	3.08%
Chicken pox	M6	0.67	17.59%
	M13	1.67	43.98%

Incorporation of environmental factors as covariates improved forecast performance for Tetanus and HIV (Table 5). Models M13 and M9, RMSE of 3.33 and 0.22, exhibited lower RMSE and MAPE values compared to models M5 (for Tetanus) and M4 (for HIV) respectively. A lower RMSE signified a more accurate model in terms of forecasting ability. Therefore, models M13 and M9 were deemed effective for predicting future time points, given their inclusion of environmental factors. In contrast, for Chicken Pox, model M6 outperformed model M13 in forecast performance with RMSE values of 0.67 and 1.67 respectively, as evidenced by a lower RMSE for M6. However, average monthly temperature and average relative humidity significantly influenced Chicken pox occurrences. According to model M6, the autoregressive term (Y_{t-1}) and seasonality (represented by Fourier terms) had a significant impact on Chicken Pox occurrences.

Now the selected models were fitted to the training dataset and forecasted occurrences for the next eighteen months (July 2021-December 2022). To evaluate forecast accuracy, these forecasts with the test dataset were compared. To visualize the models' adequacy, 'Observed vs Forecasted' values were plotted. The following (Fig 3) was the plots of observed vs fitted values for July 2021-December 2022.

4. Discussion

The results of this study suggested that the weather variability might have played a significant role in the transmission of infectious diseases (particularly Tetanus and Chicken pox) in Bangladesh either directly or through other unmeasured variables. The key determinants of the Tetanus transmission observed in this analysis included monthly average rainfall whereas Chicken pox was significantly influenced by both monthly average temperature and monthly average relative humidity. In contrast, environmental factors did not significantly affect HIV occurrences. A study in Chongquing reported that environmental factors had significant effects on contagious diseases [14]. To the best of our knowledge, this study is the first to conduct a comprehensive analysis of the long-term effects of monthly weather variations on the incidence of Tetanus and HIV utilizing the data from 2013 to 2022 in Bangladesh. A study in Jinan reported that environmental factors had significant effects on chickenpox and it was found that temperature had negative effect on chickenpox incidence [15]. Another study in West Indies showed that a chickenpox infections occur in tropical regions than in temperate regions [16]. Temperature was found to have significant positive effect on chickenpox incidence in our study. A study in Japan showed that chickenpox incidence increased at 5–20 °C which is the activation range of it and decreased at temperatures lower than 5 °C and higher than 20 °C [17]. Additionally, according to a study in Wuhan and Hong Kong of China with data from 2008 to 2015, it was found that mean annual temperature had positive correlation with chickenpox

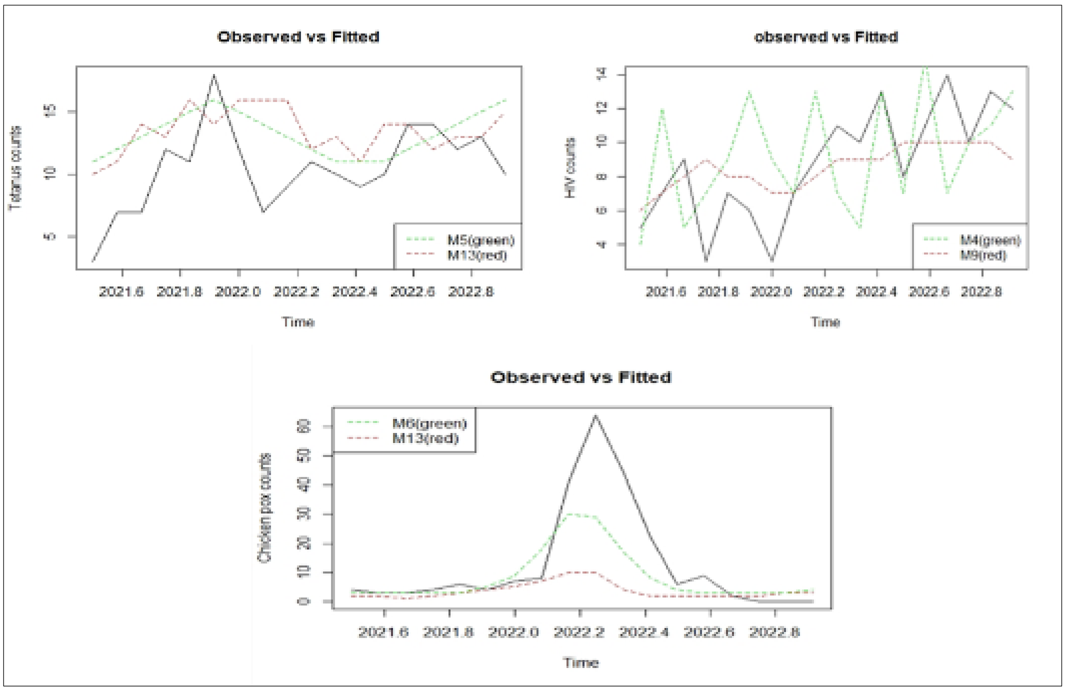

Figure 4 Plots of Observed Vs Forecasted values for July (2021) - December (2022)

Incidence [18]. Relative humidity was found to have negative effect on chickenpox incidence in our study and similar findings have been observed in Hong Kong [19] and the West Indies [16]. The precise mechanism underlying the potential link between relative humidity and the incidence and transmission of chickenpox remains unclear. It could be hypothesized that lower relative humidity might result in smaller air particles, allowing the varicella-zoster virus (VZV) to remain airborne for extended periods.

The time series models those allowed for autoregressive and moving average structure along with weather factors in the series, appeared to be more suitable in the assessment of the relationship between the weather variables and transmission of diseases with respect to goodness of fit, conformance with assumptions and predictive accuracy. The time series Poisson generalized linear model showed that temperature, relative humidity and rainfall could improve goodness of fit for Tetanus and HIV. The model with autoregressive term (at lags of 1 to 2 months) and moving average terms (at lags 11 to 13 months) for Tetanus and only with autoregressive terms (at a lag 1 month) for HIV had the lowest AIC values. To improve the goodness of fit a time series Poisson regression model with autoregressive term at lag 1, maximum temperature at lags of 1 to 3 months and relative humidity at a lag of 1 month were included for the determination of the cryptosporidiosis transmission Brisbane, the capital of Queensland State [20]. Given that data were over-dispersed, a negative binomial model rather than a Poisson model was selected to estimate the association between environmental factors and chickenpox incidence in our study where the environmental factors with autoregressive terms (at lags of 1 to 2 months) and moving average terms (at lags 11 to 13 months) had been included. Similar selection was made for the study on Jinan and Chongqing, China, to observe the meteorological effects on chickenpox incidence where different meteorological variables (temperature, relative humidity, rainfall, sunshine, atmospheric pressure and wind velocity) with autoregressive terms at lags of 1 to 3 months were included [14-15].

The long-term effect of environmental factors on infectious disease incidences indicated by this study also deserves attention. This study demonstrated the obvious presence of impact of temperature and humidity on Chickenpox and rainfall on Tetanus. There was no significant effect of environmental factors on HIV. The predictive accuracy for different models was evaluated by RMSE and MAPE values. Models with environmental factors for Tetanus and HIV were seemed to be well fitted. For Chickenpox, though model with environmental factors had lowest AIC values and better ACF for residuals compare to model without the factors, predictive accuracy showed the different pictures. This could be because of not considering other important covariates in this study.

This study relied on secondary data spanning from 2013 to 2022, covering a relatively short timeframe of 10 years. Disease counts were only available from 2013 (from 2015 for Chicken Pox) and were recorded in a handwritten notebook, not yet digitalized. Consequently, only monthly counts, sex and district variables could be accessed, which was a significant limitation. Although the selected predictors included various climate factors, other variables such as solar sunshine, wind velocity, air pressure etc. might significantly impact disease occurrences. The exclusion of these factors might lead to incomplete or slightly biased estimations.

5. Conclusion

Despite the limitations, this study demonstrated that environmental factors had significant effects on infectious disease incidences. It was clear, though rainfall had significant effect on Tetanus, Chickenpox occurrences affected by temperature and relative humidity. No significant effect of any of these three variables was found on HIV. The models were well fitted and it was evident from the goodness of fit and good predictive accuracy. In this sense, the findings of this study can be instrumental in developing precautionary measures and strategic interventions to reduce the incidence of these three diseases. Various stakeholders, including non-governmental organizations, international bodies and the government of Bangladesh can utilize the results to inform public health planning and disease prevention strategies. However, their awareness, prevention and treatment activities for these three infectious diseases should be scaled up to minimize their occurrences in Bangladesh.

Compliance with ethical standards

Acknowledgments

This research received funding through a National Science and Technology (NST) fellowship awarded by the Ministry of Science and Technology. Our heartfelt appreciation and thanks goes Dr. Ariful Bashar and Dr. Sribash Paul, Medical Officers at the Infectious Disease Hospital in Mohakhali, Dhaka, Bangladesh, for their invaluable assistance in providing the disease data for this research.

Author Contributions

Conceptualization: Khondoker Rukyath Alam, Mohammad Ahsan Uddin
Methodology: Khondoker Rukyath Alam
Formal Analysis: Khondoker Rukyath Alam
Software: Khondoker Rukyath Alam
Writing—Original Draft: Khondoker Rukyath Alam
Writing—Review & Editing: Khondoker Rukyath Alam
Supervision: Mohammad Ahsan Uddin

Disclosure of conflict of interest

All authors declare no competing interest.

References

[1] Docherty B, Giannini T. Confronting a rising tide: a proposal for a convention on climate change refugees. Harv. Envtl. L. Rev.. 2009;33:349.

[2] Hay SI, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulkader RS, Abdulle AM, Abebo TA, Abera SF, Aboyans V. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017 Sep 16;390(10100):1260-344.

[3] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England journal of medicine. 2020 Mar 26;382(13):1199-207.

[4] Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS neglected tropical diseases. 2019 Mar 28;13(3):e0007213.

[5] World Health Organization. Varicella and herpes zoster vaccines: WHO position paper, June 2014. Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire. 2014;89(25):265-87. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids

[6] UNICEF. Varicella outbreak in Bangladesh sees a sharp rise in cases. UNICEF Bangladesh Health Report. 2019.

[7] Heywood AE, Macartney KK, McIntyre PB, Grant KA. Varicella and its complications in Australia, 1998–2013. Commun Dis Intell Q Rep. 2014;38(4).

[8] Tauhid SA, Rahman MM, Chowdhury MM, Khan MSR. Impact of meteorological factors on the spread of infectious diseases in Bangladesh: A case study of COVID-19. BMC Public Health. 2021;21(1):1-12. doi:10.1186/s12889-021-11298-5.

[9] World Health Organization (WHO). Global HIV & AIDS statistics — Fact sheet. WHO HIV Report. 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids

[10] UNAIDS. Global HIV & AIDS statistics 2023. UNAIDS Report. 2023. Available from: https://www.unaids.org/en/resources/fact-sheet

[11] McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. London: Chapman & Hall; 1989.

[12] Hilbe JM. Negative Binomial Regression. 2nd ed. Cambridge: Cambridge University Press; 2011.

[13] Yien YJ, Chu CM, Wen TH. Temporal and spatial dynamics of dengue fever in Kaohsiung City, Taiwan. Am J Trop Med Hyg. 2012;86(5):884-894. doi:10.4269/ajtmh.2012.11-0513.

[14] Wang Y, Xu C, Ren J, Zhao Y, Li Y, Wang L, Yao S. The long-term effects of meteorological parameters on pertussis infections in Chongqing, China, 2004–2018. Scientific Reports. 2020 Oct 14;10(1):17235.

[15] Yang Y, Geng X, Liu X, Wang W, Zhang J. Association between the incidence of varicella and meteorological conditions in Jinan, Eastern China, 2012–2014. BMC infectious diseases. 2016 Dec; 16:1-8.

[16] Garnett GP, Cox MJ, Bundy DA, Didier JM, Catharine JS. The age of infection with varicella-zoster virus in St Lucia, West Indies. Epidemiology & Infection. 1993 Apr;110(2):361-72.

[17] Harigane K, Sumi A, Mise K, Kobayashi N. The role of temperature in reported chickenpox cases from 2000 to 2011 in Japan. Epidemiology & Infection. 2015 Sep;143(12):2666-78.

[18] Chen B, Sumi A, Wang L, Zhou W, Kobayashi N. Role of meteorological conditions in reported chickenpox cases in Wuhan and Hong Kong, China. BMC infectious diseases. 2017 Dec; 17:1-9.

[19] Chan JY, Tian L, Kwan YW, Chan W, Leung CW. Hospitalizations for varicella in children and adolescents in a referral hospital in Hong Kong, 2004 to 2008: a time series study. BMC Public Health. 2011 Dec; 11:1-7.

[20] Hu W, Tong S, Mengersen K, Connell DE. Weather variability and the incidence of cryptosporidiosis: comparison of time series poisson regression and SARIMA models. Annals of epidemiology. 2007 Sep 1;17(9):679-88.