# Environmental Factors Associated with Re-Emergency of Cholera in Tanzania using Poisson Regression Models

***Caroline Ngereza*[1], *Rosalia S. Katapa*[2] *and Ali R. Mniachi*[3]**
[1]*Lecturer, Department of Population Studies, Institute of Rural Development Planning*
[2]*Associate Professor, Department of Statistics, University of Dar es Salaam*
[3]*Senior Lecturer, Department of Statistics, University of Dar es Salaam*
*E-mail addresses: cngereza@irdp.ac.tz, rskatapa@yahoo.com, amniachi@gmail.com*

## Abstract

*In Tanzania, periodical re-emergence of cholera has been influenced by different risk factors. Specifically, the study aimed at examining risk factors associated with the number of cholera cases and determining the best model. The risk factors were examined using Poisson regression (PR) and the Geographically Weighted Poisson regression (GWPR) models. The best model was determined using the Akaike Information criterion (AIC) and Bayesian Information criterion (BIC) values. The results showed that, AIC and BIC values of GWPR model were 456.2 and 484.6, respectively; and for the PR model were 1323.5 and 13239.8, respectively. This implies that GWPR model was better for assessing risk factors of cholera cases compared to PR model since it has small AIC and BIC values. The results of the GWPR model showed that, accessing improved water source (M=-0.003), practicing hand-washing (M=-0.371) and open defecation (M=-0.169) have inverse relationship with the number of cholera cases. Accessing improved sanitation facility (M=0.156) in the household has shown a positive relationship with the number of cholera cases. The unexpected observed relationship between practicing open defecation and accessing improved sanitation facility with the number of cholera incidents could probably be due to the fact that, application of a single measure may not be enough measure of controlling cholera in a household. However, risk factors found to have significant association with the number of cholera cases vary across regions. It is concluded that the GWPR model was the better model for the purpose. Therefore, measures taken to control and prevent cholera disease should be based on the variations of the risk factors found in the regions.*

**Keywords:** Cholera, Poisson Regression, Geographically weighted Poisson Regression, Tanzania

## Background

Cholera remains a global problem to the public health and a sign of imbalance and lack of social development. Worldwide, it is estimated to have 2.9 million cases and 95,000 deaths yearly (Ali et al., 2015). For a long time, the Asian Sub-continent was home to cholera (Harris et al., 2012). Currently, there are persistent cholera outbreaks on the African Continent with large disease burden and high rate of case fatality. The majority of the reported cases are from the Sub-Saharan Africa Region, which contributes about 60% of cases and deaths globally (Ali et al., 2015).

In Tanzania, during the 7$^{th}$ pandemic, the first cholera incidence was reported in 1974, since then it has been reported almost every year with approximation of 5,800 cases (URT, 2014). By 2018, over 250,000 cholera cases and 13,078 deaths were noted in Tanzania (Lessler, 2018). In 2015, Tanzania experienced cholera outbreaks with 8,821 cases and 91 deaths. The Government has been fighting a widespread incidence of cholera since August 2015 (Narra et al., 2017) which led to a remarkable drop in the number of cholera cases and deaths; for the year 2018 there were 4,365 cases and 82 deaths (WHO, 2018). Country wise, the number of cholera cases and deaths has declined, however, the current trend of its periodical re-emergence is alarming, and the number of cases and deaths in some regions is increasing (UNICEF, 2018).

Regression model analysis has been mostly applied in cholera studies. The model is among the main statistical tools for understanding functional relationships between dependent variables and the explanatory variables (Chatterjee & Hadi, 2015). Many studies on cholera which have previously been carried out have used non-spatial regression models to determine the relationships. Poisson, Negative Binomial, Zero Inflated and Hurdlers models are often used (Rajasingham et al., 2019; Trærup et al., 2011) which are referred to as traditional regression methods, and assume that the parameters of the models remain the same for the entire study area (Oluwajana, 2018).

Non-spatial regression models were used globally in assessing risk factors of cholera by several researchers. Zero inflated models were applied by Richterman et al. (2019) across 30 countries, (Cowman et al., 2017) in Kenya, (Bwire et al., 2017) in Uganda, and (Hounmanou et al., 2019) in Tanzania. Literature also shows other non-spatial count data models used in examining risk factors of cholera are negative binomial regression model (Penrose et el., 2010; Kazaji, 2016) and Poisson regression model (Page et al., 2015).

These models were relevant since cholera cases are count data; count data are characterized as non-negative and discrete events that regularly occur within a certain time interval, tend to be highly skewed and non-normally distributed. However, the models were appropriate with the type of data used, but did not consider geographical variation between explanatory and dependent variables (Fotheringham et al., 2002). The assumption of a fixed parameter for data from a geographical reference is often incorrect (Li et al., 2011; Meng, 2014). Thus, accommodating spatial dependencies in regression models becomes important. Specifying local relationships across space requires the use of spatial regression models; spatial error models, spatial lag models, and the geographically weighted regression (GWR) models (Bernasco & Elffers, 2010).

Osei and Duker (2008) fitted spatial lag and error models in exploiting the association between cholera incidence and open space refuse dump in Ghana. Camacho et al. (2018) estimated the relationship between rainfall and the daily cholera incidence during the increasing phase of the second epidemic wave of April to June, 2017 in Yemen by fitting a spatial-temporal regression model. However, neither the spatial error nor the spatial lag model estimates a local regression equation for all data points. The models do not give an indication of how the relationship between the dependent and the explanatory variables varies across space.

Brunsdon et al. (1998) proposed Geographically Weighted Regression model to overcome this limitation. The model estimates local and separate regression equation for each data point. Nkenki and Osirike (2013) used GWR model to explore and analyze the spatial relationships between cholera occurrence and household sources of water supply. In investigating the impact of tube well user density on cholera events in Matlab, Bangladesh; household-level demographic, health, and water infrastructure data were incorporated and GWR models were constructed to identify spatial variation of relationships across the study area (Carrel et al., 2011). Geographically Weighted Poisson regression (GWPR) has been applied for analyzing disease maps (Nakaya et al., 2005; Zhou et al., 2019), also for crime and collision prediction (Oluwajana, 2018). It is evident that identifying a cause of problem at local level help in eliminating problem at high level. Despite the importance of application for spatial local model, there is limited application of GWPR model on cholera diseases. Therefore, application of GWPR model in this study was inevitable.

In this study spatial (GWPR) and non-spatial Poisson regression (PR) models were used for assessing risk factors of cholera on the number of cholera cases. Specifically, the study aimed to examine risk factors associated with the number of cholera cases and to compare the performance of PR model with GWPR model.

## Methodology

### Data Source

### Dependent Variables: Cholera Cases

Number of cholera cases in all regions of Tanzania mainland from the year 2015 to 2018 was used in the current study. The data were collected by the Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC). The current study used this data because the recent cholera outbreak started from 2015 to 2018 (WHO, 2018), in 2019 to 2020 there were few cases. The study assumed that, understanding past cholera incidences is essential for preventing future incidences, which also has been supported by (Cowman et al., 2017; Eltoukhy et al., 2020).

### Explanatory Variables

Explanatory variables which were used in this study includes percentage of households with access to improved water sources, percentage of households with improved sanitation facility, percentage of households which practiced open defecation, percentage of households which shared toilets and percentage of household practiced hand washing, both were at regional level. These variables were extracted from 2015/2016 TDHS-MIS. The survey provides data for household characteristics (households which had an access to improved water sources, households with improved sanitation facility, households which practiced open defecation, households which shared toilets and households which practiced hand washing) in Tanzania. Description of variables is provided in Table 1.

**Table 1: Description of variables**

| Variable | Description |
|---|---|
| Percentage of households which had an access to improved water sources | Improved water sources include: piped to dwelling, piped to yard/ plot, piped to neighbour, public tap/ standpipe and borehole or pump. |
| Percentage of households with improved sanitation facility | Improved sanitation facilities include: flush to septic tank, flush to pit latrine, ventilated improved pit latrine, and pit latrine with slab. |
| Percentage of households which practiced open defecation | Open defecation include: no facility/bush/field, composting toilet, bucket toilet, hanging toilet/latrine and other. |

### Population and Centroid Coordinates

Population projections of 2018 were obtained from the National Bureau of Statistics (NBS, 2019). The centroid coordinates (latitudes and longitudes) of each region was obtained from the regional shape-file extracted from NBS using ArcGIS v10.2.2. Since all data were at the regional level (cholera cases, household variables, population and centroid coordinates) were entered in one Microsoft Excel spreadsheet for analysis.

### Data Analysis

Descriptive statistics and analytical methods were done. In describing variables; mean, minimum, maximum and standard deviation were presented. Poisson and Geographically Weighted Poisson regression analysis were conducted to determine risk factors related to the number of cholera cases. Then goodness-of-fit and model selection were based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values (Akaike, 1973; Schwarz, 1978; Burnham and Anderson, 2004). Also, pseudo R squared value was used for selection of the best model. Before generating the regression models, selected explanatory variables were tested for multicollinearity to check the presence of high correlation between them. Variance Inflation Factor (VIF) values of ordinary least square regression model were used to detect multicollinearity.

Those variables with VIF value of less than 5 and a tolerance of more than 0.2 were used in fitting count data models (O'Brien, 2007).

Stata v14.0 and MGWR 1.0 software packages were used in analyzing Poisson regression models. The variable which had P<0.05 was considered to be a statistically significant risk factor associated with the number of cholera cases. In analyzing Geographically Weighted Poisson analysis, adaptive bi-square function was used to determine weights for each observation. An optimal bandwidth search from 4 to 25 regions in step of one region in the model was done using the AIC value, nine nearest neighbour regions were obtained as an optimal bandwidth which had the smallest AIC value (Su et al., 2019). The selection of an optimal bandwidth was crucial so as to minimize the mean square error of estimates (Aydın et al., 2021; Chen, 2015). To visualize estimated values and p-value of the parameters estimated in each region, a choropleth map was generated. It was normalized with the area coverage polygons (in $km^2$) by region and a five-class natural breaks (Jenks) classification method based on Nkeki and Osirike (2013) (Figures 1 and 2).

## Model Specification

Both the PR and the GWPR models were used for examining the risk factors associated with the number of cholera cases.

## Poisson Regression Model (PR)

The Poisson regression model is the standard one for count data (number of cholera cases), which is a nonlinear regression model. The model assumes that cholera cases follow Poisson distribution with a random variable $y$ as defined by Mood et al. (1974):

$$p(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0,1,2,\ldots$$

Where $y$ is the number of cholera cases for a chosen time period (2015-2018), and $\lambda$ is the expected number of cholera cases per region (parameter $\lambda$ satisfy $\lambda > 0$), where $\lambda = E(y) = Vay(y)$. Poisson regression assumes that Poisson distributed random variables $y$, and the logarithm of the expected value of $y$ can be modeled by a linear combination of the unknown parameters of explanatory variables. Since the average $(\lambda)$ must be positive, it takes a log link function for parameter $\lambda$. The link function and linear predictor determine the functional form of the model.

The log link function with linear predictor form of Poisson regression model is given by:

$$\ln E(y/x) = \ln(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2_i} + \ldots + \beta_k x_{ki}$$

Where $x$'$s$ is the set of $k$ explanatory variables of observation $i$, $(x_{1i}, x_{2i},\ldots, x_{ki})$ and $\beta_0, \beta_1,\ldots,\beta_k$ are coefficients of their corresponding explanatory variables.

**Geographically Weighted Poisson Regression (GWPR)**

The geographically weighted Poisson regression model is a modelling technique which exposes the spatial heterogeneity among the outcome and explanatory variables that allow the estimation of the regression coefficient to vary with location. According to Nakaya et al. (2005), the model takes the form of:

$$\ln(y_i) = \ln(p_i) + \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon$$

Where; $y_i$ is the number of cholera cases of region $i$, $p_i$ is the population at risk of region $i$, $(u_i, v_i)$ is the centroid coordinates (latitudes and longitudes) of region $i$, $x_{ik}$ is the $k^{th}$ explanatory variables of region $i$, $\beta_k$ is the coefficient of the $k^{th}$ explanatory variable.

GWPR uses a conditional spatial kernel weighting function for the estimation of the parameter variation of the Poisson regression method. Adaptive bi-square bandwidth was used in determining geographical weights, since it performs better when the data distribution is sparse and dense (Fotheringham et al., 2002). Adaptive bi-square kernel function is of the form:

$$w_{ij} = \begin{cases} 1 - \left(\dfrac{d_{ij}}{b_i}\right)^2 & d_{ij} \le b_i \\ 0 & d_{ij} > b_i \end{cases}$$

Where; $w_{ij}$ is the weight value of cholera cases at region $j$ for estimating the coefficient at region $i$, $d_{ij}$ is the distance between $j^{th}$ and $i^{th}$ neighbour regions, and $b_i$ is the kernel bandwidth size.

## Results and Discussion

### Results

In Table 2, the average number of cholera cases per region was found to be 1000 with the maximum number of 5657 cholera cases. The results show that, there were variations of; 19.31 from 34.89percentage of households' access improved water sources, and 18.20 from 27.89percentage of households with improved sanitation. Variation of percentage of households which practiced open defecation and percentage of households which shared toilets were 9.24 from 10.85 and, 8.78 from 25.49, respectively. Variation of the percentage of household practiced hand washing was 2.89 from 5.53.

### Multicollinearity Detection

The result shows moderate collinearity between the explanatory variables, since VIF values are less than 5 and corresponding tolerance statistic are more than 0.2. This means all explanatory variables are suitable in model fitting for the PR and GWPR since there is little influence of one explanatory variable on all other explanatory variables. The results are presented in Table 2.

**Table 2: Descriptive statistics of variables and Multicollinearity detection**

| Variable | Description (per region) | Mean | Standard Deviation | Minimum | Maximum | Multicollinearity detection | |
|---|---|---|---|---|---|---|---|
| | | | | | | VIF | Tolerance (1/VIF) |
| Number of cholera cases | Number of cholera cases | 1000.00 | 1332.00 | 0.00 | 5657.00 | - | - |
| Water | Percentage of household which had an access to improved water sources | 34.89 | 19.31 | 11.93 | 92.43 | 1.67 | 0.60 |
| Sanitation | Percentage of households with improved sanitation facility | 27.89 | 18.20 | 9.00 | 93.79 | 2.93 | 0.34 |

| Variable | Description (per region) | Mean | Standard Deviation | Minimum | Maximum | Multicollinearity detection | |
|---|---|---|---|---|---|---|---|
| | | | | | | VIF | Tolerance (1/VIF) |
| Defecation | Percentage of households which practiced open defecation | 10.85 | 9.24 | 0.03 | 27.59 | 1.26 | 0.80 |
| Toilet | Percentage of households which shared toilets | 25.49 | 8.78 | 11.93 | 45.52 | 1.45 | 0.69 |
| Hygiene | Percentage of households which practiced handwashing | 5.53 | 2.89 | 1.63 | 12.68 | 1.59 | 0.63 |

## Comparison of Model Fitting Performance of PR and GWPR

Table 3 presents the results of PR and GWPR models and their AIC and BIC values. For goodness model fit on variables, results showed that, AIC and BIC values of GWPR model were 456.2 and 484.6, respectively; and for PR model were 1323.5 and 13239.8, respectively. This means GWPR model was better for assessing risk factors of cholera cases compared to PR model since it has small AIC and BIC values. For the case of $R^2$ values, GWPR model fit data better than PR model since it has large values of 0.98 compared to 0.62 of PR. These results shown that GWPR model is the best for predicting cholera occurrence in a country compared to PR model. Therefore, interpretation and discussion was based on the results of GWPR model.

## Risk Factors Associated with the Number of Cholera Cases by GWPR

Geographically Weighted Poison estimates were interpreted based on the sign of the coefficients' mean value (Chen et al., 2020). The percentage of household which access to improved water sources has shown a negative association on the number of cholera cases (M= -0.003), while percentage of households with improved sanitation facility had a positive association on the number of cholera cases (M= 0.156). Negative association was found between number of cholera cases and percentage of households which practiced open defecation in the regions (M= -0.169). For the case of percentage of households which shared toilets it has shown positive relationship on the number of cholera cases (M=0.081), while estimated parameters values of percentage of households practiced hand washing has shown a negative association (M=-0.371) on the number of cholera cases (Table 3).

**Table 3: Parameter estimates of PR and GWPR models**

| Variables | PR | | | GWPR | | | |
|---|---|---|---|---|---|---|---|
| | Estimates | 95% CI | p-value | Mean (M) | Minimum | Median | Maximum |
| Intercept | 4.382 | 4.327-4.437 | <0.001 | 1.024 | -5.328 | 1.847 | 5.666 |
| Water | 0.004 | 0.003-0.005 | <0.001 | -0.003 | -0.255 | -0.001 | 0.385 |
| Sanitation | -0.002 | -0.003--0.001 | <0.001 | 0.156 | -0.083 | 0.084 | 0.795 |
| Defecation | -0.023 | -0.246--0.021 | <0.001 | -0.169 | -1.202 | -0.057 | 0.148 |
| Toilet | 0.074 | 0.073-0.076 | <0.001 | 0.081 | -0.478 | 0.062 | 0.493 |
| Hygiene | 0.083 | 0.077-0.088 | <0.001 | -0.371 | -2.677 | -0.157 | 0.379 |
| $R^2$ | 0.62 | | | 0.98 | | | |
| AIC | 13232.57 | | | 456.23 | | | |
| BIC | 13239.88 | | | 484.68 | | | |

## Variation of the Risk Factors across the Regions

Estimates variations of the risk factors across the regions are presented in Figure 1.A positive association between percentage of households which had access to improved water sources and number of cholera cases observed in all regions except for Ruvuma, Rukwa, Mwanza, Kigoma, Katavi, Iringa, Njombe, Lindi, Mtwara and Simiyu. All estimated parameters in regions had significant impact on number of cholera cases (p< 0.05), except for Geita and Dodoma Regions (Figure 2).

A positive association between percentage of households with improved sanitation facility and number of cholera cases observed in all regions except for Manyara, Kilimanjaro, Ruvuma, Lindi and Mtwara Regions. All estimated parameters in regions had significant impact on number of cholera cases, except for Singida Region (Figure 2).

In Arusha, Mbeya, Njombe and Tabora Regions; there were positive relationship between number of cholera cases and percentage of households practicing open defecation. However, insignificant association between percentage of households which practiced open defecation and number of cholera cases was found in Iringa, Manyara and Mwanza and Shinyanga Regions (Figure 2).

Almost all regions showed a positive association between percentage of households which shared toilets and number of cholera cases. However, in Mwanza and Simiyu Regions; the percentage of households which shared toilets had shown no significant association on number of cholera cases (Figure 2).

A negative association between the percentage of households which practiced handwashing and the number of cholera cases was observed in all the regions except in Kilimanjaro, Manyara, Mwanza, Kigoma, Rukwa, Mbeya, Iringa, Njombe and Ruvuma Regions. Percentage of households which practiced handwashing had shown no significant association on the number of cholera cases in Kigoma and Singida Regions (Figure 2).

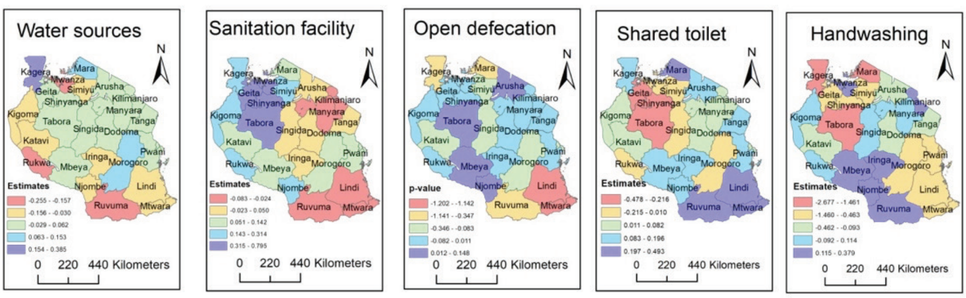


**Figure 1: Spatial distributions of environmental risk factors on cholera cases**

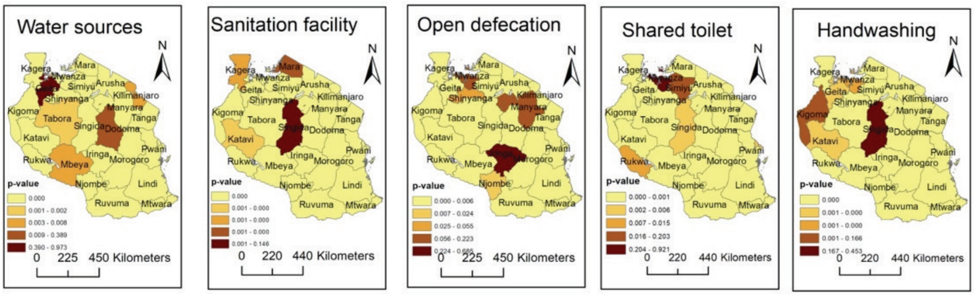


**Figure 2: Spatial distributions of significance levels of the estimates**

## Discussion

The study has shown that, the number of cholera cases and its risk factors in Tanzania mainland are spatially uneven, and the GWPR model was able to explain this variability. The GWPR model has shown to be a better fitting than PR model. This is indeed expected as the GWPR model estimates the local parameters of each unit, which can effectively explain spatial variability (Wu et al., 2013). This result is in line with previous studies (Rodrigues et al., 2014; Guo et al., 2017) that reported better fitting with GWR approach than a non-spatial regression model. Thus, regional variation can be considered vital when studying risk factors that significantly affect the likelihood of cholera occurrence, as the same variables may influence the number of cholera cases differently.

With respect to risk factors of cholera, accessing improved water source in the household was found to have inverse relationship with the number of cholera cases. Given the increasing number of households which use improved water sources from 54% in 2010 to 61% in 2015/2016, the number of households which access improved water sources in rural is still low 47.8% (MoHCDGEC et al., 2016). In 2000 the Government of Tanzania launched Tanzania Vision 2025; one of its aims is to provide high quality livelihood for its people ensuring access to safe water for all by 2025. By 2020 Tanzania government aimed to reach 75% of its population to have access to safe drinking water under Health Sector Strategic Plan IV. This line up with the Sustainable Development Goal 6 that aims in achieving access to safe and affordable drinking water for all by 2030 (WB, 2018). These could be an important measure in reducing cholera cases, which had also been identified by others as a long-term solution in controlling cholera (Nygren et al., 2014; Date et al., 2013). However, in some regions the case was different since increasing percentage of households which access improved water sources leads to an increase in the number of cholera cases. This could be attributed to water handling and other hygiene principles violation at home. Akyala et al. (2014) and Bhunia et al. (2009) stated that “improved water sources would still have risks of cholera if poorly treated”, or poor storage of clean water (Hedman, 2009).

Accessing improved sanitation facility in the household has shown a positive relationship with the number of cholera cases. This could be attributed to cultural aspects including norms where the father and daughter-in-law are not allowed to share the same toilet because it is believed it is a shame. This makes them to practice open defecation even if the toilet is found in the household (Massawe, 2017). Manyara, Kilimanjaro, Ruvuma, Lindi and Mtwara Regions; a negative association between accessing improved sanitation facility and number of cholera cases was observed. Similar result was also indicated by the multi-country study conducted by Leidnerand Adusumilli (2013). Ensuring sustainable improved sanitation facility at household could be an important measure in controlling cholera as has been among the strategies settled by cholera global roadmap to 2030. This is also supported by Tanzania National Health Policy 2007 which stated that, sustainability of sanitation is among the key mechanism to prevent diseases such as cholera in the country.

Handwashing behaviours showed a significant negative association with number of cholera cases in many regions. The reason is that hand washing reduces the ability of microorganisms to survive and hence, reduces transmission of the disease. This result is comparable with Dan-Nwafor et al. (2019) in Nigeria as well as Mahamud et al. (2012) in Kenya who identified handwashing behaviour as a preventive measure of cholera transmission. In Kilimanjaro, Manyara, Mwanza, Kigoma, Rukwa, Mbeya, Iringa, Njombe and Ruvuma Regions the case was different, handwashing behaviour leading to an increase in the number of cholera cases. This could be contributed by washing hands without soap or touching contaminated surfaces such as tap or toilet door knob after washing hands. Previous studies found that surfaces like taps and door knobs were the least cleaned areas and thus, they covered the highest bacterial load in a toilet (De Alwis et al., 2012; Carling et al., 2008). Also, this could be attributed to cultural beliefs of washing hands in the same bowl, believing that it is a sign of love and unity in society (Massawe, 2017).

Households practicing open defecation had shown negative association with the number of cholera cases; which differs with results of Cowman et al. (2017) who found that districts with a higher prevalence of open defecation were at a higher risk of cholera in Kenya. A positive correlation between cholera outbreak and practicing of open defecation by households in the Lake Zone was also found by Kessy and Mahali (2016). The difference in this observation on open defecation could probably be due to the fact that, the application of a single measure may not be a sufficient measure of controlling cholera in a household. A household needs to have an unshared toilet, improved sanitation facility, and fixed hand washing station with clean water and soap (Joint Monitoring Programme, 2017) to control cholera in the regions. However, eradicating open defecation is essential to ensure health environment and reducing infectious disease. This is underlined in 2013 the United Nations initiated a call to action on sanitation (UNICEF and WHO, 2015) and Sustainable Development Goal 6 (UN, 2018).

## Conclusions

Assessing risk factors of cholera across geographical areas is crucial especially in preventing and controlling cholera occurrence in Tanzania. By Application of GWPR model which was found to be the best model for predicting cholera occurrence compared to PR model, the study has explained the variation of the risk factors on cholera across the regions. It was found that the occurrence of cholera in the country is significantly associated with sources of water, sanitation facility, open defecation, hand washing, and shared toilets in the household. However, those risk factors were not the same throughout the country; they vary across regions. The study recommends that, since Governments and multilateral organizations such as the WHO and other agencies are engaged in eradicating cholera in Tanzania, effective policies and strategies be developed and implemented by them. Therefore, policies and strategies to control the cholera disease in Tanzania should be based on the variations of the risk factors found in the regions rather than focusing on interventions at the country level.

## References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. & Csaki, F. [Eds.], Proceedings of the 2nd international symposium on information theory, pp. 267–281.

Akyala, I. A., Shadrack, B. E., Ajumobi, O., Olayinka, A., & Nguku, P. (2014). Investigation of cholera outbreak in an urban north central Nigerian community- the Akwanga experience. *Public Health Research*, 4(1): 7–12.

Ali, M., Nelson, A. R., Lopez, A. L., & Sack, D. A. (2015). Updated global burden of cholera in endemic countries. *PLoS Neglected Tropical Diseases*, 9(6): 1–13.

Aydın, C., Gündüz, N., & Balibeyoğlu, J. (2021). Optimal bandwidth selection for a kernel density with a location-scale property. *Communications in Statistics: Theory and Methods*, 50(7): 1671-1684. doi: 10.1080/03610926.2020. 1804586

Bernasco, W., & Elffers, H. (2010). Statistical analysis of spatial crime data. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 699-724). New York: Springer.

Bhunia, R., Ramakrishnan, R., Hutin, Y., & Gupte, M. D. (2009). Cholera outbreak secondary to contaminated pipe water in an urban area, West Bengal, India, 2006. *Indian Journal of Gastroenterology*, 28(2): 62–64.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial non-stationarity. *Geographical Analysis,* 28: 281–298.

Burnham, D. R., & Anderson, K. P. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2): 261–304.

Bwire, G., Ali, M., Sack, D. A., Nakinsige, A., Naigaga, M., Debes, A. K., Ngwa, M. C., Brooks, W. A., & Orach, G. C. (2017). Identifying cholera "hotspots" in Uganda: An analysis of cholera surveillance data from 2011 to 2016. *PLoS Neglected Tropical Diseases*, 11(12): e0006118.

Camacho, A., Bouhenia, M., Alyusfi, R., & Luquero, F. J. (2018). Cholera epidemic in Yemen, 2016-18: An analysis of surveillance data. *Lancet Global Health*, 6(6): 680-690.

Carling, P. C., Parry, M. F., & Von-Beheren, S. M. (2008). Identifying opportunities to enhance environmental cleaning in 23 acute care hospitals. *Infection Control and Hospital Epidemiology*, 29(1): 1–7.

Carrel, M., Escamilla, V., Messina, J., Giebultowicz, S., Winston, J., Yunus, M., Streatfield, P. K., & Emch, M. (2011). Diarrheal disease risk in rural Bangladesh decreases as tubewell density increases: A zero-inflated and geographically weighted analysis. *International Journal* *of Health Geographics*, 10(41): 1–9.

Chatterjee, A., Dutta, P. K., & Chowdhury, R. (2007). Effect of fatty acids and cholesterol present in bile on expression of virulence factors and motility of vibrio cholera infection. *Journal of Immunology*, 75(4): 1946-1953.

Chen, S. (2015). Optimal bandwidth selection for kernel density functional estimation. *Journal of Probability and Statistics*, 10(3): 1-21.

Chen, J., Liu, L., Xiao, L., Xu, C., & Long, D. (2020). Integrative analysis of spatial heterogeneity and overdispersion of crime with a geographically weighted negative binomial model. *International Journal of Geo-Information*, 9(1): 60.

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: 110.

Dan-Nwafor, C. C., Ogbonna, U., Onyiah, P., Gidado, S., Adebobola, B., Nguku, P., & Nsubuga, P. (2019). A cholera outbreak in a rural North Central Nigerian community: An unmatched case-control study. *BMC Public Health*, 19(1): 1-7.

Date, K., Person, B., Nygren, B., Were, V., Kola, S., Ayers, T., & Quick, R. (2013). Evaluation of a rapid cholera response activity-Nyanza province, Kenya 2008. *The Journal of Infectious Diseases*, 208(1): 62-68.

De Alwis, W. R., Pakirisamy, P., Wai San, L., & Xiaofen, E. C. (2012). A study on hand contamination and hand washing practices among medical students. *International Scholarly Research Notices Public Health*, 251483: 1-5.

Eltoukhy, A., Shaban, I. A., Chan, F., & Abdel-Aal, M. (2020). Data analytics for predicting COVID-19 cases in top affected countries: Observations and recommendations. *International Journal of Environmental Research and Public Health*, 17(19): 7080.

Fotheringham, S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression*: *The analysis of spatially varying relationships*. Chichester: Wiley.

Guo, F., Su, Z., Wang, G., Sun, L., Tigabu, M., Yang, X., & Hu, H. (2017). Understanding fire drivers and relative impacts in different Chinese forest ecosystems. *Science of the Total Environment*, 605/606: 411–425.

Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T., & Calderwood, S. B. (2012). Cholera. *The Lancet*, 379(9835): 2466–2476.

Hedman, M. (2009). Women, water, and perceptions of risk : A case study made in Babati, Tanzania 2008. Bachelor Thesis, Södertörn University College. http://urn.kb.se/resolve?urn=urn:nbn:se:sh:diva-2666. Retrieved 14th November, 2020.

Hounmanou, Y. M. G., Mølbak, K., Kähler, J., Mdegela, R. H., Olsen, J. E., & Dalsgaard, A. (2019). Cholera hotspots and surveillance constraints contributing to recurrent epidemics in Tanzania. *BMC Research Notes*, 12(1): 1–6.

Joint Monitoring Programme (JMP). (2017). *WASH in the 2030 Agenda. Online brochure*. *http://www.who.int/watersanitationhealth/monitoring/coverage/jmp-2017-wash-in-the-2030-agenda.pdf?ua=1*. Retrieved 9th June, 2020.

Kazaji, N. D. (2016). Factors contributing to the prevalence of cholera during 2008 to 2009 in Yhembe District of Limpopo Province, South Africa. Masters Public Health Dissertation, University of Limpopo, South Africa.

Kessy, F., & Mahali, R. (2016). *Water, sanitation and hygiene services in Tanzania: Access, policy trends and financing*. https://www.thdr.or.tz/docs/THDR2017BP 11.pdf. Retrieved 13th December, 2020.

Leidner, A. J., & Adusumilli, N. C. (2013). Estimating effects of improved drinking water and sanitation on cholera. *Journal of Water and Health*, 11(4): 671–683.

Lessler, J., Moore, S. M., Luquero, F. J., McKay, H. S., Grais, R., Henkens, M., Mengel, M., Dunoyer, J., M'bangombe, M., Lee, E. C., Djingarey, M. H., Sudre, B., Bompangue, D., Fraser, R. S. M., Abubakar, A., Perea, W., Legros, D., & Azman, A. S. (2018). Mapping the burden of cholera in Sub-Saharan Africa and implications for control: An analysis of data across geographical scales. *The Lancet*, 391(10133): 1908–1915.

Li, Z., Lee, S. H., Lee, Y., & Valiou, E. (2011). Geographically-weighted regression models for improved predictability of urban intersection vehicle crashes: In transportation and development institute congress 2011. *American Society of Civil Engineers*, 41167(398): 1315–1329.

Mahamud, A., Ahmed, J., Nyoka, R., Auko, E., Kahi, V., Ndirangu, J., Nguhi, M., Burton, J. W., Muhindo, B. Z., Breiman, R. F., & Eidex, R. B. (2012). Epidemic cholera in Kakuma refugee camp, Kenya 2009: The importance of sanitation and soap. *Journal of Infection in Developing Countries*, 6(3): 234–241.

Massawe, E. G. (2017). The contribution of socio-cultural practices to the outbreak of cholera in Tanzania: A case study of Ukerewe District. Master Dissertation: The University of Dodoma. http://repository.udom.ac.tz/handle/20.500.12661/768?show=full. Retrieved 22nd July, 2021.

Meng, Q. (2014). Regression kriging versus geographically weighted regression for spatial interpolation. *International Journal of Advanced Remote Sensing and GIS*, 3(1):606-615.

Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of Chief Government Statistician (OCGS), & ICF. (2016). *Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16*. Dar es Salaam, Tanzania, and Rockville, Maryland, USA: MoHCDGEC, MoH, NBS, OCGS and ICF. Available at http://dhsprogram.com/pubs/pdf/FR321/FR321.pdf.

Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd Edn.). New York: McGraw-Hill.

Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, 24(17): 2695–2717.

Narra, R., Maeda, J. M., & Temba, H. (2017). Notes from the field: Ongoing cholera epidemic- Tanzania, 2015-2016. *Morbidity and Mortality Weekly Report*, 66: 177-178.

Nkeki, F. N., & Osirike, A. B. (2013). GIS-based local spatial statistical model of cholera occurrence: Using geographically weighted regression. *Journal of Geographic Information System*, 5(6): 531–

542.

Nygren, B. L., Blackstock, A. J., & Mintz, E. D. (2014). Cholera at the crossroads: the association between endemic cholera and national access to improved water sources and sanitation. *American Journal of Tropical Medicine and Hygiene*, 91(5): 1023–1028.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41: 673–690.

Oluwajana, S. D. (2018). Development of hot zone identification models for simultaneous crime and collision reduction. PhD Dissertation, York University Toronto, Ontario.

Osei, F. B., & Duker, A. A. (2008). Spatial dependency of V. cholera prevalence on open space refuse dumps in Kumasi, Ghana: A spatial statistical modelling. *International Journal of Health Geographics*, 7(1): 62.

Page, A. L., Ciglenecki, I., Jasmin, E. R., Desvignes, L., Grandesso, F., Polonsky, J., Nicholas, S., Alberti, K. P., Porten, K., & Luquero, F. J. (2015). Geographic distribution and mortality risk factors during the cholera outbreak in a rural region of Haiti, 2010-2011. *PLoS Neglected Tropical Diseases*, 9(3): e0003605.

Penrose, K., De Castro, M. C., Werema, J., & Ryan, E. T. (2010). Informal urban settlements and cholera risk in Dar es Salaam, Tanzania. *PLoS Neglected Tropical Diseases*, 4(3): 1–11.

Rajasingham, A., Hardy, C., Kamwaga, S., Sebunya, K., Massa, K., & Mulungu, J. (2019). Evaluation of an emergency bulk chlorination project targeting drinking water vendors in cholera-affected wards of Dar es Salaam and Morogoro, Tanzania. *American Journal of Tropical Medicine and Hygiene*, 100(6): 1335-1341.

Richterman, A., Azman, A. S., Constant, G., & Ivers, L. C. (2019). The inverse relationship between national food security and annual cholera incidence: A 30-country analysis. *BMJ Global Health*, 4(5): 1–8.

Rodrigues, M., de la Riva, J., & Fotheringham, S. (2014). Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Applied Geography*, 48: 52–63.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2): 461–464.

Su, Z., Hu, H., Tigabu, M., Wang, G., Zeng, A., & Guo, F. (2019). Geographically weighted negative binomial regression model predicts wildfire occurrence in the Great Xing'an mountains better than negative binomial model. *Forests*, 10(5): 377.

Tanzania, N.B.S. (2019). *Tanzania in figures 2018, National Bureau of Statistics*. Available online at www.nbs.go.tz. Retrieved 20th April, 2020.

Trærup, S. L. M., Ortiz, R. A., & Markandya, A. (2011). The costs of climate change: A study of cholera in Tanzania. *International Journal of Environmental Research and Public Health*, 8(12): 4386–4405.

United Nations (UN). (2018). *Goal 6: Ensure availability and sustainable management ofwater and sanitation for all*. From https://www.un.org/sustainabledevelopment/water-and-sanitation/. Accessed 17th January, 2021.

UNICEF. (2018). *Tanzania Humanitarian Situation Report*. From http://www.unicef.org/ethiopia/. Accessed 7th December, 2020.

UNICEF & WHO. (2015). *Progress on sanitation and drinking water: 2015 update and MDG assessment.* https://www.ircwash.org/sites/default/files/progressonsanitationanddrinkingwater.pdf. Accessed 15thJuly, 2020.

United Republic of Tanzania (URT). (2014). Water sector development programme phase II. In *Dar es Salaam: Ministry of water, United Republic of Tanzania.* http://www.tawasanet.or.tz/files/WSDPIIFinalAugust312014

World Bank (WB). (2018). Reaching for the SDGs: *The untapped potential of Tanzania's water supply, sanitation, and hygiene sector*: *WASH poverty diagnostic.* World Bank, Washington, DC.

World Health Organization (WHO). (2018). *Cholera Country Profile, United Republic of Tanzania*.http://www.who.int/csr/don/12-january-2018-cholera-tanzania/en/.Accessed 12th January, 2018.

Wu, W., & Zhang, L. (2013). Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in North Eastern Puerto Rico. *Applied Geography*, 37: 52–62.

Zhou, S., Suhong, Z., Lin, L., Zhang, M., Kang, M., Xiao, J., & Song, T. (2019). Examining the effect of the environment and commuting flow from/to epidemic areas on the spread of dengue fever. *International Journal of Environmental Research and Public Health,* 16(24): 5013.