Vol. 6 (Iss. 2) 2025, pp. 283-296 African Journal of Empirical Research https://ajernet.net ISSN 2709-2607

On Bayesian Generalized Linear Mixed Modeling of Cholera Risk Factors: A Case Study of Masvingo Province, Zimbabwe

Ever Moyo[1*]
Sipho Sibanda[2]
Alice Chimhondoro[3]
Senzenia Chakauya[4]
Davidon Jani[5]

[1]Mathematics and Statistics Lecturer, Department of Mathematical Sciences, Zimbabwe Open University, Masvingo Region, Zimbabwe, moyoe@zou.ac.zw
https://orcid.org/0009-0006-9375-2018
[2]Mathematics and Statistics Lecturer, Department of Agriculture, Sciences and Health Professions, Solusi University, Bulawayo, Zimbabwe, sibandas@solusi.ac.zw
https://orcid.org/0009-0002-7425-8672
[3]Mathematics and Statistics Lecturer, Department of Mathematical Sciences, Zimbabwe Open University, Harare Region, Zimbabwe, chimhondoroa@zou.ac.zw
https://orcid.org/0009-0007-3981-4860
[4]Mathematics and Statistics Lecturer, Department of Mathematical Sciences, Zimbabwe Open University, Mashonaland West Region, Zimbabwe
https://orcid.org/0009-0001-3632-4619
[5]Mathematics and Statistics Lecturer, Department of Mathematical Sciences, Zimbabwe Open University, National Centre, Zimbabwe, janid@zou.ac.zw
https://orcid.org/0009-0000-3149-1584

*Corresponding Author: Ever Moyo, moyoe@zou.ac.zw

https://doi.org/10.51867/ajernet.maths.6.2.25

ABSTRACT

Cholera remains a significant public health threat in Masvingo, Zimbabwe, particularly in districts with inadequate water, sanitation and healthcare infrastructure. This study applies a Bayesian Generalised Linear Mixed Model (BGLMM) to analyse key risk factors associated with cholera incidence across seven districts (Masvingo, Chivi, Zaka, Bikita, Gutu, Mwenezi and Chiredzi) in Masvingo Province, Zimbabwe, during the 2023/ 2024 outbreak. Descriptive statistics show that cholera affects a young population with an average age of 24.6 years. The average distance to a health facility is 5.95km, indicating potential challenges in accessing healthcare. The frequency distributions reveal that 60.9% of the sampled population reported cholera cases. Significant fixed-effect predictors include gender, access to health facilities and seasonal risk, with the wet season posing a substantially higher risk (Estimate = 6.143, 95% CI: 3.649–9.114). Random effects suggest district-level variations, with Masvingo showing the highest risk deviation, followed by Chiredzi and Mwenezi. These findings were supported by a geospatial cholera case density heatmap, visualising concentrated high-risk zones (dark red) in the three districts. Chivi, Zaka, Bikita and Gutu appear as lighter spots, suggesting lower cholera intensity. Model fit statistics, Widely Applicable Information Criterion (WAIC) (110.115) and Leave-One-Out Cross Validation (LOO) (112.700), validate the model`s adequacy for evaluating cholera determinants. The study highlights the urgent need for integrated public health strategies, including infrastructure development, community health education and targeted interventions to curb cholera incidence. Such efforts will enhance resilience and reduce cholera incidence among vulnerable populations in Masvingo Province.

Keywords: Cholera Risk Factors, Geospatial Analysis, Public Health Intervention, Seasonal Risk, Vulnerable Populations

I. INTRODUCTION

Cholera is highly infectious and threatens life. It is a diarrheal disease caused by the bacterium Vibrio cholerae. It can kill within hours if left untreated. The infected individuals will become severely ill with symptoms like watery diarrhoea, vomiting, rapid dehydration and abdominal cramps. Quick access to treatment is crucial. Researchers estimate that there are 1.3 to 4.0 million cases and 21000 to 143 000 deaths from cholera worldwide yearly. The population's access to safe water, basic sanitation and hygiene (WASH) is essential to prevent cholera. (WHO, 2024). The first pandemic or global epidemic was recorded in the 19th century. Since then, six pandemics have killed millions of people worldwide. The current (seventh) pandemic started in south Asia in 1961 and continues to affect populations globally.

Cholera outbreaks occur regularly in some countries. In others, they are less frequent, and it may be years between outbreaks. Cholera is linked to limited access to safe water, basic sanitation facilities and poor hygiene practices. This may be due to conflict, population displacement, climate events like cyclones, floods or drought, and lack of investment in maintaining and improving WASH services and infrastructure. The number of cholera cases reported to WHO has continued to rise in recent years. In 2023, 535 321 cases and 4007 deaths were reported to the WHO from 45 countries. Countries at risk or affected by cholera should strengthen their surveillance systems according to the revised Global Taskforce on Cholera Control (GTFCC) recommendations to quickly detect and respond to outbreaks. Community engagement involves collaborating with people and communities to develop and implement programmes designed to address their needs. Local culture, practices and beliefs are central to promoting protective practices, such as handwashing with soap and water, safe preparation and storage of food and water, and safe disposal of faeces. Funeral practices for individuals who die from cholera may need to be adapted to prevent infection among attendees. Community engagement is essential for effectively communicating the potential risks and symptoms of cholera, precautions to take to avoid cholera, when and where to report cases, and the importance of seeking immediate treatment if symptoms appear. (WHO, 2023).

Cholera is a public health concern in Zimbabwe, with Masvingo Province experiencing recurrent outbreaks in recent years, with several people hospitalised and others dying (WHO, 2024; Makoni et al., 2020), creating an environment conducive to cholera transmission (WHO, 2019). Previous studies have identified various risk factors like poor quality of water and lack of proximity to safe water sources, inadequate sanitation and hygiene practices, food handling and consumption habits, demographic factors such as age, sex and population density and environmental factors like temperature, rainfall and altitude (Mweemba et al., 2018; Ngwerume et al., 2019). Chikoto and Hughes (2019), Mugabe et al. (2021). The Bayesian Generalized Linear Mixed Models offer a robust framework for investigating cholera risk factors in Masvingo Province, Zimbabwe. The models` ability to handle hierarchical data, incorporate prior knowledge and quantify uncertainty makes them a powerful tool in epidemiological research, ultimately supporting more effective cholera control and prevention efforts (Lawson, 2021). By addressing the knowledge gap and limitations of previous studies, this research will provide valuable insights for policymakers, public health practitioners, and researchers who want to develop effective strategies for cholera control in Masvingo Province and similar settings. The analysis will use historical data from 2023 to 2024 on cholera cases, environmental conditions, socioeconomic factors, and demographic variables to develop a Bayesian Generalized Linear Mixed Model (BGLMM).

1.1 Problem Statement

Cholera remains a significant public health concern in Masvingo Province, Zimbabwe, with recurrent outbreaks causing morbidity and mortality. Despite existing research, the complex interplay of environmental, socio-economic, demographic and health-related factors contributing to cholera risk in Masvingo Province remains poorly understood. Traditional statistical models have limitations in accounting for spatial and temporal correlations, non-linear relationships and uncertainty. Hence, a robust analytical approach is needed using Bayesian Generalised Linear Mixed Model to identify significant risk factors, explore spatial patterns and inform evidence-based interventions to mitigate cholera incidence in Masvingo.

1.2 Purpose of the study

This study aims to apply the Bayesian Generalized Linear Mixed Model to analyse cholera risk factors in Masvingo Province, Zimbabwe, while accounting for spatial and temporal variations and uncertainties in the data.

1.3 Objectives of the study

This study aims:

(i) To identify significant risk factors associated with cholera incidence in Masvingo, Zimbabwe, using the Bayesian Generalized Linear Mixed Model.
(ii) To examine spatial variation in cholera incidence in Masvingo and its association with geographic characteristics.
(iii) To evaluate the effectiveness of the Bayesian Generalised Linear Mixed Model in modelling the relationship between risk factors and cholera incidence in Masvingo.
(iv) To provide evidence-based suggestions for cholera prevention and control in Masvingo.

1.4 Significance of the study

This study provides critical insights that contribute to reducing the cholera burden, improving public health policy and advancing epidemiological research in Masvingo Province, Zimbabwe, and beyond.

II. LITERATURE REVIEW

Cholera remains a significant public health concern, particularly in resource-poor settings. Understanding risk factors is crucial for effective prevention and control. Bayesian Generalised Linear Mixed Models offer a robust approach to analysing complex health data. This study aims to investigate the cholera risk factors using Bayesian Generalised Mixed Models in Masvingo, Zimbabwe.

2.1 Cholera Risk Factors in Masvingo Province

Masvingo, Zimbabwe, has experienced several cholera outbreaks, which have significantly impacted public health. The risk factors associated with these outbreaks include inadequate access to clean water, poor sanitation facilities, and limited healthcare services, which are escalated by socio-economic conditions such as poverty and population density (Mugabe et al., 2021). Additionally, environmental factors such as rainfall patterns and proximity to contaminated water sources play a crucial role in cholera transmission dynamics (Chikoto & Hughes, 2019). Again, Macheka and Chikoto (2021) assessed the efficiency of service delivery provisions to Victoria Ranch suburb residents in Masvingo Urban

2.2 Bayesian Generalised Linear Mixed Model Overview

Bayesian Generalised Linear Mixed Models (BGLMMS) extend the traditional Generalised Linear Models (GLMS) by incorporating random effects to account for hierarchical or clustered data structures. This flexibility is crucial in epidemiological studies where data often come from different levels (e.g., individual, household, community) and are subject to multiple sources of variability (Ugarte et al., 2022). Bayesian methods are particularly effective for dealing with incomplete data, which is a common challenge in epidemiological studies. Again, the Bayesian approach provides credible intervals for parameter estimates, offering a probabilistic interpretation of the results (Held & Rue, 2022). Markov Chain Monte Carlo (MCMC) Methods algorithms, such as Gibbs sampling and Metropolis-Hastings, are widely used for estimating the posterior distribution of model parameters in complex models (Fong et al., 2021). These methods are computationally intensive but provide accurate estimates for high-dimensional models. Integrated Nested Laplace Approximation (INLA) is a computationally efficient alternative to MCMC for Bayesian inference in latent Gaussian models, including BGLMMS. It is particularly useful for spatial and spatio-temporal models, such as those used to investigate Masvingo's cholera risk (Ugarte et al., 2022). Bayesian Generalised Linear Mixed Models have been used in various studies to investigate risk factors for infectious diseases: In Bangladesh and Haiti, Bayesian Generalised Linear Mixed Models have been used to model the spatial and temporal distribution of cholera, identifying significant environmental and socio-economic determinants, such as proximity to water bodies and poverty levels (Li et al., 2020). Again, Fong et al., 2021 applied Bayesian Generalised Linear Mixed Models to model the risk factors for diseases such as malaria and dengue fever, successfully accounting for spatial heterogeneity and dependencies.

2.3 Investigating cholera risk factors using other methods

Cowman et al. (2017) investigated the relationship between cholera occurrence and various environmental and demographic factors related to water, sanitation, socio-economic status, education, urbanization and availability of health facilities in Kenya from 2008 to 2013. They analysed the data using a zero-inflated negative binomial regression model and a multivariate analysis. They observed that the risk of cholera was associated with open defecation, use of unimproved water sources, poverty headcount ratio and the number of health facilities per 100 000 population. Dureab et al. (2019) conducted a case-control study in Aden, Yemen, to investigate the risk factors for cholera in the 2018 cholera outbreak using a bivariate and multivariate analysis. They discovered that risk factors associated with a cholera case included: a history of travelling and having visitors from outside Aden Province; eating outside the house; not washing fruit, vegetables, and khat (a local herbal stimulant) before consumption; using common-source water and not using chlorine or soap in the household. Elimian et al. (2020) used multivariable logistic regression to identify and quantify the factors associated with cholera-related deaths in Nigeria. They used a cross-sectional design to analyse the surveillance data from all the States that reported cholera cases during the 2018 outbreak. The adjusted multivariable model revealed that, older age, male gender, living in peri-urban areas or flooded states, infection during the rainy season, and delay in seeking health care were positively associated with cholera-related death, whereas those living in urban areas, hospitalisation in the course of illness, and presentation to a secondary hospital were negatively associated with cholera-related death Ngereza et al. (2021) examined the cholera risk factors to determine the best model, in Tanzania, between the Poisson regression (PR) and the Geographically Weighted Poisson regression (GWPR) models. The results of the GWPR model showed that accessing improved water sources, practicing hand-washing, and open defecation have an inverse relationship with the number of cholera cases. Leo et al. (2019) explored the use of machine learning techniques to model cholera epidemics with linkage to seasonal weather changes while overcoming the data imbalance problem. Adaptive Synthetic Sampling Approach (ADASYN) and Principal Component Analysis (PCA) were used to restore sampling balance and dimensionality of the dataset. Overall results improved the understanding of the significant roles of machine learning strategies in healthcare data.

Ayling et al. (2023) used data on individual cases, where the household location and time of each case were known, to explore factors associated with cholera transmission and to retrospectively generate detailed maps of weekly transmission risk across the epidemic in Harare, Zimbabwe, between September 2018 and January 2019. Their study results showed that several potential risk factors and people living near the sewer network with high access to piped water were at higher risk because the sewer would burst, leading to the contamination of the piped water network. Cholera is a concern in Zimbabwe. As of 11 April 2024, a total of 31,705 cholera cases and 683 deaths with a cumulative case fatality rate (CFR) of 2.2 per cent have been reported from 63 districts across the 10 provinces. The cumulative cholera cases, approximately 31 per cent are children aged below 15 years, and 14 per cent are children under five years. (UNICEF, 2024)

III. RESEARCH METHODOLOGY

3.1 Area of study

Masvingo Province is in southeastern Zimbabwe. It has an area of 56 566 km^2 and a population of 1.638 million as of the 2022 census, ranking fifth out of Zimbabwe's ten provinces. The province is divided into seven districts: Bikita, Chivi, Zaka, and Masvingo in the centre of the province, Gutu in the north, and Mwenezi and Chiredzi in the south and east, respectively. The province experienced cholera outbreaks in 2008/2009, 2018/2019 (Munyenyiwa et al., 2022) and 2023/2024 (WHO, 2023, 2024). Figure 1 below is a map showing Masvingo Province.

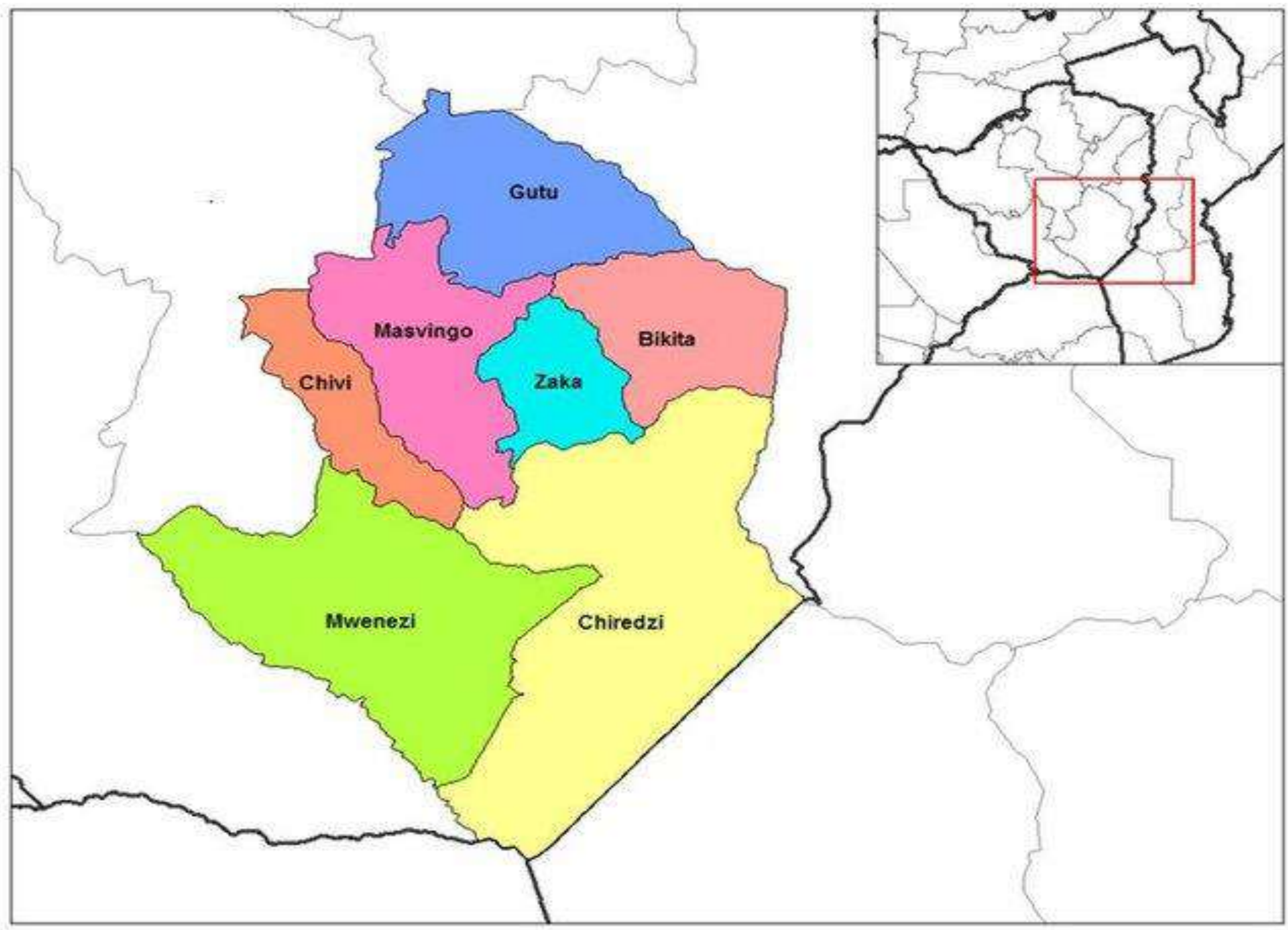

Figure 1
Masvingo Province Map

3.2 Data Collection

The study utilized secondary cholera case data from the Zimbabwe Ministry of Health and Child Care (MoHCC) and World Health Organisation (WHO) databases from 2023/ 2024 cholera outbreak in Masvingo Province. The data consisted of 350 participants across 7 districts in Masvingo Province with 50 participants from each district. These data typically included information on the number of confirmed cholera cases data, patient demographics (age, sex), and socio-economic data (access to cholera education, access to sanitation, access to health facility and geographic location (district). Environmental data, such as access to clean water, season risk period, average distance to health facility, overcrowding and cholera education, were collected from Meteorological Department (Zimbabwe Meteorological Services Department, 2023), UNICEF, WHO and Ministry of Health and Child Care (MoHCC, 2023) situation reports on internet. Data on socio-economic factors (education and sanitation facilities) and demographic and geographical data were sourced from national census records (ZIMSTAT, 2022).

3.3 Data Preprocessing

Data analysis was done using JASP (version 0.19.3.0) and Python software. It involves several steps, from data preparation to model interpretation. The dataset contains cholera cases and relevant predictors. The response variable was binary (1 = Cholera case, 0 = No case). Predictor variables were categorical (access to clean water, health facility and sanitation, overcrowding, flood risk, cholera education, season risk period) and continuous (average distance to health facility and age). Before analysis, data was cleaned by handling missing values, detecting outliers and standardising variables. The descriptive statistics were computed for the continuous variables and frequency tables were produced for all categorical data, to understand their distributions and identify potential outliers and for data quality issues. A binomial distribution in a logit link function for the binary outcomes was used. The data was grouped into fixed effects (access to sanitation, clean water and health facilities, overcrowding, flood risk, cholera education, average distance to health facility and season risk period) and random effects (districts: Masvingo, Chivi, Zaka, Bikita, Gutu, Mwenezi and Chiredzi) to model using Bayesian Generalised Linear Mixed Model (BGLMM).A geospatial analysis was carried out and produced a cholera case density heatmap using Python to visualise and analyse the geographic distribution of cholera cases.

3.4 Model Formulation and Fitting

A Bayesian Generalized Linear Mixed Model (BGLMM) was developed to investigate the relationship between cholera risk and various covariates. In the Bayesian framework, model parameters were estimated using Markov Chain Monte Carlo (MCMC) methods (Fong et al., 2021), requiring the specification of prior distributions. After fitting the model, convergence diagnostics, the Gelman-Rubin statistic (R-hat) and Estimated Sample Size (ESS) ensured that the MCMC chains have stabilized. Posterior predictive checks compared simulated and observed data to assess model fit, while goodness-of-fit criteria was done using the Widely Applicable Information Criterion (WAIC) and Leave-One-Out Cross Validation (LOO).

3.4.1 Binomial Model Specification

In this study, a Bayesian Generalized Linear Mixed Model (BGLMM) with a binomial response (number of cholera cases out of a total population at risk), was formulated as follows:

$$Y_i \sim \text{Binomial}(N_i, p_i)$$

$$\text{logit}(p_i) = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + u_i$$

$$u_i \sim N(0, \sigma_u^2)$$

Where:

Y_i= number of cholera cases in a district i.
N_i = total population at risk in district i.
P_i = probability of a cholera case occurring in district i.
β_0 = intercept.
X_{ij}= predictor variables (rainfall, temperature, sanitation, water access).
β_j = fixed effect coefficients.
$u_i \sim N(0, \sigma_u^2)$ = random effect for clustering at district level.

3.4.2 Bayesian Priors

In Bayesian modelling, priors are assigned to parameters:

$\beta_j \sim N(0,10)$ $\sigma^2 \sim \text{Half-Cauchy}(0,2.5)$

This model estimates the probability of cholera infection given environmental and socio-economic risk factors.

Random effects account for unobserved variability across districts.
Results provide posterior distributions for model parameters, allowing uncertainty quantification.

3.4.3 Model Formulation

The Bayesian Generalised Linear Mixed Model can be specified as:

$$\text{Log}\ (\lambda_{ij}t) = \beta_0 + X_{ij}t\beta + u_i t + vt + w_i t$$

Where:

$\lambda_{ij}t$ is the expected number of cholera cases in district 'i' at time" t".
β_0 is the intercept term.
$X_{ij}t$ represents the matrix of fixed effects (covariates) for district i at time t.
β is the vector of coefficients for the fixed effects.
u_i is the spatial random effect for district i, capturing the unobserved spatial heterogeneity.
vt is the temporal random effect for time t, capturing the unobserved temporal heterogeneity.
$w_i t$ is the spatiotemporal interaction term, capturing specific interactions between district i and time t.

3.5 Spatial Analysis and Heatmaps

Spatial analysis involves examining the locations, attributes, and relationships of features in spatial data through various analytical techniques. It is used to identify spatial patterns and relationships, assess geographic distribution and clustering of events. Spatial analysis was done using a geospatial heatmap. It represented cholera case density across the seven districts using colour gradients. The geospatial cholera case density heatmap was made using latitude and longitude coordinates and the number of cholera cases in the seven districts. The districts with more cholera cases appeared darker, indicating higher intensity, while regions with fewer cases would appear lighter, providing clear and actionable insights.

IV. DATA ANALYSIS & DISCUSSION

4.1 Descriptive Statistics

Table 1
Descriptive Statistics for Age and Average Distance to a Health Facility

	District	Age(year)	Average distance to health facility (km)
Valid	350	350	350
Missing	0	0	0
Mean		24.60	5.95
Standard Deviation		16.08	2.13
Minimum		1	1
Maximum		64	11

Table 1 shows that the dataset consisted of 350 individuals with no missing values across all variables. The mean age of the population is 24.6 years, with a wide range (1 to 64 years**)** and a high standard deviation (16.08), indicating a mix of younger and older individuals, since cholera can affect people differently based on age. Further analysis is required to determine if certain age groups are more vulnerable. The descriptive statistics for the average distance to a health facility indicate that, on average, individuals in the study area are approximately 5.95 km away from the nearest healthcare centre. The minimum distance recorded is 1 km, indicating that some individuals have convenient access to health services, while the maximum distance is 11 km, highlighting that others may face significant challenges in reaching medical care. This disparity suggests potential inequalities in healthcare accessibility, which could be particularly concerning during cholera outbreaks when timely medical intervention is critical. Those living farther from health facilities may experience delays in seeking treatment, increasing the risk of severe health outcomes. These findings underscore the need for targeted interventions, such as mobile clinics, improved transportation infrastructure, or community health outreach programs, to ensure that individuals in remote areas receive adequate healthcare services. Additionally, while proximity to health facilities is important, ensuring that these centres are well-equipped with cholera treatment and prevention resources is equally vital in managing disease outbreaks effectively.

4.2 Frequency Results (Tables 2 to 10)

Table 2
Frequency Results for Cholera Cases

Cholera Case	Frequency	Percentage
0 (No Cholera Case)	137	39.1%
1 (Cholera Case)	213	60.9 %
Missing	0	0%
Total	**350**	**100%**

Table 2 above shows that 0-(No Cholera) has 39.1% (137 individuals) and 1- (cholera case) has 60.9% (213 individuals). This result indicates that the majority (60.9%) of individuals in the dataset had a cholera case, while 39.1% did not. This suggests a high prevalence of cholera in the study population.

Table 3
Frequency Results for Gender

Gender	Frequency	Percentage
Female	198	56.6%
Male	152	43.4%
Missing	0	0%
Total	**350**	**100%**

The frequency results for gender in Table 3, indicate that 56.6% of the study population is female, while 43.4% is male. This suggests that women are more vulnerable to cholera than men, due to their roles in water collection, hygiene practices and caregiving responsibilities. Additionally, future research should assess whether gender-based differences in exposure and health-seeking behaviour contribute to variations in cholera risk.

Table 4
Frequency Results for Access to Clean Water

Access to Clean Water	Frequency	Percentage
0	166	47.4%
1	184	52.6%
Missing	0	0%
Total	**350**	**100%**

Table 4 shows that, 47.4% of the population does not access clean water, which is a significant proportion. This lack of access increases vulnerability to cholera. 52.6% of individuals have access to clean water, which is slightly more than those without. While this is positive, it still means a large population segment is at risk. These results reinforce the importance of strengthening WASH infrastructure, particularly in rural and peri-urban areas with poor water access (Azman et al., 2020).

Table 5
Frequency Results for Access to Sanitation

Access to Sanitation	Frequency	Percentage
0	192	54.9%
1	158	45.1%
Missing	0	0%
Total	**350**	**100%**

Table 5 shows a complete dataset with 54.9% lacking access to sanitation. This suggests that more than half of the study population faces sanitation-related challenges, which could contribute to cholera risk. Fewer have access, and 45.1% have access to sanitation. This is a significant proportion but still lower than those without access. The findings indicate that poor access to sanitation is a significant issue, with more than half of the individuals lacking proper sanitation. Policymakers should prioritize investments in safe water supply, improved latrine coverage, and waste management systems to reduce the burden of waterborne diseases (Luquero et al., 2019).

Table 6
Frequency Results for Access to Health Facility

Access to Health Facility	Frequency	Percentage
0	198	56.6%
1	152	43.4%
Missing	0	0%
Total	**350**	**100%**

Table 6 shows that the majority,56.6% of individuals in the study, do not have access to a health facility, while a smaller proportion have access. This suggests that access to health facilities is a significant concern, hence there is a need for improved healthcare infrastructure.

Table 7
Frequency Results for Overcrowding

Overcrowding	Frequency	Percentage
0	143	40.9%
1	207	59.1%
Missing	0	0%
Total	**350**	**100%**

Table 7 shows that the majority, 59.1% of individuals, live in overcrowded conditions, while 40.9% are not crowded. Overcrowding is associated with poor sanitation, higher transmission rates of infectious diseases and increased vulnerability to cholera outbreaks. Measures such as improving sanitation infrastructure, promoting hygiene practices and reducing population density in high-risk areas could help mitigate cholera spread. Similar findings have been reported in studies from the Democratic Republic of Congo, where socio-economic deprivation contributes significantly to cholera risk (Moore et al., 2018).

Table 8
Frequency Results for Flood Risk

Flood Risk	Frequency	Percentage
0	150	42.9%
1	200	57.1%
Missing	0	0%
Total	**350**	**100%**

Table 8 shows that over half, 57.1% of the population, lives in flood-prone areas, increasing their risk of cholera infection. Flooding can contaminate water supplies, spreading bacteria and escalating cholera outbreaks. Flood mitigation strategies, such as improved drainage systems and pre-flood preparedness programs, should be introduced to reduce cholera risks.

Table 9
Frequency Results for Cholera Education

Cholera Education	Frequency	Percentage
0	162	46.3%
1	188	53.7%
Missing	0	0%
Total	**350**	**100%**

Table 9 shows that nearly half, 46.3% of individuals have not received cholera education, while 53.7% have received it. Lack of knowledge about cholera prevention may contribute to poor hygiene practices and increased disease spread. Strengthening cholera education campaigns, particularly in high-risk communities, can improve prevention and response strategies, reducing infection rates.

Table 10
Frequency Results for Season Risk Period

Season Risk Period	Frequency	Percentage
Dry	157	44.9%
Wet	193	55.1%
Missing	0	0%
Total	**350**	**100%**

Table 10 above shows that cholera risk appears to be higher during the wet season, 55.1%, likely due to increased water contamination from floods and poor drainage systems. This aligns with previous research by Moore et al. (2018) and Rebaudet et al., (2019), who indicated that heavy rainfall could lead to flooding, contamination of water sources, and increased cholera transmission. These findings emphasise the need for early warning systems based on meteorological data to predict and mitigate outbreaks before they escalate (Sasaki et al., 2018).

4.3 Fixed Effects Results

Table 11
Fixed Effects Estimate Results

Term	Estimate	95% CI Lower	95% CI Upper	R-hat	ESS
Intercept	5.634	1.852	10.061	1.000	2949.65
Age	-0.065	-0.164	0.008	1.004	1889.60
Gender (1)	-5.513	-11.342	-1.228	1.002	3079.85
Gender (2)	-3.632	-7.173	-0.948	1.002	2687.32
Access to clean water (1)	-0.939	-2.419	0.253	1.003	1648.17
Access to sanitation (1)	-0.142	-1.316	1.050	1.000	3743.76
Access to health facility (1)	-1.039	-2.383	0.086	1.000	2834.07
Overcrowding (1)	0.266	-0.953	1.478	1.000	4141.51
Flood risk (1)	0.691	-1.286	3.000	1.003	2067.72
Cholera Education (1)	-0.309	-1.631	0.858	1.000	2820.47
Average distance to health facility (km)	0.003	-0.592	0.530	1.001	2592.14
Season Risk Period (1)	6.143	3.649	9.114	1.001	2579.23

The Bayesian Generalised Linear Mixed Model based on Table 11 results, the fixed effects estimate is:

$$Log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)$$

$$= 5.634 - 0.065(Age) - 5.513(Gender\ 1) - 3.632(Gender\ 2)$$

$$-0.939\ (Access\ to\ clean\ water) - 0.142\ (Access\ to\ sanitation)$$

$$-1.039\ (Access\ to\ health\ facility) + 0.266(Overcrowding) + 0.691\ (Flood\ risk)$$

$$-0.309\ (Cholera\ Education) + 0.003(Average\ distance\ to\ health\ facility)$$

$$+6.143\ (Season\ Risk\ Period) + u_j \qquad (1)$$

Table 11 shows the fixed effects estimates from the Bayesian Generalized Linear Mixed Model (BGLMM), and equation (1) is the fixed effects model. Both provide valuable insights into factors influencing cholera risk. The intercept estimate of 5.634 (95% CI: 1.852 to 10.061) suggests that the baseline log-odds of cholera occurrence are relatively high when all predictors are at their reference levels. The Effective Sample Size (ESS) of 2949.645 indicates a stable estimate and the R-hat value of 1.000 confirms good model convergence.

Age has a negative estimate of -0.065 (95% CI: -0.164 to 0.008), suggesting that as age increases, cholera risk slightly decreases, though this effect is not statistically significant. Similarly, gender is a significant predictor, with -5.513 for females (gender (1)) and -3.632, for males (Gender (2)), indicating that females are more vulnerable to the effects of cholera. Both estimates have relatively high ESS values (3079.845 and 2687.319) and R-hat values close to 1, ensuring reliable results.

Access to clean water, sanitation, and health facilities generally shows protective effects, with estimates of -0.939, -0.142, and -1.039, respectively. However, their confidence intervals include zero, suggesting that these effects are not statistically significant. Despite this, the high Effective Sample Size (ESS) values and R-hat values around 1.000 indicate stable estimates. Overcrowding and flood risk have positive estimates (0.266 and 0.691), suggesting a potential increase in cholera risk, but their confidence intervals also include zero, making their effects uncertain. Among all predictors, the seasonal risk period has the strongest and most statistically significant effect, with an estimate of 6.143 (95% CI: 3.649 to 9.114). This suggests that the rainy season significantly increases cholera risk. The Effective Sample Size (ESS) of 2579.225 ensures stability, and the R-hat value of 1.001 confirms reliable convergence.

4.3.1 Model Validation

Table 12

Model Fit Statistics

WAIC	SE(WAIC)	LOO	SE (LOO)
110.12	21.21	112.7	21.17

From Table 12 above, WAIC (Widely Applicable Information Criterion) and LOO (Leave-One-Out Cross Validation) were used for model validation. WAIC (110.12, SE = 21.21) and LOO (112.7, SE = 21.17) provide measures of model fit. The relatively low values of WAIC and LOO suggest that the model fits the data well, meaning it effectively captures the relationship between the predictors and cholera risk.

4.4 Random Effects Results

Table 13

Districts Random Effects Estimates

District	Bikita	Chiredzi	Chivi	Gutu	Masvingo	Mwenezi	Zaka
Intercept	-0.016	0.038	-0.038	-0.005	0.092	-0.023	-0.013
Age	-0.016	0.063	-0.013	-0.016	0.003	0.008	-0.004
Gender (Female)	-0.008	0.000567	-0.002	-0.002	0.008	0.000123	-0.004
Gender (Male)	-0.005	0.046	-0.043	0.004	0.094	-0.047	-0.021
Access to Clean Water	-0.020	0.223	-0.024	-0.009	-0.002	-0.008	-0.108
Access to Sanitation	-0.048	0.059	-0.062	-0.024	0.251	-0.025	-0.080
Access to a Health Facility	-0.010	0.004	-0.003	-0.020	0.132	-0.033	-0.041
Overcrowding	0.022	0.021	-0.011	0.027	-0.111	-0.003	0.052
Flood Risk	0.042	-0.016	0.058	0.043	-0.201	0.056	0.003
Cholera Education	0.025	-0.134	0.028	0.023	-0.042	-0.002	0.080
Average Distance to Health Education(km)	-0.123	0.136	-0.099	-0.072	0.569	-0.080	-0.014
Season Risk Period (wet)	0.026	-0.018	0.055	0.035	-0.140	0.040	0.011

The random effects estimate from the Bayesian Generalized Linear Mixed Model (BGLMM) reveals significant district-level variations in cholera risk across Masvingo Province. Below is an interpretation of each risk factor based on the random effect estimates provided in Table 13. Masvingo (0.092) and Chiredzi (0.038) districts exhibit the highest baseline cholera risks, suggesting that population density and sanitation challenges contribute to more severe outbreaks. Conversely, districts like Chivi (-0.038), Mwenezi (-0.023), Zaka (-0.013), Bikita (-0.016) and Gutu (-0.005) show lower baseline risks.

Chiredzi exhibits the highest positive association (+0.063) in terms of age, indicating that increasing age correlates with a higher risk of cholera in this district. This suggests that older individuals in Chiredzi may be more vulnerable to cholera due to weakened immunity, prolonged exposure to contaminated water sources or limited access to healthcare. Conversely, districts like Bikita (-0.016), Chivi (-0.013), Gutu (-0.016) and Zaka (-0.004) show negative values, implying that older individuals in these areas may have a slightly lower risk. This could mean that younger individuals in these districts face a greater burden of cholera, due to increased mobility, exposure at communal water points or lack of awareness.

The results indicate a notable difference in cholera risk between males and females across the districts. In Masvingo, males (+0.094) are significantly more at risk compared to females (+0.008), and a similar trend is observed in Chiredzi, where males (+0.046) also exhibit higher susceptibility than females (+0.00057). In contrast, Chivi (-0.002 for females, -0.043 for males) and Zaka (-0.0004 for females, -0.021 for males) have negative values, suggesting a lower cholera risk for both genders, with minimal influence of gender on disease transmission. The increased risk for males in Masvingo and Chiredzi could be attributed to occupational exposure, particularly in outdoor activities such as farming or fishing, which involve frequent contact with contaminated water sources.

Flooding emerges as a key driver of cholera outbreaks, with districts such as Bikita (+0.042), Gutu (+0.043), Chivi (+0.058) and Mwenezi (+0.056 showing positive associations between flood risk and cholera cases. This highlights the role of waterborne transmission, where contaminated floodwaters may spread Vibrio cholerae bacteria. Masvingo (+0.251) shows a strong positive association with access to sanitation, meaning individuals with access have a higher risk due to inadequate sanitation infrastructure or shared facilities. Zaka (-0.080), Chivi (-0.062), Bikita (-0.048), Mwenezi (-0.025) and Gutu (-0.024) show negative associations, suggesting better sanitation access reduces cholera risk. In most districts, sanitation reduces risk. However, in Masvingo, sanitation infrastructure might be inadequate despite access.

Access to clean water in Chiredzi (+0.223) correlates with higher cases. This suggests that while water availability has improved, water quality remains a critical issue, potentially due to contamination at the source or during distribution. Zaka (-0.108), Bikita (-0.020), Chivi (-0.024), Mwenezi (-0.008), Gutu (-0.009) and Masvingo (-0.002) have negative estimates, meaning better water access reduces cholera risk. Overcrowding and healthcare accessibility also significantly influence cholera cases. The higher population density in districts like Zaka (+0.052), Gutu (+0.027), Bikita (+0.022) and Chiredzi (+0.021) is associated with increased outbreaks, reinforcing the need for improved sanitation infrastructure in densely populated areas. Additionally, greater distances to health facilities, particularly in Masvingo (+0.569) and Chiredzi (+0.136), increase cholera risk by delaying timely treatment and containment efforts.

Cholera education reduces cases in Chiredzi (-0.134) and Masvingo (-0.042), but its impact is less pronounced in other districts. In areas like Zaka (+0.080), Bikita (+0.025), Chivi (+0.028) and Gutu (+0.023), education campaigns alone are not enough without infrastructure improvements. Seasonal trends further influence cholera dynamics, with most districts experiencing heightened cases during the wet season due to increased water contamination, except for Masvingo (-0.140) and Chiredzi (-0.018), where dilution effects may play a role in lowering the risk.

4.5 Spatial Analysis Results

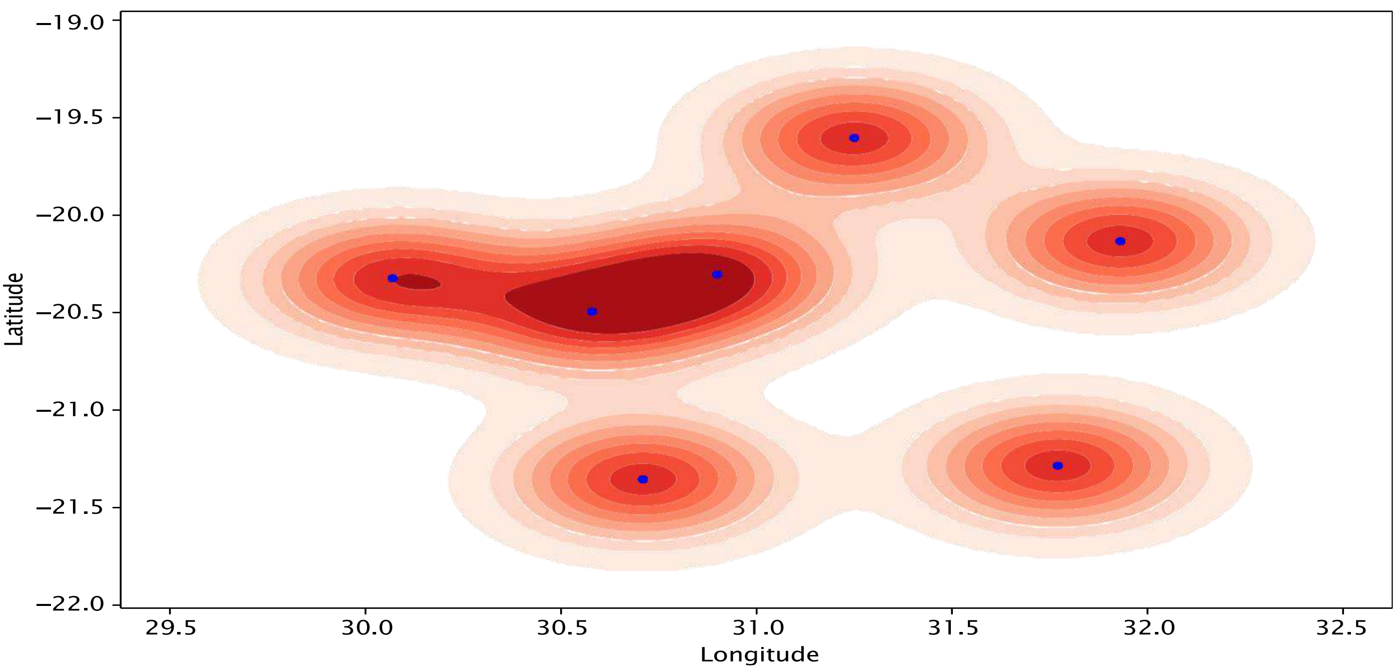

Figure 2
Geospatial Cholera Case Density Heatmap

The heatmap in Figure 2 above visually represents cholera case density across Masvingo, using a colour gradient. Areas shaded in light pink or white signify regions with little to no reported cases, while light red areas reflect low to moderate case densities. The darkest red regions highlight the most severe outbreaks, indicating locations with the highest cholera case concentrations. The blue dots scattered across the map represent Masvingo, Bikita, Chivi, Chiredzi, Mwenezi, Gutu and Zaka districts, serving as the foundation for generating the density estimation. This process results in the red-shaded intensity distribution, illustrating how cases are spatially spread. The smooth, contour-like patterns are produced using kernel density estimation (KDE), a common technique for illustrating spatial intensity. The red shading is darker in areas where blue dots are densely clustered, whereas regions with fewer cases appear in lighter tones. This heatmap effectively identifies outbreak hotspots and areas at risk. Based on the intensity ranking, Masvingo and Chiredzi are in the darkest red regions, indicating the highest cholera cases. Chiredzi shows the highest concentration of cases due to poor sanitation, high population density, irrigation-related water contamination and cross-border movements. The region's location in the Lowveld, along with its extensive sugarcane plantations and water canals, further contributes to the disease spread. Masvingo district is home to the provincial capital, which faces urban water supply and sewer management challenges. Mwenezi has a severe cholera outbreak due to its reliance on boreholes and unprotected water sources, especially during droughts, which increases its vulnerability. Moderate-risk areas include Chivi and Gutu, which report significant but less severe cholera outbreaks than Chiredzi, Masvingo and Mwenezi. These rural districts have limited access to clean water and sanitation, with communities often relying on seasonal rivers, increasing their susceptibility to cholera. In contrast, Bikita has a lower cholera case density due to better access to clean water or a lower population density in affected zones. Zaka records the lowest case density in the province but still needs preventive measures to avoid escalation**.** While some districts, such as Gutu, Bikita and Zaka, experienced lower-case densities, their sporadic cases highlight the potential for localized outbreaks if preventive measures are not sustained. The results suggest the need for tailored intervention strategies for district-specific vulnerabilities (Diggle & Giorgi, 2019). It supported the study results. The spatial analysis and Bayesian Generalised Linear Mixed Model results show that public health interventions are essential. One of the study's key strengths is the use of Bayesian Generalized Linear Mixed Models (BGLMM), which allowed for a more robust analysis of both fixed and random effects. This approach effectively accounted for spatial and temporal heterogeneity in cholera risk, reducing biases associated with traditional regression models (Luquero et al., 2019).

The findings are consistent with previous studies on cholera risk factors in Africa and other regions. For example, similar associations between rainfall, proximity to water sources, socio-economic status, and cholera risk have been reported in studies from Bangladesh (Moore et al., 2018) and the Democratic Republic of the Congo (Rebaudet et al., 2019). This suggests that the factors driving cholera transmission are likely to be universal across different contexts, although local variations exist.

V. RESEARCH IMPLICATIONS

This study has several important research implications. The Bayesian Generalized Linear Mixed Model (BGLMM) highlights the effectiveness of advanced statistical methods in analysing infectious disease patterns, particularly in resource-limited settings. It is a valuable tool for future epidemiological research because it accounts for both fixed and random effects. Again, the model discovered significant risk factors such as healthcare access, gender and seasonal variation that provide critical evidence to inform targeted public health policies and interventions. The observed district-level variations in cholera incidence underscore the need for localised, context-specific strategies, particularly in high-risk areas like Masvingo, Chiredzi and Mwenezi. Additionally, the introduction of geospatial heatmaps demonstrates the importance of spatial analysis in mapping disease distribution, offering a powerful approach for the future. This study also establishes a foundation for longitudinal analyses, enabling researchers to track the impact of interventions over time. The findings encourage a multi-sectorial approach that integrates public health, environmental management and socio-economic development to effectively address the complex determinants of cholera transmission in Zimbabwe and other cholera-prone regions.

VI. CONTRIBUTIONS TO SCIENTIFIC COMMUNITY AND FUTURE RESEARCH

This study provides valuable insights into the factors driving cholera transmission in Masvingo Province, Zimbabwe. It offers a foundation for more effective public health policies and interventions in similar settings. Future research should prioritize high-resolution data collection, longitudinal studies, and the integration of behavioural factors to enhance cholera risk assessment and intervention planning. Future research should focus on improving data quality, refining model specifications, and using advanced computational methods to enhance the applicability of Bayesian Generalised Linear Mixed Models for cholera risk modelling.

VII. CONCLUSION

The Bayesian Generalized Linear Mixed Models (BGLMM) results highlight the complex interplay between environmental, socio-economic and infrastructural factors in driving cholera transmission. Rainfall was identified as a significant predictor of outbreaks, reinforcing the role of seasonal variations in cholera epidemiology. Districts with inadequate sanitation and limited access to clean water, such as Chiredzi, Masvingo and Mwenezi, exhibited the highest cholera risk, underscoring the urgent need for improved WASH infrastructure. The spatial analysis revealed distinct clustering patterns, with high-risk areas requiring targeted interventions such as enhanced surveillance, vaccination programs, and improved sanitation services. Despite its strengths, the study acknowledges limitations, including data quality constraints and the need for finer-scale spatial and temporal resolution.

VIII. RECOMMENDATIONS

To effectively reduce cholera outbreaks in Masvingo Province, Zimbabwe, a comprehensive and integrated approach is needed to mitigate cholera outbreak. Strengthening water, sanitation, and hygiene (WASH) infrastructure should be a priority, ensuring expanded access to clean water and improved sanitation facilities, particularly in high-risk districts. Household-level interventions such as water treatment and safe storage should also be promoted. Public health interventions must focus on targeted seasonal cholera vaccination campaigns, early warning systems for floods and outbreaks, and enhanced community-based surveillance to facilitate rapid response. Additionally, improving healthcare accessibility by expanding health infrastructure, training healthcare workers in cholera management, and deploying mobile clinics to remote areas is essential in reducing mortality and morbidity. Community education and behaviour change initiatives should be intensified to promote proper hygiene practices, including handwashing, food safety and sanitation habits. Engaging local leaders, schools, and community groups will help drive sustainable behaviour change. Environmental management strategies should also be implemented to mitigate flood risks, including improving drainage systems, developing climate-resilient infrastructure, and promoting proper waste management to prevent water contamination. Furthermore, district-specific interventions should be designed based on localized risk factors, ensuring that resources are allocated efficiently to high-burden areas while encouraging community-driven solutions. By implementing these strategies, stakeholders can work collaboratively to reduce cholera incidence, improve public health resilience, and safeguard vulnerable populations from future outbreaks.

IX. FUNDING

This research paper received no internal or external funding.

X. ACKNOWLEDGEMENT

All sources were acknowledged.

XI. CONFLICT OF INTEREST

This manuscript was not submitted nor is it being reviewed by another journal or other publishing centres.

REFERENCES

Ayling, S., Milusheva, S., Maidei Kashangura, F., Hoo, Y. R., Sturrock, H., & Joseph, G. (2023). A stitch in time: The importance of water and sanitation services (WSS) infrastructure maintenance for cholera risk: A geospatial analysis in Harare, Zimbabwe. *PLOS Neglected Tropical Diseases, 17*(6), e0011353. https://doi.org/10.1371/journal.pntd.0011353

Azman, A. S., Moore, S. M., Rumunu, J., Abubakar, A., West, H., Sack, D. A., & Lessler, J. (2020). Micro-hotspots of risk in urban cholera outbreaks. *The Lancet, 395*(10232), 120–124. https://doi.org/10.1016/S0140-6736(19)32937-6

Chikoto, G., & Hughes, G. (2019). Spatial analysis of cholera outbreaks in Zimbabwe. *African Journal of Epidemiology, 14*(3), 221–231.

Cowman, G., Otipo, S., Njeru, I., Achia, T., Thirumurthy, H., Bartram, J., & Kioko, J. (2017). Factors associated with cholera in Kenya, 2008–2013. *Pan African Medical Journal, 28*(1), 156.

Diggle, P. J., & Giorgi, E. (2019). *Model-based geostatistics for global public health: Methods and applications*. CRC Press.

Dureab, F., Jahn, A., Krisam, J., Dureab, A., Zain, O., Al-Awlaqi, S., & Müller, O. (2019). Risk factors associated with the recent cholera outbreak in Yemen: A case-control study. *Epidemiology and Health, 41*, e2019015. https://doi.org/10.4178/epih.e2019015

Elimian, K. O., Musah, A., Ochu, C. L., Onwah, S. S., Oyebanji, O., Yennan, S., Fall, I. S., Yao, M., Chukwuji, M., Ekeng, E., Abok, P., Omar, L. H., Balde, T., Kankia, A., Williams, N., Mutbam, K., Dhamari, N., Okudo, I., Alemu, W., Peter, C., … Ihekweazu, C. (2020). Identifying and quantifying the factors associated with cholera-related death during the 2018 outbreak in Nigeria. *The Pan African medical journal*, *37*, 368. https://doi.org/10.11604/pamj.2020.37.368.20981

Emch, M., Yunus, M., Escamilla, V., Feldacker, C., & Ali, M. (2017). Local environmental predictors of cholera in Bangladesh and their implications for global control. *Journal of Infectious Diseases, 218*(Suppl 3), S180–S186. https://doi.org/10.1093/infdis/jiy520

Fong, Y., Rue, H., & Wakefield, J. (2021). Bayesian inference for generalised linear mixed models. *Biostatistics, 22*(1), 69–82. https://doi.org/10.1093/biostatistics/kxz010

Held, L., & Rue, H. (2022). Conditional and intrinsic autoregressive models for disease mapping. *Statistical Methods in Medical Research, 31*(2), 202–217. https://doi.org/10.1177/09622802211048248

Lawson, A. B. (2021). *Bayesian disease mapping: Hierarchical modelling in spatial epidemiology* (3rd ed.). CRC Press.

Leo, J., Luhanga, E., & Michael, K. (2019). Machine learning model for an imbalanced cholera dataset in Tanzania. *The Scientific World Journal, 2019*, 9397578. https://doi.org/10.1155/2019/9397578

Li, Y., Wakefield, J., & Touloumi, G. (2020). A comparison of Bayesian spatio-temporal models for disease mapping. *Statistical Methods in Medical Research, 29*(5), 1230–1253. https://doi.org/10.1177/0962280219861582

Luquero, F. J., Banga, C. N., Remartinez, D., Palma, P. P., Baron, E., Grais, R. F., & Checchi, F. (2019). Estimating the impact of emergency cholera vaccination in Haiti: A modelling study. *PLOS Medicine, 16*(2), e1002747. https://doi.org/10.1371/journal.pmed.1002747

Moore, S. M., Azman, A. S., Zaitchik, B. F., Mintz, E. D., Brunkard, J. M., & Legros, D. (2018). El Niño and the shifting geography of cholera in Africa. *Proceedings of the National Academy of Sciences, 115*(9), 2210–2215. https://doi.org/10.1073/pnas.1711313115

Mugabe, T., Masuku, S., & Sibanda, P. (2021). Cholera outbreaks in Zimbabwe: A review of risk factors. *African Journal of Public Health, 15*(2), 203–215.

Munyenyiwa, A., Musuka, G., Makuvaza, M. J. M., Mucheri, T., & Midzi, N. (2022). A review of the 2008/2009 and 2018/2019 cholera outbreaks in Zimbabwe: Call for public health action. *Journal of Infectious Diseases and Epidemiology, 8*, 269. https://doi.org/10.23937/2474-3658/1510269

Ngereza, C., Katapa, R. S., & Mniachi, A. R. (2021). Environmental factors associated with re-emergency of cholera in Tanzania using Poisson regression models. *Tanzania Journal of Sociology, 7*(1), 80–100.

Rebaudet, S., Sudre, B., Faucher, B., & Piarroux, R. (2019). Environmental determinants of cholera outbreaks in inland Africa: A systematic review. *The Journal of Infectious Diseases, 219*(12), 1935–1944. https://doi.org/10.1093/infdis/jiz129

Rinaldo, A., Blokesch, M., Bertuzzo, E., Mari, L., & Gatto, M. (2020). Modelling and analysing the transmission dynamics of cholera at two spatial scales. *Infectious Disease Modelling, 5*, 689–705. https://doi.org/10.1016/j.idm.2020.08.001

Traerup, S. L., Ortiz, R. A., & Markandya, A. (2020). The health impacts of climate change: A study of cholera in Tanzania. *Global Environmental Change, 60*, 102042. https://doi.org/10.1016/j.gloenvcha.2019.102042

World Health Organisation. (2021). *Ending cholera: A global roadmap to 2030*. World Health Organisation.

World Health Organisation. (2023, December). *WHO Zimbabwe Cholera Update*.

World Health Organisation. (2024, December). *Cholera Report*.

Zimbabwe Meteorological Services Department. (2023). *Annual rainfall and temperature data*. Zimbabwe Meteorological Services Department.

Zimbabwe Ministry of Health and Child Care. (2023). *Annual report on communicable diseases*. Ministry of Health and Child Care.

Zimbabwe National Statistics Agency. (2022). *Poverty, income, consumption, and expenditure survey 2022*. Zimbabwe National Statistics Agency.