International Journal of Health Economics and Policy
2021; 6(2): 56-64
http://www.sciencepublishinggroup.com/j/hep
doi: 10.11648/j.hep.20210602.14
ISSN: 2578-9295 (Print); ISSN: 2578-9309 (Online)

Spatial Variability of Typhoid Disease Incidences in Uganda Using Geographically Weighted Regression Approach

Kamukama Ismail[1, 2, *], Maiga Gilbert[1], Ssebuggwaawo Denis[2], Nabende Peter[1], Ali Mansourian[3]

[1]Department of Information Systems, Makerere University, Kampala, Uganda

[2]Department of Computer Science, Kyambogo University, Kampala, Uganda

[3]Department of Physical Geography, Lund University, Lund, Sweden

Email address:

Kamukama@cis.mak.ac.ug (K. Ismail), gmaiga@cit.ac.ug (M. Gibert), dssebuggwawo@kyu.ac.ug (S. Denis), pnabende@cis.mak.ac.ug (N. Peter), ali.mansourian@nateko.lu.se (A. Mansourian)

[*]Corresponding author

To cite this article:

Kamukama Ismail, Maiga Gilbert, Ssebuggwaawo Denis, Nabende Peter, Ali Mansourian. Spatial Variability of Typhoid Disease Incidences in Uganda Using Geographically Weighted Regression Approach. *International Journal of Health Economics and Policy.* Vol. 6, No. 2, 2021, pp. 56-64. doi: 10.11648/j.hep.20210602.14

Received: April 8, 2021; **Accepted**: April 21, 2021; **Published**: May 8, 2021

Abstract: The spatial variability of typhoid disease incidences has not been accounted for, most especially in developing countries, which makes its surveillance inefficient and expensive. This research aimed at (i) exploring possible risk factors of typhoid disease and (ii) accounting for spatial variability of typhoid disease incidences using GWR approach. The research first explored possible risk factors of typhoid disease using global regression-Ordinary Least Squares (OLS). Geographically Weighted Regression (GWR) model was used to account for spatial variability of typhoid disease incidences. Moran's Index was used to confirm spatial patterns in the data. The research revealed that; poor handwashing practice, rainfall and poor drainage (floods effect) were responsible for spatial variability of typhoid disease locally ($P<0.05$). GWR model revealed that poor handwashing practice was mainly responsible for typhoid disease occurrences in Northern, Northwestern and Mid-central parts of the country while excessive rainfall was mainly responsible for occurrence of the disease in the Eastern and Western regions. Poor drainage was mainly influencing disease occurrences in Eastern and Southwestern parts of the country. In the analysis, GWR model performed better than global OLS model (R-squared=0.37, R-squared=0.25 respectively). A combination of poor handwashing practice, excessive rainfall and poor drainage accounts for spatial variability of typhoid disease incidences in Uganda. This knowledge is very essential for surveillance teams to (i) enforce preventive measures, (ii) prepare for outbreaks and (iii) make targeted interventions to eventually reduce operational costs.

Keywords: Geographically Weighted Regression (GWR), Global Regression, Ordinary Least Squares (OLS), Typhoid, Uganda

1. Introduction

Typhoid disease has remained a global burden, influenced by various environmental and demographic factors most especially in developing countries of Africa and Asia. It is estimated that between 11 and 21 million cases of typhoid disease occur globally, causing between 128000 to 161000 death, most especially children and elderly [1]. In Uganda, almost all districts are typhoid endemic most especially in urban areas [2]. It is caused by "*salmonella typhi*" bacteria whose reservoir is humans. Its route of transmission is faecal contamination of water and food due to poor waste disposal. Although the disease is transmitted by faecal contamination of water and food, the routes leading to contamination differ from area to area [3].

Research studies have associated typhoid disease occurrences with many environmental and demographic factors. Mirembe *et al.* [4] associated the disease with poor sanitation and hygiene in the Ugandan district of Kasese. It was also found to be associated with poor handwashing practice in Fiji [5]. Another study in Kenya revealed that topography of areas was a risk factor, where people in valleys were more at risk than those on hills [6]. Another study by

Breiman *et al.* [7] associated the disease with urbanization in Kenya, where the disease was more pronounced in towns than in villages. Poor drainage and floods were also associated with typhoid disease in Fiji [8]. Rainfall and temperature were correlated with typhoid disease incidences in Bangladesh [9], and more others as presented in table 1. According to Crump [3], these risk factors differ at global, regional, national and sub-national levels. This shows that to get reliable results about the disease, models that cater for local spatial variabilities of such risk factors should be used [10, 11].

Many research studies at local level have used the conventional global regression approach-Ordinary Least Squares (OLS)- to explore epidemiological characteristics of typhoid disease. This is an approach that describes the whole study region with one equation [12]. If the phenomenon being studied is uniform across the whole area of study, the model gives reliable results. But if the phenomenon being studied varies across geographic space, the model gives biased and unreliable results [13]. Typhoid disease incidences and its risk factors vary across geographic space and there is no research that has accounted for its spatial variability locally. This called for application of Geographical Information Systems (GIS) to cater for the local variabilities of the disease to give improved results.

Geographically Weighted Regression (GWR) approach is one of the GIS approaches that is used to explore phenomena across geographic scope while catering for spatial variabilities. It is an approach where every spatial unit in the area of study has its own independent equation [12]. This unique and powerful feature enables GWR models to account for spatial variabilities of events such as disease incidences across the areas under study. However, according to Charlton & Fotheringham [14], precautions should be taken before fitting GWR models. Spatial dependence must first be confirmed in the data before using the model: otherwise, the model can give incorrect and misleading results. The model can also be affected by multi-collinearity amongst covariates and bandwidth [15].

This research aimed at (i) exploring possible risk factors of typhoid disease and (ii) using GWR approach to account for spatial variability of typhoid disease incidences in Uganda. This knowledge is very essential in planning and decision-making during surveillance, where authorities can enforce preventive measures, plan for outbreaks and make targeted interventions to reduce operational costs.

2. Materials and Methods

2.1. Study Area

Uganda is one of the countries of Africa with a current population of approximately 44.27 million people and 241,038 Square Kilometers of land (WPR Report, 2019). Figure 1 shows the map of Ugandan with 2012 district boundaries [16].

2.2. Data Sources

A total of clinically confirmed 1,263,923 typhoid disease cases for the period 2012 to 2017 were obtained from Uganda Ministry of Health (MOH). To protect patients' identities and avoid breach of patients' data confidentiality law, disease cases were collected already aggregated and organized per month per district. The Uganda shape files and population data were obtained from Uganda Bureau of Statistics (UBOS). Environmental data, that is, average rainfall, average maximum temperature and average minimum temperature was obtained from Uganda National Metrological Authority (UNMA). The data was collected by 18 weather stations, well identified for national representation. This was monthly data collected from 2012 to 2017. Basing on the available data, the variables that were investigated, their descriptions, literature references and respective data sources are summarized in table 1.

Data Pre-Processing

Clinical Data

To avoid population effect, total disease incidences per district were computed for the period 2012 to 2017 and then corrected to the population to give disease incidence rates per district using the formular:

$$Disease\ incidence\ rate\ (ATypy\_Rate(i)) = \frac{Incidence_i}{Population_i} \quad (1)$$

Where $ATypy\_Rate(i)$ is the disease incidence rate for district i, $Incidence_i$ is the average disease incidence of $district_i$ and $Population_i$ is the average population of $district_i$ for the period 2012 to 2017.

Figure 1. *Map of Uganda and district boundaries.*

Environmental data

Environmental data was collected by 18 stations well distributed nationally and Ordinary Kriging method of interpolation under ArcGIS platform was used to estimate weather values for each district. According to Ozturk and

Kilic [22], ordinary Kriging uses the formular

$$Z(S_0) = \sum_{i=1}^{N} \lambda_i Z(S_i) \quad (2)$$

Where $Z(S_i)$ is the measured value at the i^{th} location, λ_i is the unknown weight for the measured value at the i^{th}location, S_0 is the prediction location and *N* is the number of measured values. From formular 2, the average rainfall, average maximum temperature and average minimum temperature for each of the districts were determined.

Demographic data

This data was collected already organized per district and did not require pre-processing. This was data about annual district populations, hygiene, topography, urbanization and poor drainage (floods).

2.3. Methods

The general mapping, exploratory regression and Geographically Weighted Regression (GWR) were done using ArcGIS software (version 10.5, ESRI Inc.; Redlands, CA, USA). The overall research methodology has been organized under three categories; (i) Exploratory regression and Model selection, (ii) Global regression (OLS) and Geographically Weighted Regression.

Table 1. *Exploratory variables of the study.*

S/N	Variable	Description	Risk factor studied	Literature References	Data Source
1	HH_wash	Percentages of households that practice hand washing per district.	Hygiene	[5, 17, 18]	UBOS
2	PH_lands	Percentages of households that stay in highlands per district	Topography	[6]	UBOS
3	Urban_Level	Percentages urbanization level per district	Urbanization	[7, 19, 20]	UBOS
4	ARainfall	Average annual rainfall for period 2012 to 2017 per district	Rainfall	[8, 9, 11, 20]	UNMA
5	Temp_Max	Average annual maximum temperature for period 2012 to 2017 per district	Temperature	[9, 11]	UNMA
6	Temp_Min	Average annual minimum temperature for period 2012 to 2017 per district.	Temperature	[9, 11]	UNMA
7	Pn_Floods	Percentages of households affected by floods per district.	Poor drainage	[8, 20]	UBOS
8	Hhsize	Average household size per district	Household size	[21]	UBOS
9	ATypy_Rates	Typhoid Disease Incidence Rates per district (cases per unit population) from 2012 to 2017.	N/A (Dependent variable)	N/A (Dependent variable)	Computed

2.3.1. Exploratory Regression and Model Selection

Exploratory Regression (with all variables in table 1) and model selection were performed using ArcGIS with *ATypy_Rates* as a dependent variable while the rest of the variables as independent variables. The purpose of this task was to establish (i) variables that have significant influence on typhoid incidence rates, (ii) spatial influence in the data sets and (iii) best possible models. Under ArcGIS platform, multi-collinearity among explanatory variables is automatically detected. The possible models are suggested with their respective parameter values. Multi-collinearity affects performance of regression models and is measured by Variance Inflation Factor (VIF) parameter whose value must be less than 7.5 to avoid its effects. The parameter that was used for model selection was Corrected Akaike Information Criterion (AIC_c), and the results of exploratory regression and model selection are presented in table 2 and table 3 respectively.

2.3.2. Global Regression

Global regression is a statistical modelling method that is used to study the influence of one or many independent variables on a dependent variable [23]. If x_1, x_2, ……, x_n are *n* independent variables and y is the dependent variable in the study,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon \quad (3)$$

Where $\beta_i' s$ are the coefficients and ε is the random error term in the equation, $\epsilon \sim N(0, \partial^2)$.

This method uses a technique of Ordinary Least Squares (OLS) sometimes referred to as the Best Linear Unbiased Estimator (BLUE). According to Cheng and Jian [24],

$$\epsilon = \sum_{i=1}^{n} (y_i - Y_i), \quad (4)$$

Where Y_i are observed Y data points while y_i are the regressed Y data points and n is the number data points used on the regression line. This technique aims at minimizing errors in the dependent variable to get the best estimates. OLS techniques work with the following assumptions;

(i) The random errors have a mean of 0, (ii) the random errors have constant variance and are uncorrelated, and (iii) the random errors have normal distribution. If any of the above assumptions is not fulfilled, global OLS model becomes biased and the coefficients become inefficient. This modelling approach was used to validate the GWR model by comparing results of both models and results are presented in table 5.

2.3.3. Geographically Weighted Regression

According to Comba, *et al.* [25], the general standard regression model for spatial data is

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i \quad (5)$$

Where $i = 1, 2, \ldots, n$ are observation indices, y_i is the dependent variable, x_{ij} is the value of the j^{th} independent variable, m is the number of independent variables. β_0 is the intercept term, β_j is the regression coefficient for the j^{th}independent variable and ε_i is the random noise. This model treats all observations with equal weights and does not consider local variability in different spatial units [26]. According to Xiao, *et al. [27]*, GWR model therefore extends this model by adding geographical coordinates, and equation 5 becomes:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{m} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i \quad (6)$$

Where (u_i, v_i) are the geographical coordinates of the i^{th} observation, $\beta_j(u_i, v_i)$ is the realization of the continuous function $\beta_j(u, v)$ at point i.

According to Oshan and Fotheringham [13], the model uses weighted least estimation to compute a set of coefficients for all independent variables of every spatial unit. The matrix form of this set of coefficients is:

$$\hat{\beta}(i) = [X'W(i)X]^{-1}X'W(i)y_i \quad (7)$$

Where $\hat{\beta}(i)$ is a vector of coefficients, X is the matrix of independent variables, X' is the transpose of matrix X, $W(i) = diag[w_1(i), \dots\dots, w_n(i)]$ is a diagonal matrix of weights of each observation based on its distance from spatial unit i, and $y_i$is a vector of observations of dependent variable. The weights matrix of the observations is computed by a kernel function using equation 8.

$$W_j(i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, if\ j \epsilon \{N_i\} \\ 0, if\ j \notin \{N_i\} \end{cases} \quad (8)$$

Where d_{ij} is the distance between locations i and j, b is the distance between the N^{th} nearest neighbor and $\{N_i\}$ is a set of all observations that are no more than b distance from i.

From equation 8, an observation at location i would have a weight of 1 and an observation that is farther than distance d would have a weight of 0.

GWR models require optimal estimation of number of nearest neighbors, which is used to estimate parameter b in the kernel function before fitting the model. This is the bandwidth estimation. There are two types of kernels in the weighting, that is, fixed and adaptive. For a fixed kernel, distance is fixed but the number of neighbors varies where Gaussian function becomes secure and better. With adaptive kernel, distance varies but the number of neighbors remains constant, where the use of bisquare kernel function is recommended. Regardless of the type of kernel used, if regression points are evenly distributed, a bisquare function becomes better [27].

In the GWR models, Akaike Information Criterion (AICc) is used as a minimization routine to select optimal bandwidth for goodness-of-fit of the model. AICc of the model is computed by:

$$AICc = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{\frac{n + tr(s)}{n - 2 - tr(s)}\right\}, \quad (9)$$

Where n is the number of spatial units, $\hat{\sigma}$ is the standard deviation of the error term and $tr(s)$ is the trace of the hat matrix [13].

After specifying the bandwidth, the weights matrix is computed by the kernel function using equation (8), a GWR model can then be fitted to compute local coefficients using equation (7) and the average of local coefficients of determination R_i^2 gives the overall R^2 of the model.

3. Research Results

Results of this study are presented using three themes (i) Exploratory regression and model selection results (ii) Global regression Results and GWR Results.

3.1. Exploratory Regression and Model Selection Results

The exploratory OLS regression under ArcGIS platform with *ATypy_Rates* as a dependent variable while the rest of the variables in table 1 as independent variables yielded the results presented in table 2.

Table 2. *Results of Exploratory Regression.*

S/N	Variable	Coefficient	Moran's I	VIF	P-Value
1	Intercept	-598.7621	-	-	0.4613
2	HH_Wash	-89.4560	0.2107	1.30	0.0002
3	ARainfall	24.30	0.7092	1.12	0.0443
4	Pn_Floods	416.46	0.3666	1.68	0.0088
5	Temp_Max	208.72	0.8371	2.48	0.3570
6	Urban_level	79.24	0.1174	1.65	0.0180
7	Temp_Min	13.19	0.7432	4.09	0.9611
8	Hhsize	664.19	0.0395	1.16	0.0381
9	PH_lands	-198.96	0.6925	2.24	0.0462
10	ATypy_Rates	(dependent variable)		-	-

Table 3. *Possible models, parameter values and model variables using ArcGIS.*

AdjR2	AICc	JB	K (BP)	VIF	Possible Models
0.26	1793.75	0	0.21	1.48	-HH_Wash +ARainfall + Pn_Floods
0.23	1795.3	0	0.13	1.1	-HH_Wash +ARainfall-Temp_Min
0.24	1793.94	0	0.04	1.12	-HH_Wash +ARainfall-Temp_Max

From table 2, *HH_Wash* and *PH_lands* have negative correlation with typhoid disease incidence rates while the rest of the variables have positive correlation with the disease rates. Variables *HH_Wash, ARainfall, Pn_Floods, Urban_level, Hhsize* and *PH_lands* are significant (P<0.05) while the rest of the variables are insignificant (P> 0.05). All variables have a positive Moran's Index value which confirms spatial dependence in the datasets and this called for spatial approaches such as GWR to solve the problem. All values of VIF are far less than 7.5 threshold which implies that there is no any risk of multi-collinearity amongst covariates. All significant variables were subjected to model selection and the possible models are presented in table 3.

In table 3, *AdjR2* is the, *AICc* is Akaike's Information Criterion, *JB* is the *Jarque-Bera's* p-value, The *adjusted R-squared* statistic is used as a measure of model fit, the bigger its value, the better the model fit. AICc is used for comparing performances of different models. The model with lest value of AICc is always the best. The JB statistic is used to test for normality of residuals in the OLS model. Since the value of JB statistic is positive and significant ($P < 0.05$), it confirmed non-normality of the residuals in the model which violets assumptions of OLS models stated in section 2.3.2. This implies that the OLS model coefficients are biased and unreliable. Breusch-Pagan K (BP) test was significant ($P < 0.05$) implying that there was heteroscedasticity in the data due to spatial effects and which also called for the application of GWR modelling to cater for the local spatial effects for improved and reliable results.

From table 3, the best model for further spatial diagnosis was the one with the least *AICc* or highest *AdjR2* values with variables *HH_Wash, ARainfall* and *Pn_Floods*. The OLS equation is presented as equation 10.

$$ATypy\_Rate = \beta_0 - \beta_1 HH\_Wash + \beta_3 ARainfall + \beta_5 Pn\_Floods + \varepsilon, \quad (10)$$

where $\beta_1, \beta_2, \ldots, \beta_5$ are variable coefficients and ε is the error term.

3.2. Global Regression Results

Global regression (OLS) in equation 10 was fitted under ArcGIS platform and the results are presented in table 4.

Table 4. *Global regression Results.*

Variable	Coefficient	Std. Error	Z-Value	P-Value
Constant	1525.95	810.98	1.88	0.05
HH_Wash	-30.0003	7.05	-4.25	0.00
ARainfall	11.8547	5.02	2.35	0.02
Pn_Floods	40.1585	14.42	2.78	0.01
AICc=1996.0661				
R-squared=0.2491, Adjusted R2=0.2282				

From table 4, all the three covariates *HH_Wash, ARainfall* and *Pn_Floods* were significant ($P<0.05$) and the constant term was insignificant ($P>0.05$). HH_Wash was negatively correlated with disease incidence rates (*ATypy_Rates*). *ARainfall* and *Pn_Floods* were positively correlated.

3.3 Geographically Weighted Regression Results

From equation 8, the GWR model that was finally fitted under ArcGIS platform is:

$$ATypy\_Rate(i) = \beta_{0i}(u_i, v_i) + \beta_{1i}(u_i, v_i) HH\_Wash_i + \beta_{2i}(u_i, v_i) ARainfall_i + \beta_{3i}(u_i, v_i) Pn\_Floods_i + \varepsilon_i \quad (11)$$

Where $i = 1, 2, 3, \ldots, 112$, (number of districts), $\beta_{0i}(u_i, v_i)$ is constant for district with coordinates (u_i, v_i), β_{1i} is a coefficient of *HH_Wash* covariate for district i, β_{2i} is a coefficient of *ARainfall* covariate for district i and β_{3i} is a coefficient of *Pn_Floods* covariate for district i.

A fixed kernel was used due to its slightly higher model fit *(R2=0.37)* compared to adaptive kernel *(R2=0.35)* and AICc was used for optimal bandwidth selection. There was improved model fitting with GWR model compared to classic global OLS regression model and the results are presented in table 5.

Table 5. *Results of fitted GWR and Global Regression models.*

Parameter	Fitted GWR Model	OLS Model
AICc	1785.29	1996.07
R2	0.37	0.25
Adjusted R2	0.31	0.23

From table 5, it is evident that GWR model had a better model fit than the global regression (OLS) model due to its lower AICc value of 1785.3 compared to AICc of OLS model of 1996.1. This is again consistent with R-squared parameter. GWR model has much higher R-squared of 0.37 compared with 0.25 for global OLS model. The spatial variability of the coefficients of GWR model are presented in figures 2, 3 and 4.

From figure 2, it is evident that poor handwashing practice is mainly responsible for typhoid disease occurrences in the Northern and Central parts of the country. This influence of the disease can also be observed in Southeastern parts of the country.

Figure 3 shows that excessive rainfall is responsible for most typhoid disease occurrences in the Eastern and Southwestern parts of the country. Lower influences can also be seen in mid-central, Southern and Northeastern parts of the country.

Figure 4 shows that poor drainage (floods effect) is mainly responsible for typhoid disease occurrences in both Eastern and Southwestern parts of the country. Significant influences can also be seen in the Central parts of the country. This factor influences the disease least in Northen parts of the country. Figure 5 continues to present spatial variability of predicted disease incidence rates by GWR model.

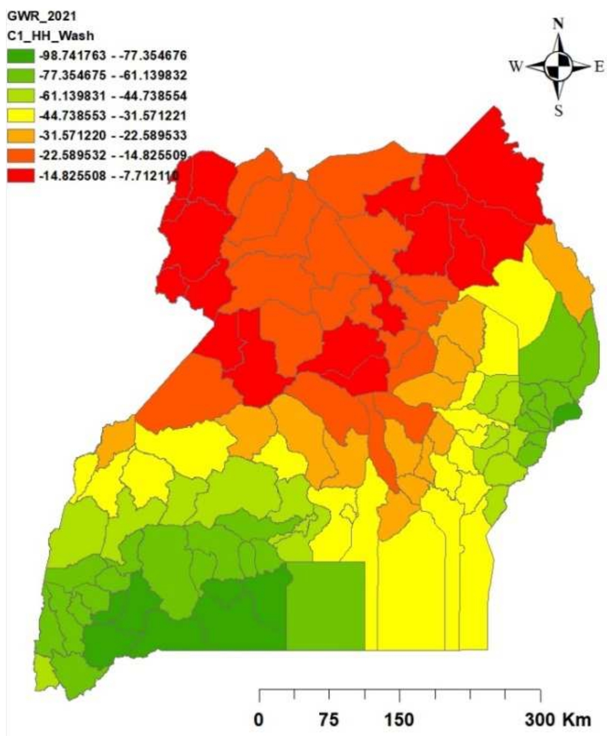

Figure 2. *Spatial variability of coefficient of HH_Wash.*

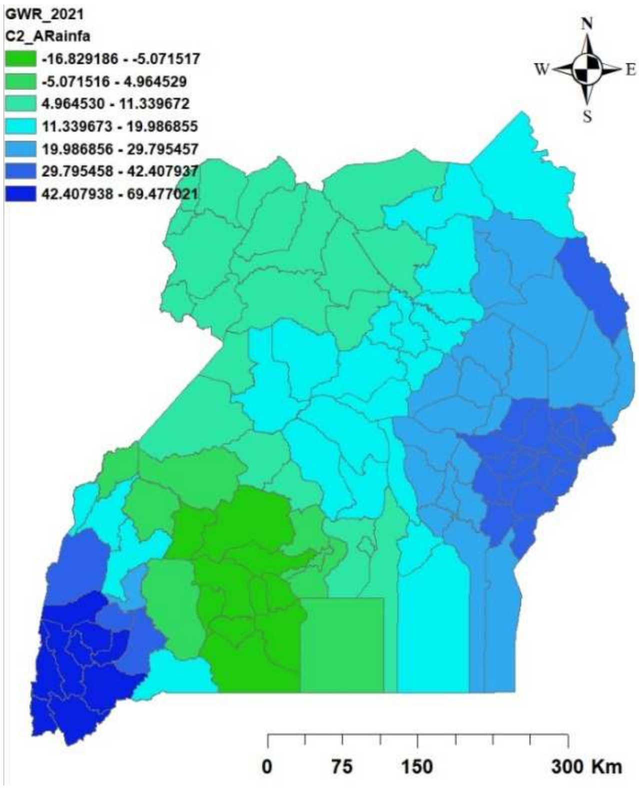

Figure 3. *Spatial variability of coefficient of ARainfall.*

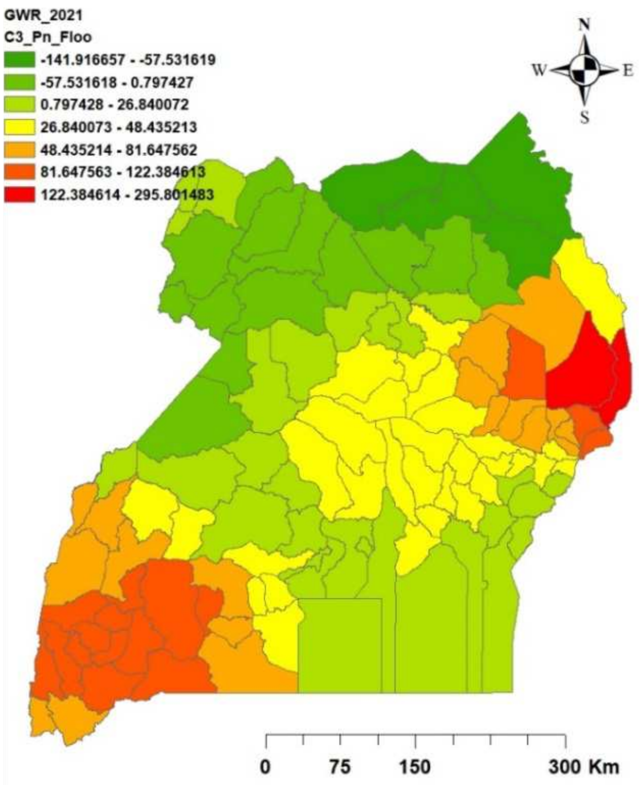

Figure 4. *Spatial variability of coefficient of Pn_Floods.*

From figure 5, the model predicted high incidences in Southern, South-Western and Eastern parts of the country. The Northern part of the country is predicted to have least incidences of the disease.

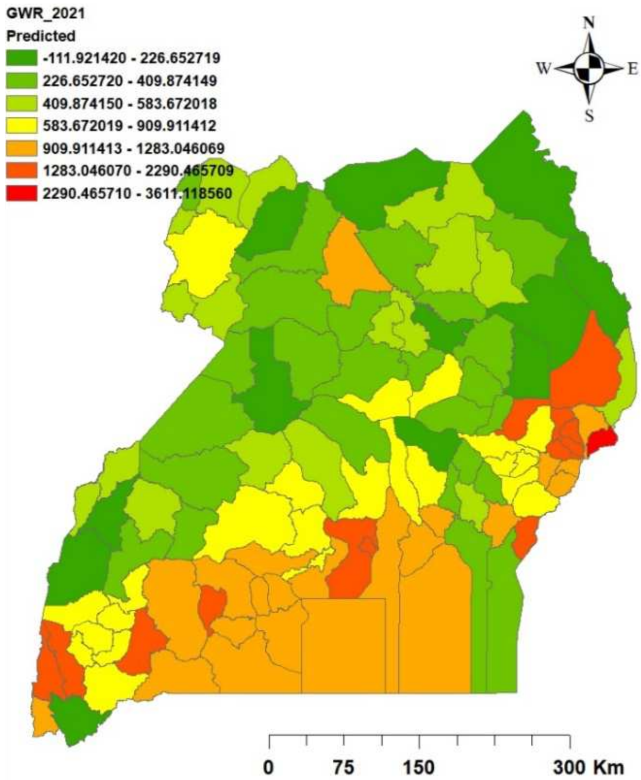

Figure 5. *Spatial variability of predicted disease incidence rates by GWR model.*

The GWR model was evaluated by using out-of-sample data technique. The model was fitted with data from 2012 to 2017. The spatial patterns of the actual clinical data of 2018 (out-of-sample data) were compared with spatial patterns of the predicted values of GWR model using Moran's Index, the results are presented in table 6.

Table 6. *Spatial patterns for predicted values of GWR model and those of actual clinical data of 2018.*

Datasets	Moran's I	Variance	P – value
Predicted values	0.117	0.0290	0.030
Actual clinical data of 2018	0.076	0.0023	0.047

Table 6 shows that the predicted values and actual values of 2018 have positive and significant autocorrelation values (Moran's I=0.117, P=0.03) and (Moran's I=0.076, P=0.047) respectively. These values are almost the same which is a characteristic of a good model. The spatial patterns are again presented graphically in figure 6.

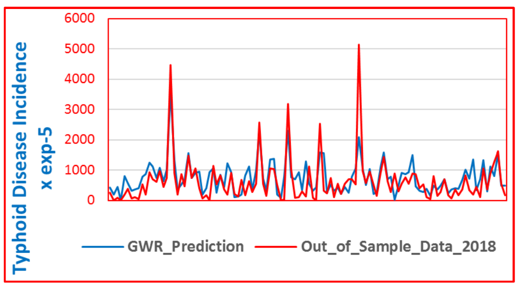

Figure 6. *Spatial patterns of out of sample data and those of predicted incidences of GWR model.*

Figure 6 shows that the patterns are close, implying that GWR model is capable of identifying both high and low-risk areas which is the major property of good models. The results are discussed in the next section.

4. Discussion

This research aimed at (i) exploring possible risk factors of typhoid disease and (ii) using GWR approach to account for spatial variability of typhoid disease incidences in Uganda.

The study revealed that typhoid disease occurrence is influenced by many demographic and environmental risk factors. There was a negative and significant relationship between typhoid disease incidences and handwashing ($P<0.05$). This implies that handwashing practice significantly reduces the risk of typhoid infection. Similarly, the populations that stay in highlands were at a lower risk of getting typhoid disease than those that stayed in valleys. Rainfall, Floods, Urbanization level and household size were revealed to have positive and significant influence on the occurrence of the disease ($P<0.05$). This implies that districts that receive high rainfall and affected by floods were at high risks of getting the infection. Also districts with high proportions of populations in towns were at a higher risk of getting infected by the disease. Average household size yielded a positive significant relationship with typhoid disease ($P<0.05$). This is because household size directly affects hygiene and sanitation due to increased costs of disposal and living requirements. These increased costs are usually unaffordable to many households, consequently increasing disease risk. The research did not find significant relationship between typhoid disease incidences with temperature (both average maximum and minimum) ($P>0.05$).

GWR model results revealed that the disease occurrence in different areas is influenced by different factors. Figure 2 shows that most of typhoid disease occurrences in the Northern, Mid-Central and South-Eastern parts of the country was as a result of poor handwashing practice in those areas. This fact was gain evidenced by Hirai *et al.* [28] that poor handwashing practice in these areas opens the route of disease transmission to eventually penetrate the population. In addition, it is reported that these areas have higher proportions of drinking surface water without treatment and open defecation, which are additional routes of disease transmission. A combination of these factors provides conducive environments for easy spread of harmful bacteria including *salmonella typhi (bacteria that causes typhoid)* [28].

From figure 3, the model revealed that excessive rainfall was highly influencing typhoid disease occurrences in Eastern and Southwestern regions of the country. From figure 7, these are mountainous areas. The Eastern region includes Mount Elgon and the Southwestern region includes Mount Rwenzori and Mount Mufumbiro [29].

In figure 4, the model associates much of the typhoid incidences in the Eastern, Western and Central parts of the country to the effect of floods. When it rains, floods mainly affect high population densities in valleys and this was revealed by other studies [30]. From figure 7, Mount Elgon is surrounded by districts Bududa, Bulambuli, Mbale, Sironko, Manafwa, Bukwo, Kween and Kapchorwa. These districts are affected by floods and mudslides every year, which has resulted into temporally camps and permanent relocations. Poor conditions in these camps every year result into waterborne diseases including typhoid [31].

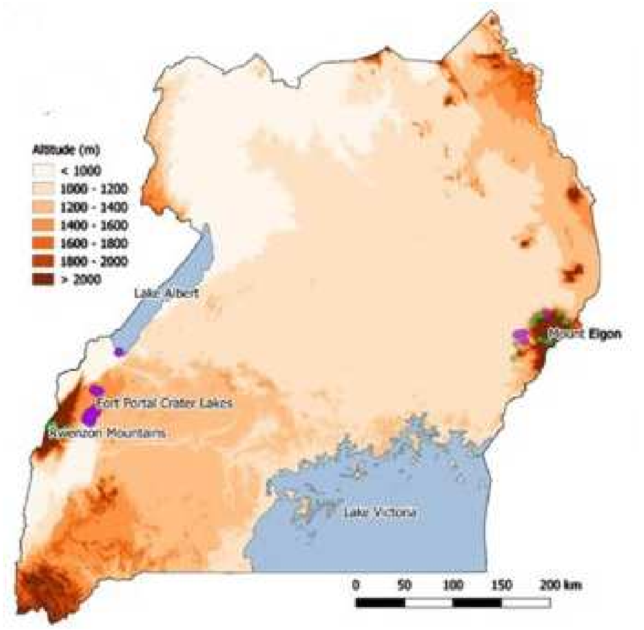

Figure 7. *A map of Uganda altitudes (Source: [29]).*

Figure 4 also indicate high effect of floods in Southwestern region. These are mountainous areas characterized by cultivation in the hillsides and human settlements largely in valleys. The highest densities in these valleys are severely affected by floods eventually contaminating water sources. The Southern and Central parts of the country have high rate of urbanization (including capital city Kampala), where many settlements have taken place without proper planning. People have settled in wetlands, cut vegetation on hilltops and hillsides, blocked waterways and many households are affected by floods. These floods destroy, fill-up latrines and sewage reservoirs eventually contaminating water sources. This fact was reported by other researchers [32]. Many research studies have also associated floods with typhoid disease and other waterborne diseases [4, 33, 34].

From table 5, GWR model performed better than OLS regression model (R^2=0.37, R^2=0.25 respectively). This is because GWR model catered for local spatial variabilities of disease incidences and other covariates, which OLS model could not do. In Figure 5, the model predicted high incidences of the disease in Southern parts of the country. This came from the combination of the three risk factors: poor handwashing practice, high rainfall and the effect of excessive floods in highly and ill-planned urbanized areas.

Limitations of the Study

(i) Some data could have been missed out as a result of some hospitals, health centers and clinics' failure to submit their data to the national MOH database, which could have direct impact on the results.

(ii) During treatment, some few district cross-border referrals could have occurred with chances of duplicating records in the system, and this could have some effects on the results.

(iii) Some patients resort to self-medication without reporting to any health facilities, and such cases could have been missed out, which could have some effect on the results.

5. Conclusions

The main goal of the research was to (i) explore risk factors of typhoid disease and (ii) use GWR to account for spatial variability of typhoid disease incidences in Uganda. Global regression revealed that typhoid disease is influenced by many spatially varying environmental and demographic factors. GWR model revealed that poor handwashing practice influences typhoid disease mainly in Northwestern, Northern and Northeastern parts of the country. High rainfall was most responsible for disease incidences in the Eastern, Central and Southern parts of the country. Floods were mainly influencing disease in Western, Central and Southern parts of the country. A combination of poor hygiene, excessive rainfall and floods accounts for spatial variability of typhoid disease incidences in Uganda. In the study, GWR model performed better than OLS regression model.

The research revealed that typhoid disease can be locally controlled through (i) promoting handwashing practice (ii) enforcing proper disposal to avoid contamination by excessive rainfall and (iii) well planned cultivation and settlements in both mountainous and urban areas to avoid blocking waterways, which is a major cause of poor drainage for water source contamination. This knowledge is very essential for concerned authorities to change most of the focus from treatment to prevention. It also helps authorities to effectively plan and effect targeted interventions, which eventually reduces costs of surveillance. The use of GIS explores spatially varying epidemiological characteristics of diseases to support targeted interventions and preventions. This eventually reduces surveillance costs and it should be the future direction most especially for the developing countries.

Acknowledgements

The research is sponsored by Makerere-SIDA Phase (IV) Programme and supported by Kyambogo University. Uganda National Meteorological Authority provided environmental data, Ministry of Health provided typhoid disease clinical data and Uganda Bureau of Statistics (UBOS) provided demographic data and Uganda shape files.

References

[1] WHO. (2019). Immunization, Vaccines and Biologicals, Typhoid, https://www.who.int/immunization/diseases/typhoid/en/. (Accessed on 22/08/2020).

[2] Kabwama, S., N., Bulage, L., Nsubuga, F. *et al.* (2017). A large and persistent outbreak of typhoid fever caused by consuming contaminated water and street-vended beverages: Kampala, Uganda, January – June 2015. *BMC Public Health* 17, 23 (2017) doi: 10.1186/s12889-016-4002-0.

[3] Crump, J., A., (2017). Progress in Typhoid Fever Epidemiology, Supplementary articles, Clinical Infectious Diseases, 2019; 68 (S1): S4–9, Centre for International Health, University of Otago, Dunedin, New Zealand.

[4] Mirembe, B., B., Mazeri, S., Callaby, R., Nyakarahuka, L., Kankya, C., Muwonge, A. (2019). Temporal, spatial and household dynamics of Typhoid fever in Kasese district, Uganda. PLoS ONE 14 (4): e0214650.https://doi.org/10.1371/journal.pone.0214650.

[5] Greenwell, J., et al. (2013). Typhoid fever: hurdles to adequate hand washing for disease prevention among the population of a peri-urban informal settlement in Fiji. Western Pacific Surveillance and Response Journal, 2013, 4 (1): 41–45. doi: 10.5365/wpsar.2012.3.4.006.

[6] Akullian, A., Ng'eno, E., Matheson, AI, Cosmas, L., Macharia, D., Fields, B., et al. (2015). Environmental Transmission of Typhoid Fever in an Urban Slum. PLoS Negl Trop Dis 9 (12): e0004212. https://doi.org/10.1371/journal.pntd.0004212.

[7] Breiman, R., F., Cosmas, L., Njuguna, H., Audi, A., Olack, B., Ochieng, J., B.,… Feikin, D., R. (2012). Population-based incidence of typhoid fever in an urban informal settlement and a rural area in Kenya: implications for typhoid vaccine use in Africa. *PloS one*, *7* (1), e29119. doi: 10.1371/journal.pone.0029119.

[8] Jenkins, A., P., Jupiter, S., D., Jenney, A., Rosa, V., Naucukidi, A., Prasad, N., Vosaki, G., Mulholland, K., Strugnell, R., Kama, M., Crump, J., A., Horwitz, P. (2019). Environmental Foundations of Typhoid Fever in the Fijian Residential Setting, International Journal of Environmental Research and Public Health.

[9] Dewan, A., M., Corner, R., Hashizume, M., Ongee, E., T. (2013). Typhoid Fever and its association with environmental factors in the Dhaka Metropolitan Area of Bangladesh: a spatial and time-series approach. *PLoS neglected tropical diseases*, *7* (1), e1998. doi: 10.1371/journal.pntd.0001998.

[10] Kim, O., S., Nugent, J., B., Y, Newell, J., P., Curtis, A., J. (2017). A Mixed Application of Geographically Weighted Regression and Unsupervised Classification for Analyzing Latex Yield Variability in Yunnan, China. *Forests*, *8* (5), 162. https://doi.org/10.3390/f8050162.

[11] Saad, N., J., Lynch, V., D., Antillón, M. *et al.* (2018). Seasonal dynamics of typhoid and paratyphoid fever. *Sci Rep* 8, 6870 (2018), doi: 10.1038/s41598-018-25234-w.

[12] Kala et al. (2017). A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. PeerJ 5: e3070; DOI 10.7717/peerj.3070.

[13] Oshan, T., M., Fotheringham, A., S. (2018). A Comparison of Spatially Varying Regression Coefficient Estimates Using Geographically Weighted and Spatial-Filter-Based Techniques, Geographical Analysis (2018) 50, 53–75.

[14] Charlton, M., Fotheringham, A., S., (2019). Geographically Weighted Regression, A Tutorial on using GWR in ArcGIS 9.3, National Centre for Geocomputation, National University of Ireland Maynooth. https://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr_arcgis/GWR_Tutorial.pdf.

[15] Anselin, L. (2005). Exploring Spatial Data with GeoDa: Workbook.

[16] UBOS, 2012. Statistical Abstract 2012, Uganda Bureau of Statistics, http://www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/2012StatisticalAbstract.pdf.

[17] Nahimana, M., R., Candide, T., N., Olushayo, O., Nyamusore, J., Isiaka, A., Ndahindwa, V., Dassanayake, L., Rusanganwa, A. (2017). Knowledge, attitude and practice of hygiene and sanitation in a Burundian refugee camp: implications for control of a Salmonella typhi outbreak, Pan African Medical Journal, http://www.panafrican-med-journal.com/content/article/28/54/full.

[18] Brainard, J., D'hondt, R., Ali, E., Van denBergh, R., De Weggheleire, A., Baudot, Y., et al. (2018). Typhoid fever outbreak in the Democratic Republic of Congo: Case control and ecological study. PLoSNegl Trop Dis 12 (10): e0006795. https://doi.org/10.1371/journal.pntd.0006795.

[19] Reyes, R., Ahn, R., Thurber, K., Burke, T., F. (2013). Urbanization and Infectious Diseases: General Principles, Historical Perspectives, and Contemporary Challenges. In: Fong I. (eds) Challenges in Infectious Diseases. Emerging Infectious Diseases of the 21st Century. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-4496-1_4.

[20] Mourad, K., A., Habomugisha, V., Sule, B., F. (2019). Assessing Students' Knowledge on WASH-Related Diseases, *Int. J. Environ. Res. Public Health* 2019, *16* (11), 2052; https://doi.org/10.3390/ijerph16112052.

[21] Sur, D., Ali, M., von Seidlein, L., *et al.* (2007). Comparisons of predictors for typhoid and paratyphoid fever in Kolkata, India. *BMC Public Health* **7,** 289 (2007). https://doi.org/10.1186/1471-2458-7-289.

[22] Ozturk, D., & kilic, F. (2016). Geostatistical Approach for Spatial Interpolation of Meteorological, Anais da Academia Brasileira de Ciências (2016) 88 (4): 2121-2136.

[23] Shyti, B., Valera, D. (2018). The Regression Model for the Statistical Analysis of Albanian Economy, Department of Mathematics, FNS, University of Elbasan Albania, International Journal of Mathematics Trends and Technology (IJMTT) – Volume 62 Number 2 – October2018.

[24] Cheng, W., Jian, Z., Y., (2018). Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting, Atmospheric Measurement Techniques, EGU. https://doi.org/10.5194/amt-11-1233-2018.

[25] Comber, A., Wang, Y., Lü, Y., Zhang, X., Harris, P. (2018). Hyper-local geographically weighted regression: extending GWR through local model selection and local bandwidth optimization. Journal of Spatial Information Science, Number 17 (2018), pp. 63–84, doi: 10.5311/JOSIS.2018.17.422.

[26] Goschin, Z. (2018). "Regional patterns of Romanian emigration. A Geographically Weighted Regression Model," Romanian Journal of Economics, Institute of National Economy, vol. 46 (1 (55)), pages 60-74, June.

[27] Xiao, D., Xu, X., Duan, L. (2019). Spatial-Temporal Analysis of Injury Severity with Geographically Weighted Panel Logistic Regression Model, Hindawi Journal of Advanced Transportation Volume 2019, Article ID 8521649, 15 pages, https://doi.org/10.1155/2019/8521649.

[28] Hirai, M., Roess, A., Huang, C., Graham, J. (2016). Exploring geographic distributions of high-risk water, sanitation, and hygiene practices and their association with child diarrhea in Uganda. Glob Health Action, 2016; 9: 10.3402/gha.v9.32833, doi: 10.3402/gha.v9.32833.

[29] Stanton, M., C., Adriko, M., Arinaitwe, M., Howel, A., Davies, J., Allison, G., E., LaCourse, J., Muheki, E., Kabatereine, N., B., Stothard, J., L. (2017). Intestinal schistosomiasis in Uganda at high altitude (>1400m): Malacological and epidemiological surveys on Mount Elgon and in Fort Portal crater lakes reveal extra preventive chemotherapy needs, Infectious Diseases of Poverty (2017) 6: 34 DOI 10.1186/s40249-017-0248-8.

[30] Akankwasa, B. (2018). Uganda Red Cross supporting communities affected by cholera, floods and landslides in Western Uganda*, https://www.unicef.org/uganda/stories/uganda-red-cross-supports-communities-affected-cholera-floods-and-landslides-western-uganda.* (Accessed 25/03/2020).

[31] Atuyambe, L., M., Ediau, M., Orach, C., G., Musenero, M., Bazeyo, W. (2011). Land slide disaster in eastern Uganda: rapid assessment of water, sanitation and hygiene situation in Bulucheke camp, Bududa district, *Environmental health: a global access science source*, *10*, 38. doi: 10.1186/1476-069X-10-38.

[32] Garcia, J., & Markandya, A. (2015). Economic Assessment of the Impacts of Climate Change in Uganda, Kampala Case-Study, Ministry of Water and Environment, https://cdkn.org/wp-content/uploads/2015/12/Uganda_CC-Economics_Kampala_Case-study2.pdf.

[33] NEMA. (2016). State of the Environment Report for Uganda 2014. National Environment Management Authority (NEMA), Kampala.

[34] Liu, Z., Lao, J., Zhang, Y., Liu, Y., Zhang, J., Wang, H., Jiang, B. (2018). Association between floods and typhoid fever in Yongzhou, China: Effects and vulnerable groups, Environmental Research Volume 167, November 2018, Pages 718-724, https://doi.org/10.1016/j.envres.2018.08.030.