ASSESSING MALARIA AND TYPHOID FEVER TRENDS USING CORRELATION AND COVARIANCE: CASE STUDY OF ADAMAWA REGION (CAMEROON)

Manuscript Info

……………………….

Manuscript History

Received: xxxxxxxxxxxxxxxx
Final Accepted: xxxxxxxxxxxx
Published: xxxxxxxxxxxxxxxx

Key words:-

Health Time Series, Correlation, Covariance, Malaria and Typhoid Fever Dynamics, Adamawa Region of Cameroon.

Abstract

………………………………………………………………

Malaria and Typhoid Fever are two diseases classified as potentially epidemiological in Cameroon, and where cases of coinfection are often reported in Health Facilities. To assess the degree and direction of this interdependence, correlation and covariance are specifically used in this work. A set of statistical approaches is applied using the Python programming language to a dataset of weekly cases for both diseases in the Adamawa Region of Cameroon, spanning from January 2021 to December 2024 (four years). The proposed analytical framework encompasses graphs and algebraic approaches to correlation, including cross-correlation, cross-covariance, and their corresponding time lags, as well as rolling window functions. First and foremost, the stationarity of each series is examined. The values obtained for the correlation coefficients are 0.73 for Pearson and 0.63 for Spearman, both of which exceed 0.5, indicating strong correlations. There is a strong peak at lag 0 for cross-correlation, suggesting a significant contemporaneous relationship. The time lag cross-correlation consistently shows high values (between 0.8 and 1) for all lags. At lag zero, the series vary together and the time lag cross-covariance remains above zero. Overall, the two diseases exhibit the same directionality with an immediate correlation, and peaks are explicitly observed in mid-2023 and the beginning of 2024. This work provides statistical knowledge for both the population and stakeholders, helps predict disease trends, and informs strategies for the joint management of the diseases. It opens up ways for examining causalities and multivariate analysis.

……………………………………………………………………………………………….....

Introduction:-

In Cameroon, around twenty diseases are classified as potentially epidemiological, requiring close monitoring to anticipate any large-scale contamination [1]. Among them, malaria [2] and typhoid fever [3] appear to be two predominant infectious diseases that substantially affect population health, and where cases of coinfection are regularly encountered in Health facilities [4]. Various measures are then undertaken, including weekly data collection on occurring cases, for disease monitoring.

To carry out this statistical assessment, the paper focuses on health data considered as a time series [5]. Time series refers to a sequence of events observed and recorded over a period of time [6], [7]. The Adamawa Region is one of the ten regions in Cameroon, located at the crossroads between the South and the North of the country. With a population of about 1.18 million and an area of 63,701 km2, the Region is bordered on the West by Nigeria and on the East by the Central African Republic. The climate is temperate, and its savannah vegetation is situated in a hilly area, making the Region a suitable sample for these experiments [4], [8].

The primary motivation for this paper is, firstly, to pursue works undertaken on epidemiological prevention using time series data and methods. The second motivation stems from the observation that several cases of coinfections are frequently reported, which requires a better understanding of some factors, including the degree of correlation and covariance, the co-evolution, the causality and so forth [9]. Lastly, research has revealed several studies based on the analysis of malaria and typhoid fever coinfections [4], [10], but very few on their interdependence. The challenge is to fill this gap and provide the various stakeholders with more statistical data on which to base decisions and actions.

The primary purpose of this work is therefore to carry out an exhaustive and comprehensible statistical analysis of both malaria and typhoid fever in the Adamawa Region of Cameroon, based on an approach involving correlation and covariance [11], [12]. This objective involves collecting data over a significant period for experiments, followed by statistical analysis using the selected approaches, and ultimately providing valuable insights and recommendations for stakeholders and decision-makers.

Various studies focused on statistical analysis of disease-related time series data. The subsequent paragraph presents some relevant ones.

To assess Google Trends' accuracy for epidemiological surveillance of dengue and yellow fever and compare their incidence on the population of São Paulo state, the work in [13] was carried out. The correlation was calculated using Pearson's coefficient and the cross-correlation function. The study in [14] investigated the transmissibility and death distribution of COVID-19 and its association with meteorological parameters to study the propagation pattern of COVID-19 in UK regions. The correlation and regression analysis between COVID-19 variables and meteorological parameters was performed. To identify potential predictors of new health system overloads, [15] analysed Twitter and emergency services data, comparing it to daily infected time series through wavelet and cross-correlation analysis. Using real-world data and machine learning models, [16] conducted a retrospective study from 2010 to 2020 to analyse the trends and characteristics of Multidrug-resistant bacteria (MDRB) infections. Combining 39 hospital indicators, the authors used a random forest model and cross-correlation analysis. The study's aim in [4] was to determine the prevalence of malaria and typhoid fever, as well as their coinfection among febrile patients at Ngaoundere Regional Hospital, Adamawa, Cameroon. A cross-sectional and descriptive study was conducted on 208 febrile patients suspected of Malaria and/or typhoid fever from September to November 2019. A similar work was conducted in a University Hospital in Nigeria by different authors in [10]. In [17], correlation tools were applied to open-source COVID-19 data from different countries. A longitudinal time series study was carried out with a cross-correlation analysis of Temporary Incapacity (TI) and COVID-19 cases, as reported by the work of [18]. [8] used weekly collected surveillance data from health facilities in the Adamawa Region from January 2018 to December 2022 and applied key statistical metrics for central tendency, data spread, distribution shape, and variable dependence. The objective in [19] was to identify and estimate the autocorrelation and cross-correlation of time series of hospitalisation rates for syphilis and HIV/AIDS in the State of Bahia from 2000 to 2020 by using Detrended Fluctuation Analysis (DFA) and cross-correlation coefficient.

The main contribution of this work is a comprehensive description of the correlation and covariance of the diseases, based on a relevant set of applied statistical approaches [5], [20]. This work introduces others on causalities and multivariate analysis.

2. Materials and methods

The present work aims to analyse two disease-related time series. Stationarity is a key property to check before starting a statistical assessment of a time series.

2.1 Stationarity of time series

For significance correlation analysis, the time series should be stationary, meaning that their statistical properties (mean, variance, autocovariance) are constant over time [21]. Non-stationary series can produce misleading correlation results and poor forecasts [22]. Several statistical tests assess stationarity in a time series. Among them, the Augmented Dickey-Fuller (ADF) test tests the null hypothesis that a unit root is present in the time series, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test assesses the null hypothesis that the time series is stationary around a deterministic trend. If a non-stationarity is found in a time series, some techniques can be employed to transform it. These techniques include differencing (subtracting the previous observation from the current observation) and detrending (removing trends from the data) [23], [24].

2.2 Correlation and cross-correlation

Correlation of time series refers to the statistical relationship between two or more time series, indicating how changes in one series relate to changes in another over time [25]. Understanding this correlation is crucial for analysing and predicting the behaviour of interrelated time series.

There are two ways to assess time series correlation: graphs and algebraic approaches. The graphs approach includes time series plots and scatter diagrams. Meanwhile, algebraic approaches are based on coefficients of correlation [12].

The first step in testing for correlation between time series is to plot them in a common plan or referential and inspect their appearance and aspect [25]. The scatter diagram is a graphical representation of the relationship between two quantitative variables [26], [27]. For a positive correlation, points trend upwards from left to right, indicating that as one variable increases, the other also increases. A negative correlation shows a downward trend in points from left to right, indicating that as one variable increases, the other decreases. No correlation is when the points are scattered randomly, revealing no discernible relationship between the variables. A trend line (or line of best fit) is added to summarise the relationship between variables. This line helps to visualise the general direction of the data and is considered a regression line [28].

Besides, algebraic approaches include coefficients of correlation, statistical measures that quantify the strength and direction of the linear relationship between two series [29]. The most common measure is Pearson's correlation coefficient, which ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation) [30]. A value close to zero indicates no correlation, showing that the series do not move together. The Pearson correlation coefficient is calculated using the formula:

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

where n is the number of pairs, x and y are correlated variables.

The other correlation measure used is Spearman's rank coefficient, a non-parametric measure that assesses how well a monotonic function can describe the relationship between two variables [31]. It also ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation). A value around 0 exhibits no predictable relationship between the variables [32]. The coefficient is obtained via the formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (2)$$

where d_i is the difference between the ranks of each pair of observations, and n is the number of observations.

The Cross-correlation function (CCF) measures the correlation between two series as a function of the time lag applied to one of them [33]. The cross-correlation at lag k is mathematically expressed as:

$$C(k) = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(Y_{t+k} - \bar{Y})}{\sqrt{\sum_{t=1}^{n-k}(X_t - \bar{X})^2 \sum_{t=1}^{n-k}(Y_{t+k} - \bar{Y})^2}} \quad (3)$$

$\bar{X}$ and $\bar{Y}$ are the means of the series X and Y, respectively, and n is the number of observations. A positive value of CCF indicates that as one time series increases, the other tends to increase after the specified lag. A negative CCF suggests that as one series increases, the other decreases after the specified lag [34], [35].

The time lagged cross correlation (TLCC) function measures the correlation between two series at different time lags [36]. This technique helps identify how one time series may influence or relate to another over time, accounting for potential relationship delays.

The rolling windowed correlation (RWC) function computes the correlation coefficient over a moving window, providing insights into how the relationship between the series evolves [37], [38].

2.3 Covariance and cross-covariance

The covariance of the two series measures how much they change together [9]. It can take any value and is calculated using the following formula:

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \qquad (4)$$

where n is the total number of observations in the time series, $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively [39]. A positive covariance indicates that the two series tend to increase or decrease together, while a negative covariance suggests that when one time series increases, the other tends to decrease. A covariance close to zero implies no relationship between the series' movements.

Cross-covariance extends the concept of covariance and measures the relationship between two series at different time lags applied to one of them [9]. Itis a statistical measure that assesses the degree to which two series change together over time.

Time lag cross-covariance measures the joint variability of two series at different lags [40], [41]. It helps to identify how one time series may influence or relate to another over various time delays.

Rolling cross-covariance is used to analyse the time-varying relationship between two series over a specified window.

2.4 Data and programming environment

The dataset used encompasses weekly cases of malaria and typhoid fever from Health Districts of the Region, stored via an online platform[1] and managed by the Health Information Unit of the Ministry of Public Health. The data, aggregated at the region level from January 2021 to December 2024, comprise 208 records used for experiments.

To perform experimentations, the scientific programming language Python[2] is used via Google Colaboratory. It is adapted for statistics, through several specialised libraries including Statistics for descriptive statistics; Pandas for numerical computing; Matplotlib combined with Seaborn for graphics and data visualisation[42].

2.5 Methodology

The methodological approach defined involves six main stages:

1. Data collection and data set construction;
2. Stationarity tests;
3. Statistical description of the data set;
4. Correlation analysis:
   - Graphs approach (time series and scatter diagrams plot);
   - Algebraic approach (Pearson and Spearman coefficients);
5. Cross-correlation, time lag cross-correlation and rolling correlation analysis;
6. Cross-covariance, time lag cross-covariance, and rolling covariance analysis.

3. Results

We assume that the dataset is already built.

3.1 Stationarity tests of time series

The stationarity test for the malaria series reveals a non-stationary with a stochastic trend, giving a p-value of 0.20 for ADF. However, the series is stationary in a deterministic trend with a p-value of KPSS = 0.10. In order to preserve memory as much as possible and render the series stationary, fractional differentiation is used instead of integer one [21]. The following values are obtained:

Differentiation order: 0.20, ADF p-value: 4.70 %, Correlation with original series: 0.93.

For the typhoid fever series, the tests indicate full stationarity: *ADF p-value = 0.00, KPSS p-value = 0.10.*

3.2 Data description

Table 1 contains the basic statistical properties of the series.

Table 1*: Descriptive statistics of series*

Indicator	Malaria	Typhoid fever
Mean	1813.55	736.21
Standard deviation	609.87	193.28
Minimum	11.98	513.00
1st quartile	1407.39	622.75
2nd quartile	1697.08	677.50
3rd quartile	2069.42	760.25
Maximum	3941.31	1609.00
Kurtosis	1.51	4.35
Skewness	0.97	2.08

3.3 Correlation and covariance analysis

The first assessment of the correlation is the graph approach. Figure 1 depicts the joint curves for Malaria and Typhoid Fever, and the fractionally differentiated version of the malaria series.

[1]https://dhis-minsante-cm.org/
[2]www.python.org

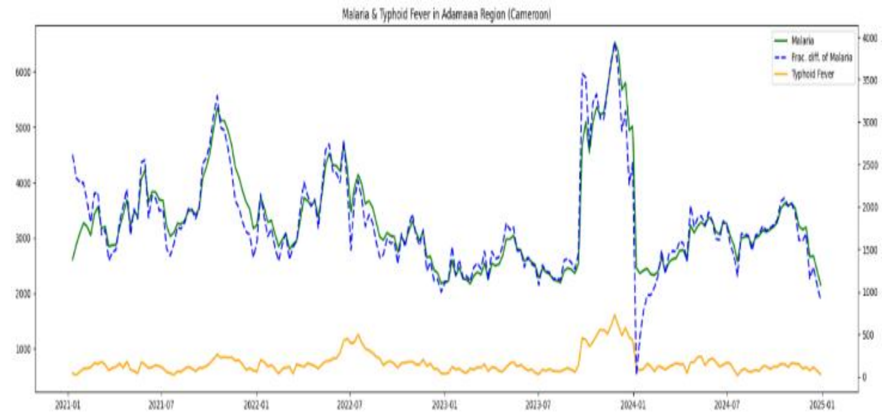

Figure 1*: Malaria and Typhoid Fever graphs*

Figures 2 and represent a scatter plot of the variables, associated with their regression line.

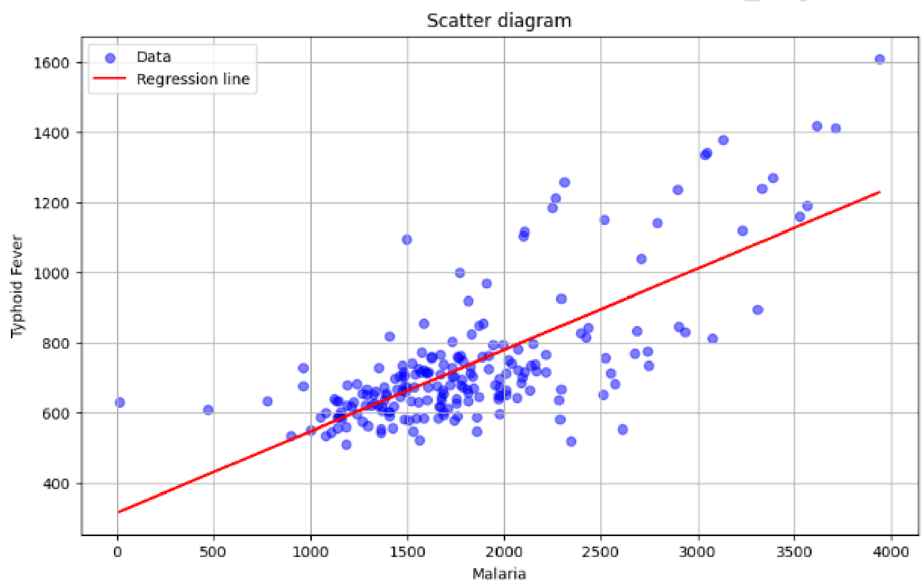

Figure 2*: Regression of Malaria over Typhoid Fever*

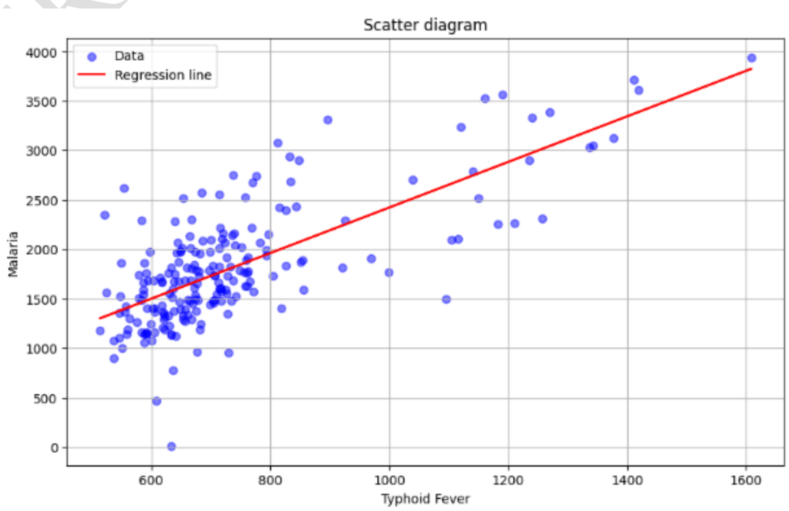

Figure 3*: Regression of Typhoid Fever over Malaria*

For Figure 2, the slope of the curve is *0.23*, and the intercept is *316.30*. The equation for the regression curve is therefore:

Typhoid Fever cases = 0.23(Malaria cases) + 316.30*.

The slope of the curve for Figure 3 is *2.30*, and the intercept is *116.46*. Thus, the equation of the regression curve obtained is:

Malaria cases = 2.30(Typhoid Fever cases) + 116.46*.

The coefficient of determination R^2 for predicting Malaria cases from linear regression is $R^2 = 0.53$, slightly higher than the one from the AutoRegressive Moving Average (ARIMA) prediction: $R^2 = 0.27$. This result suggests that linear regression can be a viable option for estimating future cases.

Concerning the algebraic approach for the two series taken together, the values of the correlation coefficients are *0.73* for Pearson and *0.63* for Spearman. They are all above 0.5, unveiling strong correlations between series.

The cross-correlation, time lag cross-correlation, and rolling correlation functions produce the diagrams in Figures4, 5 and 6. The curves are symmetric for both series, so calculating one is sufficient for analysis.

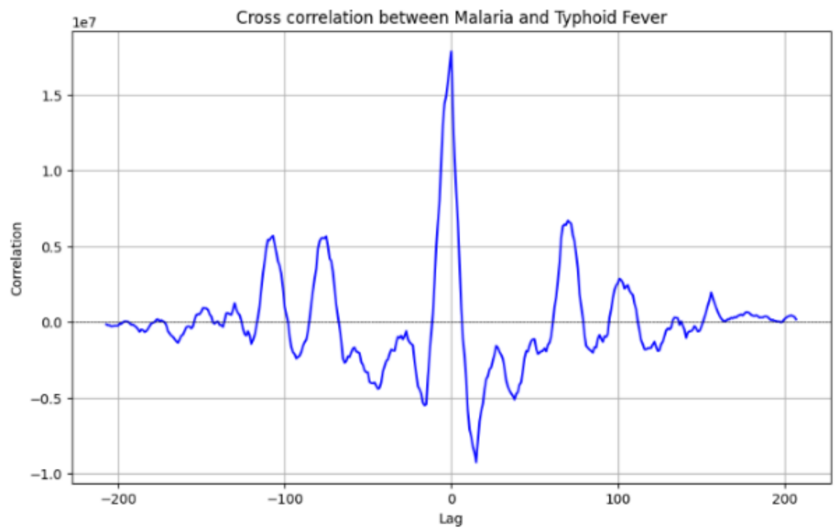

Figure 4: *Cross-correlation*

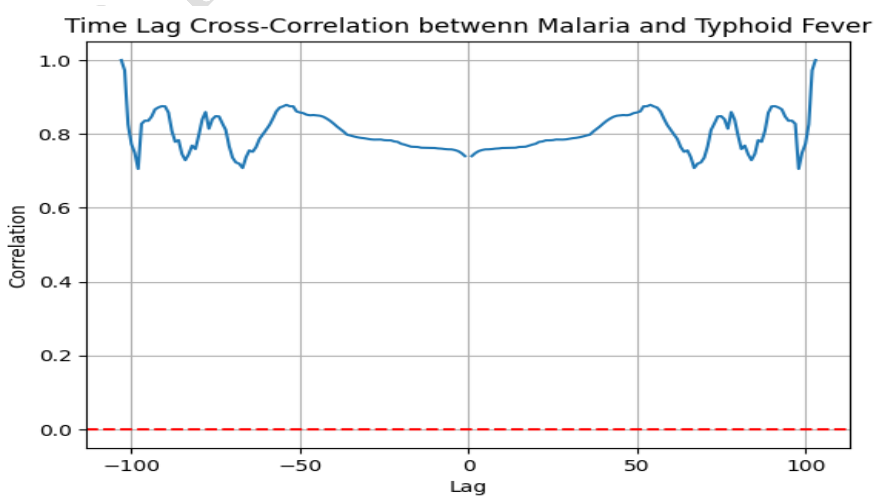

Figure 5: *Time lag cross-correlation*

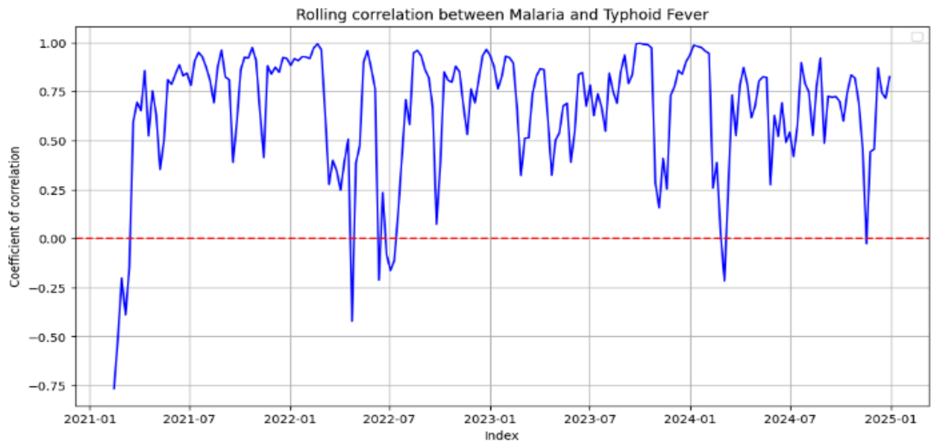

***Figure 6**: Rolling correlation*

Figure 4 shows the cross-correlation between Malaria and Typhoid Fever. The lag values of Typhoid Fever range from -200 to 200, indicating how the correlation changes over time, both before and after the current observation of Malaria. The larger lags have been chosen to appreciate the changes over the period better. The strength of the correlation at each lag is sometimes above zero, showing a positive correlation at this specific lag, or under zero, revealing a negative correlation. The strong peak at lag 0 suggests a significant contemporaneous correlation between Malaria and Typhoid Fever, meaning that when cases of one disease are high, the other cases are also simultaneously high. The time lag cross-correlation is presented in Figure 4, with values ranging from 0 to 100. The plot shows consistently high correlation values (around 0.8 to 1) across most lags, suggesting a strong positive relationship between Malaria and Typhoid Fever over time when the two series are shifted. Figure 6 displays the rolling correlation plot over the studied period. The window size is 6, representing the two series' minimum Autocorrelation function (ACF). There are periods where the correlation coefficient approximates 1, suggesting a strong positive relationship. According to the plot, the relationship is generally positive.
The cross-covariance, time lag cross-covariance, and rolling covariance functions yield the diagrams of Figures7, 8 and 9. Similarly, the curves are symmetric for both series, so calculating one is sufficient to perform the analyses.

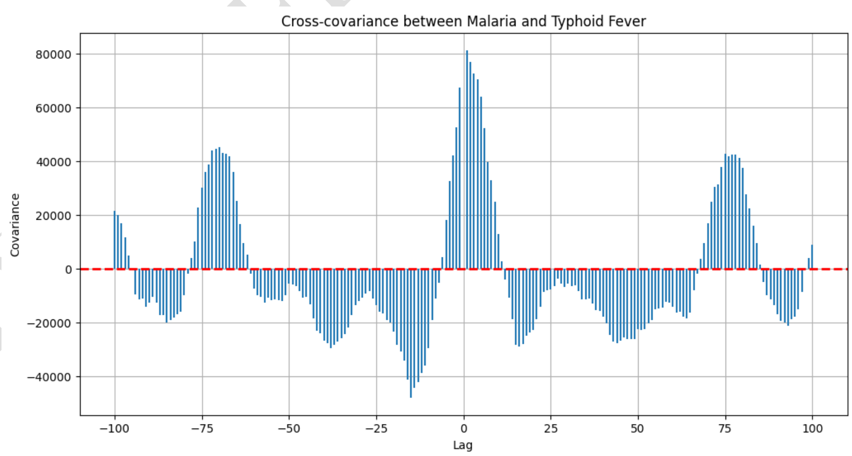

***Figure 7**: Cross-covariance*

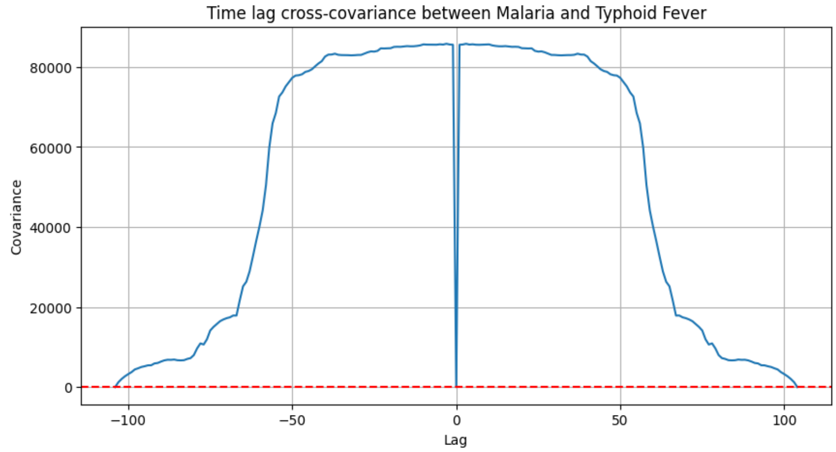

Figure 8: *Time lag cross-covariance*

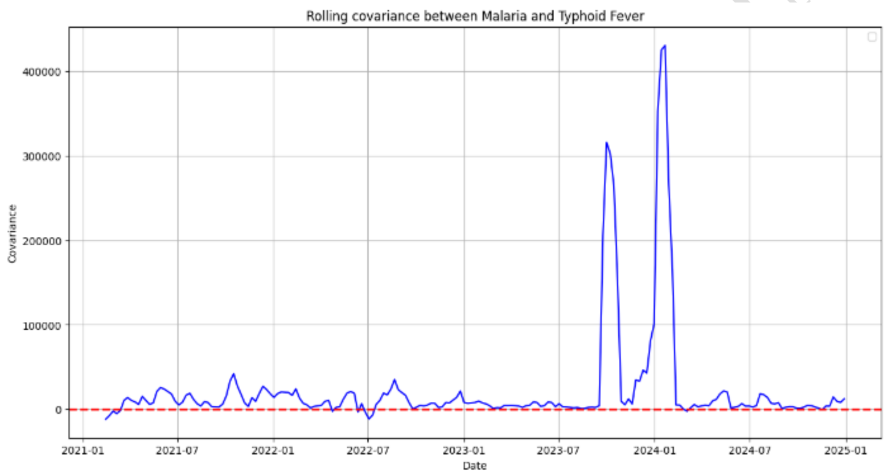

Figure 9: *Rolling covariance*

Figure 7 illustrates how the covariance between Malaria and Typhoid Fever changes over various lags. Lags range from -100 to 100, indicating the time lags at which the cross-covariance is calculated. Negative lags represent past values of Typhoid Fever affecting current values of Malaria, while positive lags indicate the opposite. The cross-covariance values are either positive, associated with peaks, or negative, characterised by troughs. At lag zero, the series strongly vary together. Figure 8 is related to the time lag cross-covariance plot ranging from -100 to 100. Negative lags show the effect of past values of Typhoid Fever on current values of Malaria, while positive lags show the effect of past Malaria values on Typhoid Fever. The plot reveals a relatively flat region with high positive covariance values (from about -50 to +50), unveiling that fluctuations in one disease are consistently associated with fluctuations in the other over this range. The cross-covariance remains well above zero for most lags. Finally, Figure 9 presents the rolling covariance plot between Malaria and Typhoid Fever over time, covering the studied period, with a window size of 6. Overall, the two diseases tend to occur together. Peaks in covariance are explicitly observed in mid-2023 and the beginning of 2024.

4. Discussion

This work first involved a stationarity test. As the malaria series was identified as non-stationary, it has been differentiated. The plot of the curves showed similar trends over several periods, confirming interdependence. The scatter diagram indicates that points trend upwards from left to right, mainly around the regression line, leading to the conclusion of a positive relationship between the series. The coefficients of correlation confirmed this notorious relationship, as they are well above the positive mean (0.5). The cross-correlation shows a highest peak at lag zero between the two series, revealing an immediate relation. For the joint lag, the cross-correlations remain between 0.8 and 1. The sliding correlation analysis for a window of size 6 reveals a correlation almost always above the positive mean (0.5). Thus, incorporating time dynamics in the analysis confirms a significant relationship. The two series vary similarly together, with concomitant peaks. Following the combined lag, this peak remains constant and high between -50 and 50. Finally, the rolling covariance stays above zero most of the time, with many infections observed in mid-2023 and early 2024. Overall, this analysis, based primarily on correlation and covariance, reveals a substantial relationship between Malaria and Typhoid Fever with a notable contemporaneous correlation. The relation is strong and stable across the examined lags.

The work presented in this paper used statistical approaches to understand some common epidemiological phenomena. When compared to others, the work of [8] is based in the same geographical area as the present study but focuses solely on one disease for the statistical analysis. [4] on his side, carried out a study in Ngaoundere, the Adamawa Region Capital, focusing only on prevalence assessment. Papers [4], [10] also tackle Malaria and Typhoid Fever coinfections. Most of the work combined correlation assessment with another method: regression analysis in [14], wavelet analysis in [15], random forest in [16] and detrended fluctuation analysis in [19]. All the researchers limited their study to cross-correlation, leaving out time lag and rolling analysis. None of them focused on both correlation and covariance approaches.

The work carried out in this study is distinctive because it considers a wide range of statistical tools to assess the correlation and covariance between two diseases, unlike other studies, which use only one or two tools. In addition to correlation, covariance is used to understand the joint variation of both diseases better. In this geographical area, no studies have focused on statistically explaining the correlation and covariance of these two diseases.

The limitation of this work mainly lies in the availability of data. Only weekly cases from the last 5 years were available. Furthermore, obtaining data on gender, age, climate, environment, and socio-economic considerations should provide more insightful information on causalities and facilitate a multivariate analysis. Clinical cases may also be considered.

Awareness of this valuable statistical information makes it impactful and worthwhile to:

- Help understand the dynamics of the diseases and inform interventions.
- Monitor the trend of one disease and provide insights on the trend of the other, valuable for resource allocations.
- Monitor diseases in tandem to help predict trends and inform outbreak management strategies.

5. Conclusion

The main objective of this work was to assess the degree and direction of malaria and typhoid fever, two diseases classified as potentially epidemiological and for which coinfection cases are often reported in Health facilities. To that end, correlation and covariance approaches were applied to time series data of the Adamawa Region of Cameroon, spanning from January 2021 to December 2024. The results revealed a strong and constant relationship between these two diseases over time, which may help in the joint implementation of surveillance and response policies. The outlook includes obtaining more data for casualty analysis and multivariate analysis.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:**We acknowledge the Health Information Unit of the Ministry of Public Health of Cameroon, and the Adamawa Regional Delegation of the same Ministry for providing us with data and relevant information for this work.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author.

References

[1] J. Moskolaï Ngossaha, A. Ynsufu, A. BatoureBamana, R. Djeumen, S. Bowong Tsakou, and A. Ayissi Eteme, 'Towards a Flexible Urbanisation Based Approach for Integration and Interoperability in Heterogeneous Health Information Systems: Case of Cameroon', in *Safe, Secure, Ethical, Responsible Technologies and Emerging*

Applications, vol. 566, F. Tchakounte, M. Atemkeng, and R. P. Rajagopalan, Eds., in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 566. , Cham: Springer Nature Switzerland, 2024, pp. 258–275. doi: 10.1007/978-3-031-56396-6_16.

[2] K. K. S. Garcia *et al.*, ‘Malaria time series in the extra-Amazon region of Brazil: epidemiological scenario and a two-year prediction model’, *Malar. J.*, vol. 21, no. 1, p. 157, Dec. 2022, doi: 10.1186/s12936-022-04162-1.

[3] C. L. Kim *et al.*, ‘The Burden of Typhoid Fever in Sub-Saharan Africa: A Perspective’, *Res. Rep. Trop. Med.*, vol. Volume 13, pp. 1–9, Mar. 2022, doi: 10.2147/RRTM.S282461.

[4] F. S. Sohanang Nodem, D. Ymele, M. Fadimatou, and S.-P. C. Fodouop, ‘Malaria and Typhoid Fever Coinfection among Febrile Patients in Ngaoundere (Adamawa, Cameroon): A Cross-Sectional Study', *J. Parasitol. Res.*, vol. 2023, p. 5334813, 2023, doi: 10.1155/2023/5334813.

[5] A. BatoureBamana, Y. Sokdou Bila Lamou, and A. Abdoulaye, ‘Benchmark Analysis of Time Series Models for Malaria Trends in the Adamawa Region (Cameroon)’, in *Research in Computer Science*, vol. 2462, D. Aissani, K. Barkaoui, and M. Roche, Eds., in Communications in Computer and Information Science, vol. 2462. , Cham: Springer Nature Switzerland, 2025, pp. 61–79. doi: 10.1007/978-3-031-88226-5_5.

[6] A. BatoureBamana, M. Shafiee Kamalabad, and D. L. Oberski, ‘A systematic literature review of time series methods applied to epidemic prediction’, *Inform. Med. Unlocked*, vol. 50, p. 101571, 2024, doi: 10.1016/j.imu.2024.101571.

[7] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, Third print edition. Melbourne, Australia: Otexts, Online Open-Access Textbooks, 2021.

[8] A. BatoureBamana, E. Dangbe, H. Abboubakar, and M. Shafiee Kamalabad, ‘A comprehensive statistical analysis of Malaria dynamics in the Adamawa region of Cameroon, from 2018 to 2022’, *Braz. J. Biom.*, vol. 42, no. 3, pp. 289–306, Aug. 2024, doi: 10.28951/bjb.v42i3.703.

[9] J. Jeong and W. Chang, ‘Analysis of East Asia Wind Vectors Using Space–Time Cross-Covariance Models’, *Remote Sens.*, vol. 15, no. 11, p. 2860, May 2023, doi: 10.3390/rs15112860.

[10] T. A. Olowolafe, O. F. Agosile, A. O. Akinpelu, N. Aderinto, O. Z. Wada, and D. B. Olawade, 'Malaria and typhoid fever coinfection: a retrospective analysis of University Hospital records in Nigeria', *Malar. J.*, vol. 23, no. 1, p. 220, Jul. 2024, doi: 10.1186/s12936-024-05052-4.

[11] S. Chatterjee, ‘A New Coefficient of Correlation’, *J. Am. Stat. Assoc.*, vol. 116, no. 536, pp. 2009–2022, Oct. 2021, doi: 10.1080/01621459.2020.1758115.

[12] H. Kato and A. Takizawa, ‘Time series cross-correlation between home range and number of infected people during the COVID-19 pandemic in a suburban city’, *PLOS ONE*, vol. 17, no. 9, p. e0267335, Sep. 2022, doi: 10.1371/journal.pone.0267335.

[13] V. U. Monnaka and C. A. C. D. Oliveira, ‘Google Trends correlation and sensitivity for outbreaks of dengue and yellow fever in the state of São Paulo’, *Einstein São Paulo*, vol. 19, p. eAO5969, Jul. 2021, doi: 10.31744/einstein_journal/2021AO5969.

[14] M. Sarmadi, S. Rahimi, D. Evensen, and V. Kazemi Moghaddam, ‘Interaction between meteorological parameters and COVID-19: an ecological study on 406 authorities of the UK’, *Environ. Sci. Pollut. Res.*, vol. 28, no. 47, pp. 67082–67097, Dec. 2021, doi: 10.1007/s11356-021-15279-2.

[15] B. A. Rivieccio *et al.*, ‘CoViD-19, learning from the past: A wavelet and cross-correlation analysis of the epidemic dynamics looking to emergency calls and Twitter trends in Italian Lombardy region’, *PLOS ONE*, vol. 16, no. 2, p. e0247854, Feb. 2021, doi: 10.1371/journal.pone.0247854.

[16] Y. Chen *et al.*, ‘Epidemiology and prediction of multidrug-resistant bacteria based on hospital level’, *J. Glob. Antimicrob. Resist.*, vol. 29, pp. 155–162, Jun. 2022, doi: 10.1016/j.jgar.2022.03.003.

[17] M. Anwer and F. Jahan, ‘Interdependency and cross-dependencies of COVID-19 time-series parameters using autocorrelation and cross-correlation’, in *2023 8th International Conference on Smart and Sustainable Technologies (SpliTech)*, Split/Bol, Croatia: IEEE, Jun. 2023, pp. 1–6. doi: 10.23919/SpliTech58164.2023.10193321.

[18] D. C. Lévano, S. E. López, A. Gherasim, J. O. Domínguez, M. T. D. Rico, and M. G. Gómez, ‘The Temporary Incapacity (TI) register as a complementary system to traditional epidemiological surveillance during the COVID-19 pandemic in Spain’, *PLOS ONE*, vol. 19, no. 5, p. e0301344, May 2024, doi: 10.1371/journal.pone.0301344.

[19] S. J. A. D. Jesus, E. M. D. Araújo, and A. M. D. Silva Filho, 'Autocorrelation and Cross-Correlation of Hospitalisations for Syphilis and HIV/AIDS in the State of Bahia’, *Rev. Gest. Soc. E Ambient.*, vol. 18, no. 11, p. e09853, Nov. 2024, doi: 10.24857/rgsa.v18n11-147.

[20] B. Leouro Mbaiossoum, A. Ibrahim Doutoum, A. BatoureBamana, N. Batouma, L. Dionlar, and D. Noudjibarem, ‘Self-Efficacy Prediction Model Using Bayesian Networks’, *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 37s, pp. 530–536, Apr. 2025, doi: 10.52783/jisem.v10i37s.6447.

[21] A. B. Bamana, Y. S. B. Lamou, D. J. Fotsa-Mbogne, and M. S. Kamalabad, ‘Addressing Non-stationarity with Stochastic Trend in the Context of Limited Time Series Data: An Experimental Survey’, Mar. 25, 2025. doi: 10.21203/rs.3.rs-6289779/v1.

[22] R. Moraffah*et al.*, ‘Causal inference for time series analysis: problems, methods and evaluation’, *Knowl. Inf. Syst.*, vol. 63, no. 12, pp. 3041–3085, Dec. 2021, doi: 10.1007/s10115-021-01621-0.

[23] A. Aieb, A. Liotta, A. Jacob, and M. A. Yaqub, ‘Short-Term Forecasting of Non-Stationary Time Series’, in *ITISE 2024*, MDPI, Jul. 2024, p. 34. doi: 10.3390/engproc2024068034.

[24] D. Batabyal, D. Bandopadhyay, B. Sadhukhan, N. Das, and S. Mukherjee, ‘Exploring Stationarity and Fractality in Stock Market Time-series’, in *2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, Silchar, India: IEEE, Feb. 2023, pp. 1–6. doi: 10.1109/ISACC56298.2023.10084056.

[25] R. Ahsan, R. Neamtu, M. Bashir, E. A. Rundensteiner, and G. Sarkozy, ‘Correlation-Based Analytics of Time Series Data’, in *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA: IEEE, Dec. 2020, pp. 4482–4491. doi: 10.1109/BigData50022.2020.9378155.

[26] A. Dokumentov and R. J. Hyndman, ‘STR: Seasonal-Trend Decomposition Using Regression’, *Inf. J. Data Sci.*, vol. 1, no. 1, pp. 50–62, Apr. 2022, doi: 10.1287/ijds.2021.0004.

[27] M. Joseph, S. Moonsammy, H. Davis, D. Warner, A. Adams, and T. D. Timothy Oyedotun, ‘Modelling climate variabilities and global rice production: A panel regression and time series analysis’, *Heliyon*, vol. 9, no. 4, p. e15480, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15480.

[28] R. Espinosa, F. Jiménez, and J. Palma, ‘Multi-surrogate assisted multi-objective evolutionary algorithms for feature selection in regression and classification problems with time series data’, *Inf. Sci.*, vol. 622, pp. 1064–1091, Apr. 2023, doi: 10.1016/j.ins.2022.12.004.

[29] A. E. Yuan and W. Shou, ‘A rigorous and versatile statistical test for correlations between time series’, Jan. 28, 2022. doi: 10.1101/2022.01.25.477698.

[30] G. Li, A. Zhang, Q. Zhang, D. Wu, and C. Zhan, ‘Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction’, *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 69, no. 5, pp. 2413–2417, May 2022, doi: 10.1109/TCSII.2022.3160266.

[31] K. Ali Abd AlHameed, ‘Spearmans correlation coefficient in statistical analysis’, *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, Jan. 2022, doi: 10.22075/ijnaa.2022.6079.

[32] H. Rahadian, S. Bandong, A. Widyotriatmo, and E. Joelianto, ‘Image encoding selection based on Pearson correlation coefficient for time series anomaly detection’, *Alex. Eng. J.*, vol. 82, pp. 304–322, Nov. 2023, doi: 10.1016/j.aej.2023.09.070.

[33] A. Iorio *et al.*, ‘Cross-correlation of virome–bacteriome–host–metabolome to study respiratory health’, *Trends Microbiol.*, vol. 30, no. 1, pp. 34–46, Jan. 2022, doi: 10.1016/j.tim.2021.04.011.

[34] S. P. Shrestha, W. Chaisowwong, M. Upadhyaya, S. P. Shrestha, and V. Punyapornwithaya, ‘Cross-correlation and time series analysis of rabies in different animal species in Nepal from 2005 to 2018’, *Heliyon*, vol. 10, no. 3, p. e25773, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25773.

[35] G. F. Zebende, A. A. Brito, and A. P. Castro, ‘DCCA cross-correlation analysis in time-series with removed parts’, *Phys. Stat. Mech. Its Appl.*, vol. 545, p. 123472, May 2020, doi: 10.1016/j.physa.2019.123472.

[36] J. M. Polanco-Martínez and J. L. López-Martínez, ‘NonParRolCor: An R package for estimating rolling correlation for two regular time series’, *SoftwareX*, vol. 22, p. 101353, May 2023, doi: 10.1016/j.softx.2023.101353.

[37] T. Ferdousi, L. W. Cohnstaedt, and C. M. Scoglio, ‘A Windowed Correlation-Based Feature Selection Method to Improve Time Series Prediction of Dengue Fever Cases’, *IEEE Access*, vol. 9, pp. 141210–141222, 2021, doi: 10.1109/ACCESS.2021.3120309.

[38] Z. Sun and T. J. Fisher, ‘Testing for correlation between two time series using a parametric bootstrap’, *J. Appl. Stat.*, vol. 48, no. 11, pp. 2042–2063, Aug. 2021, doi: 10.1080/02664763.2020.1783519.

[39] M. Stehlík *et al.*, ‘On asymmetric relations and robustified cross-correlation approach to surveillance based on detection of SARS-CoV-2 in wastewater in Chile and Peru’, *Chemom. Intell. Lab. Syst.*, vol. 242, p. 104987, Nov. 2023, doi: 10.1016/j.chemolab.2023.104987.

[40] A. Uryga *et al.*, ‘Exploration of simultaneous transients between cerebral hemodynamics and the autonomic nervous system using windowed time-lagged cross-correlation matrices: a CENTER-TBI study’, *Acta Neurochir. (Wien)*, vol. 166, no. 1, p. 504, Dec. 2024, doi: 10.1007/s00701-024-06375-6.

[41] R. Yu *et al.*, ‘Time series cross-correlation network for wind power prediction’, *Appl. Intell.*, vol. 53, no. 10, pp. 11403–11419, May 2023, doi: 10.1007/s10489-022-04004-2.
[42] I. Persson and J. Khojasteh, ‘Python Packages for Exploratory Factor Analysis’, *Struct. Equ. Model. Multidiscip. J.*, vol. 28, no. 6, pp. 983–988, Nov. 2021, doi: 10.1080/10705511.2021.1910037.