International Journal of Data Science and Analysis
2024; 10(2): 33-40
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20241002.12
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Multiple Linear Regression Analysis of Factors Contributing to the Spread of Cholera in Kenya from 2000 to 2022

Samuel Okinyi, Joseph Njuguna Karomo*, Nzioka Dorcas Mbinya, Mohammed Kassim Thoya

Department of Pure and Applied Sciences, Kirinyaga University, Nairobi, Kenya

Email address:

okinyisamuel4@gmail.com (Samuel Okinyi), josekaromo@gmail.com (Joseph Njuguna Karomo),
mbinyadorcas33@gmail.com (Nzioka Dorcas Mbinya), mohammadthoya@gmail.com (Mohammed Kassim Thoya)
*Corresponding author

To cite this article:

Samuel Okinyi, Joseph Njuguna Karomo, Nzioka Dorcas Mbinya, Mohammed Kassim Thoya. (2024). Multiple Linear Regression Analysis of Factors Contributing to the Spread of Cholera in Kenya from 2000 to 2022. *International Journal of Data Science and Analysis, 10*(2), 33-40. https://doi.org/10.11648/j.ijdsa.20241002.12

Received: 14 May 2024; **Accepted:** 6 June 2024; **Published:** 19 June 2024

Abstract: Cholera is a deadly disease caused by consumption of either food or water that is contaminated with bacterium known as *Vibro cholerae*. In the 19^{th} century, Cholera had spread across the globe from its original source in Ganges Delta, India. This research aimed to investigated the factors contributing to the spread of Cholera in Kenya from 2000 to 2022. The key objective was to fit a multiple linear regression model which aid in determining the goodness of fit as well as examining the relationship between different types of drinking water and spread of Cholera. Secondary data was obtained from United Nations International Children's Emergency Fund (UNICEF) and World Health Organization (WHO). The results showed a strong positive linear relationship between Surface water and proportion of population affected by Cholera. 88% total variation in the proportion of population being affected by Cholera can be explained by the Unimproved Basic water (UBW), Surface water (SW) and Least Basic drinking water (LBW). Significantly this research encouraged the use of the least basic drinking water which clearly showed an inverse proportion to the proportion of population affected by Cholera. In addition to that, the researchers recommend improved access to clean water and create awareness on the dangers of using untreated water.

Keywords: Cholera, Multiple Linear Regression, Coefficient of Determination, Correlation Coefficient

1. Introduction

Cholera is a deadly disease caused by consumption of either food or water that is contaminated with bacterium known as *Vibro cholerae* . There are several serogroups of *Vibro cholerae* [1] that exist but only two namely 01 and 0139 causes Cholera outbreak. The *Vibro cholerae* 01 has caused most of the recent epidemics of Cholera. The *Vibro cholerae* 0139 was first spotted in Bangladesh in 1992 and caused several outbreaks in the past, but recently it has been identified in irregular cases.

In the 19^{th} century, Cholera had spread across the globe from its original source in Ganges delta, India[5].The seventh epidemic commenced in South Asia around 1961 and extended to Africa in 1971 then spread out to America in 1991 [17]. Cholera transmission is jointly linked to the use of unclean water. The areas that are at high risk to be affected by cholera are the slums and refugee camps where there is minimal requirement of clean treated water and sanitation are not upheld.

World Health Organization(WHO) has continuously reported a higher number of cholera cases in the last few years. In the year 2020, 323,369 cases of cholera outbreak had been reported and 857 deaths were announced from 24 countries across the globe.

Kenya has experienced various cholera waves since 1971 [10]. The largest wave of cholera was experienced between 1997 to 1999 which caused 26,901 cases and 1,362 deaths. In addition to that, the next wave occurred between 2007 to 2009 where 16616 cases and 454 deaths were reported. The result showed that 42 percent of the reported cases occurred in children aged below 15 years.

The most affected counties were Nairobi, Migori, Homabay, Kisumu, Turkana, Wajir and Mombasa. These areas experienced different environmental, demographic and climatic factors. Nairobi has been greatly affected due to it's dense population; Turkana and Wajir has been affected since large refugee camps are located in the counties; Migori, Homabay and Kisumu have been affected due to floods that occur during the long rainy seasons. Strategies have been adopted to minimize the spread of cholera through use of Integrated Disease Surveillance and Response(IDSR). Measures have been put in place by Ministry of Health (MOH) which have led to the decline in cholera cases recently.

Based on study conducted by [13], results showed that two sub counties within western region of Kenya experienced an outbreak of severe watery diarrhea which occurred on February 2015. The *Vibro cholerae* which had caused most of the recent outbreaks was isolated from 26 cases and from water sample collected from the river (surface water) used by the population within the two sub-counties.

Regression analysis is a statistical model that shows the correlation between explanatory variables and the response variable [9]. The main objective of regression analysis is to determine the effects of independent variables on the response variable. There exists both simple linear regression model and multiple linear regression model. Each is selected based on the kind of explanatory variables that one is dealing with. When dealing with one explanatory variable, simple linear model is used and when the explanatory variables are more than one then a multiple linear regression model is used. There exist various diagnostic test performed on statistical regression. These: linearity, normality, homoscedasticity, randomness of the stochastic term and non multicollinearity. homoscedasticity is the property of consistency of the error term. Multicollinearity occurs when there exist a relationship between the explanatory variables. These assumptions must be justified in order to perform any regression analysis.

2. Literature Review

A study with the aim to investigate how surface water systems increase the risk of cholera infection was done by [11]. A review was conducted based on 54 previous articles published between 1995 to 2019 August. The study found definite relationship between surface water interaction and waterborne diseases including Cholera. However, weak relationships were found between agricultural land use changes and cholera transmission.

The refugee camps in Kenya are vulnerable of being affected by Cholera due to poor sanitation, nutrition by refugees and also due to dense population which leads to over crowding. A study done by [16] concerning risk factors and importance of sanitation on Kakuma refugee camp in Kenya. The camp uses water distributed from boreholes (surface water). Although the water is chlorinated in steel reservoirs before being distributed, the study showed that the distribution system and water sanitation was greatly increasing the of cholera transmission.

Based on a report written by [15], in the 20^{th} century, the Sub-Saharan Africa has recorded the maximam outbreak of Cholera ever within high percentage of mortality rate due to Cholera in the world. In June 1997 to March 1998 an outbreak of Cholera was recorded in the western Kenya region around the Lake Victoria, $14,275$ patients were admitted to various hospitals in Nyanza region alone in which 547 died marking a case fatality rate of 4. A case-controlled study was conducted on 61 Cholera patients with different age, sex and clinical matched controls. From a multivariate analysis of the study, a conclusion was made that among the risk factors for the increase in transmission was drinking unpurified water from Lake Victoria or from streams, eating the same food with people having watery diarrhea especially in overcrowded areas like feasts from funerals, weddings and parties.

Cholera being one of the waterborne and environmental diseases in which its study and research is scarce in the third world countries like Kenya [14]. In most of the outbreaks recorded in Africa from 2009 to 2011 three quarters of the outbreaks occurred in inland regions, this provide a suggestion that these outbreaks are mostly influenced by rainfall or lack of enough purified sources of water. The *Vibro cholerae* had been taking place in areas around reservoirs, lakes and rivers.

A study was conducted to determine the association between Cholera and different demographic and environmental factors related to water and other factors [4]. The obtained data was analyzed using the zero inflated binomial regression model. The outcome of multivariate analysis showed that the danger of being affected by Cholera was linked to use of unimproved water sources. The finding showed that there was no statistical significance between poverty and number of healthcare facilities within $100,000$ persons. The gap identified is that the study failed to check on specifically the relationship between Cholera and unimproved water. The improved access to clean treated drinking water in remote areas may have greatly led to decrease in the cases of Cholera [12]. The study utilized correlation and regression analysis which helped in identifying main factors which facilitated the spread of Cholera by taking a keen look into the factors that hinders access to water testing. The high presence of springs in various locations had great significant positive correlation with the high spread of cholera infection rate in Northern Corridor.

There had been a high need for adequate availability of drinking water [8]. An estimation was carried out concerning the outbreak of Cholera using a pooled data on Asia and Africa between 1990 and 2008. The case rates that existed were mortality rates linked to Cholera and other two disease control mechanism, drinking water and sanitation services. The results showed that a statistical significance and negative effects were encountered between drinking water services and Cholera. In addition to that a weak statistical association was observed between Cholera outbreak and sanitation services.

A study was conducted on impact of water and sanitation services on Cholera [3]. The study investigated the correlation between Cholera rates per population and lack of basic services of drinking water and sanitation in the sub-Saharan African countries, where incident cases of Cholera have been regularly reported to the WHO since 1991. They study used the Spear-man's correlation method to analyze their data and the results showed a convincing correlation between the lack of basic water provision and Cholera rates.

A study on how climate change affect the spread of Cholera was conducted by [6]. Several studies all of which mentioned more than one risk factor of spread of Cholera. Some studies reported several sanity-related risks which were the most reported risk factor, accounting for over 30% of the total Cholera risks. The study also indicated factors like poor access and availability of clean water, people not washing hands before eating, poor latrine access, drinking unchlorinated water, not storing water in a sealed container and drinking unboiled water also increased the contamination and spread of Cholera.

3. Methodology

3.1. Descriptive Statistics

The study made use of 92 observations, three explanatory variables and one response variable for analysis. The 3 predictor variables were: Least Basic Drinking water(X_1), Unimproved Drinking Water(X_2) and Surface Water(X_3). The response variable was the proportion of population affected by Cholera.

3.2. Target Population

The target population was the proportion of population affected by cholera from 2000 to 2022 in Kenya (Y). The data was obtained from WHO and UNICEF.

3.3. Data Presentation

The data was presented using scatter plot and q-q plot. The scatter plot helped in visualizing the relationship between two variables and detecting presence of outliers. The Q-Q plot was used in testing the normality of the data.

3.4. Data Analysis

The data was analyzed using a qualitative research design and R statistical software version 4.3.2.

3.5. Model Formulation

3.5.1. General Model

The general multiple linear regression model equation was of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_p X_p + \varepsilon \qquad (1)$$

Where:
$Y =$ Dependent variable/response variable
$X_i =$ Independent variable/explanatory variable sample equation

3.5.2. The specific Multiple Linear Regression Model

The specific model contained three explanatory variables (Least Basic Drinking water, Unimproved Water Sources and Surface water) and one response variable(Proportion of population affected by Cholera).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \qquad (2)$$

where:
$Y =$ Proportion of population affected by cholera
$X_1 =$ Proportion of population using Least Basic Drinking Water (LBW)
$X_2 =$ Proportion of population using Unimproved water services (UBW)
$X_3 =$ Proportion of population using Surface water sources (SW).
ε (error term) ∼N$(0, 1)$

The estimates of $\beta_0, \beta_1, \beta_2$ and β_3 were the unknown parameters that explained how Least Basic Drinking Water, Unimproved Water Services and Surface Water contributed to cholera.

3.6. Testing Goodness of Fit

R^2 (coefficient of determination) is defined as the proportion of the total variation in Y that can be explained by X_1, X_2 and X_3 ;

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} \qquad (3)$$

Where;

$$\sum (Y_i - \bar{Y})^2 = TotalSumofSquares(SST) \qquad (4)$$

$$\sum (Y_i - \hat{Y})^2 = SumofSquaresduetoError(SSE) \qquad (5)$$

$$\sum (\hat{Y}_i - \bar{Y})^2 = SumofSquaresduetoRegression(SSR) \qquad (6)$$

$$AdjustedR^2 = 1 - \frac{SSE/(n-k)}{SST/(n-1)} \qquad (7)$$

$$= 1 - \frac{SSE(n-1)}{SST(n-k)} \qquad (8)$$

Where;
n=Sample size
K=Number of variables

3.7. Anova Table

Table 1. Anova Table.

Source of Variation	df	SS	MSS	F-statistic
Regression	$k-1$	SSR	$\frac{SSR}{K-1}$	$F-statistic = \frac{MSSR}{MSSE}$
Error	$n-k$	SSE	$\frac{SSE}{n-k}$	
Total	$n-1$	SST		

4. Results and Discussions

4.1. Descriptive Statistics

Table 2. Median, 1st quantile and 3rd quantile.

Population	Median	1st quantile	3rd quantile
LBW	4.0746	4.0389	4.1086
UBW	3.3471	3.2039	3.4865
SW	2.7873	2.7313	2.8425

From Table 2 it the median, 1st quantile and 1st quantile of each of the proportion of population using different kind of water respectively. The median values refers to the value below and above which half of the observation are ranked in order of size. The lower quantile aids in measuring the statistical dispersion that divided the data set into four equal sections . The lower quantile represented the value below which 25% of the data lies. The upper quantile aids in measuring the statistical dispersion that divided the data set into four equal sections .The upper quantile represented the value below which 75% of the data lie.

Table 3. Maximum , minimum, mean and standard deviation of the explanatory variables.

Population	N	Maximum	Minimum	Mean	SD
LBW	23	4.1409	4.0015	4.0733	0.0453
UBW	23	3.6222	3.0567	3.3441	0.1835
SW	23	2.8969	2.6748	2.7867	0.0721

From Table 3 it indicated the sample size, maximum, minimum, mean and standard deviation of the proportion of population using different kinds of water respectively. The average number of proportion using Least basic water was 4.0733, the proportion of population using Least basic water indicated that there was a deviation of 0.0453 from the mean. The average number of proportion using Unimproved Water was 3.3441. Based on the standard deviation, the proportion of population using Unimproved Water indicated that there was a deviation of 0.1835 from the mean. The average number of proportion using Surface water was 2.7867. From the column of standard deviation, the proportion of population using Surface Water indicated that there was a deviation of 0.0721 from the mean.

4.2. Diagnostic Test of Multiple Linear Regression Assumptions

Diagnostic test for normality, heteroscedasticity and linearity were conducted before fitting a multiple linear regression model. when the mentioned diagnostic test hold through, then a multiple linear regression model is fitted [5].

However, in some situations these assumptions may fail to be satisfied, as reported by [2] making it difficult to fit the linear regression model.

4.2.1. Test for Normality

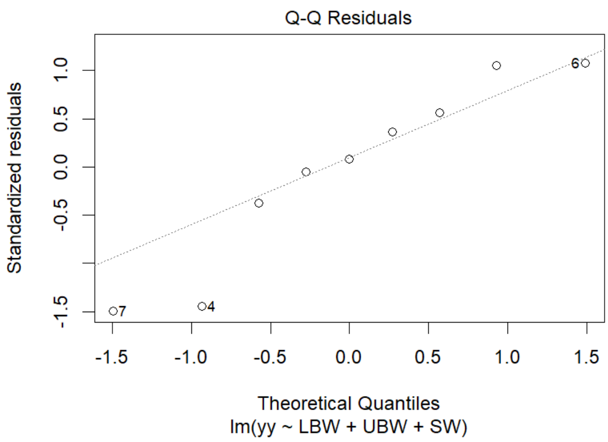

Figure 1. Q-Q plot of standard residuals against quantiles.

The Q-Q plot was used to test for normality. The normal Q-Q plot aid in assessing the normality of residuals. As indicated in the Figure 1 that the points on the plot were not deviating significantly from the straight line thus assuming normality exist.

4.2.2. Test for Heteroscedasticity

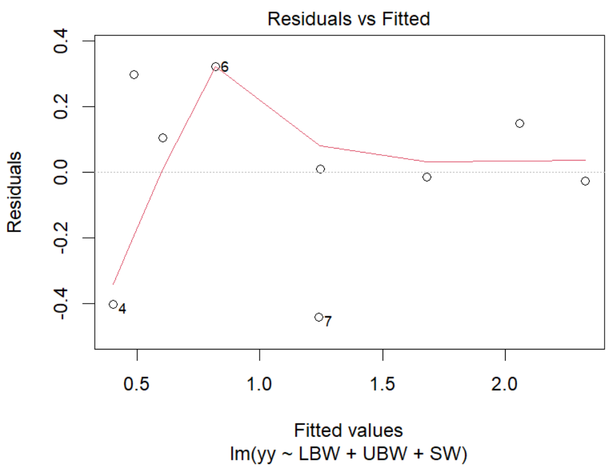

Figure 2. Plot of the Residuals against the fitted values.

Based on Figure 2, the plot of standard residual against fitted values formed a funnel shape or discernible pattern thus suggesting presences of heteroscedasticity.

4.2.3. Linearity

A scatter plot of each explanatory variable against the response variable was plotted to find out if there exist a linear correlation between the independent variables and dependent variable used in analysis shown in Figure 3, Figure 4, and Figure 5.

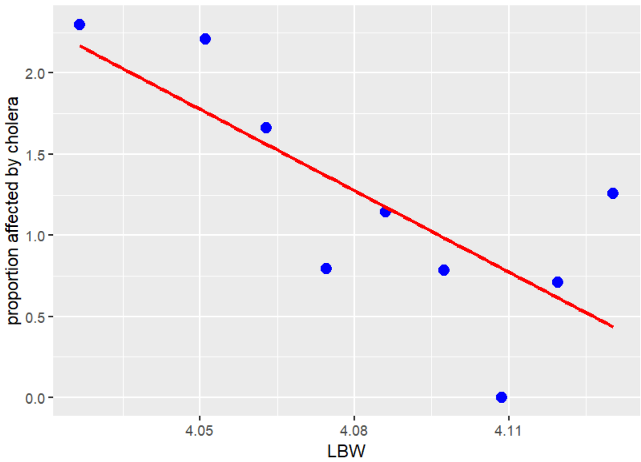

Figure 3. *The proportion of population affected by Cholera against proportion of population using LBW.*

From Figure 3, as the proportion of population using Least Basic Water increases, the number of Cholera cases decreases.

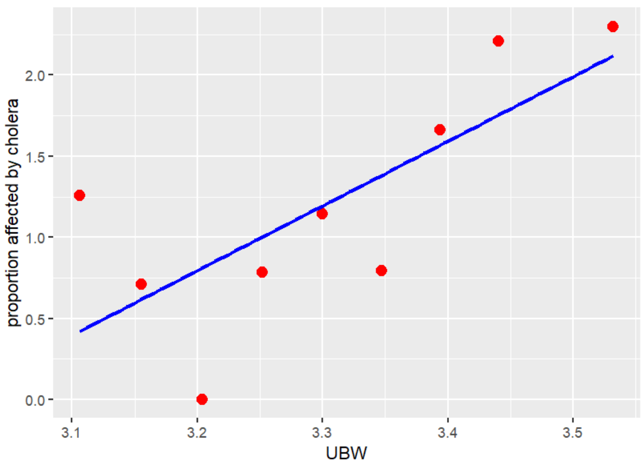

Figure 4. *The proportion of population affected by Cholera against the proportion of population using UBW.*

From Figure 4, as the proportion of population using Unimproved Water increases, the number of Cholera cases increases.

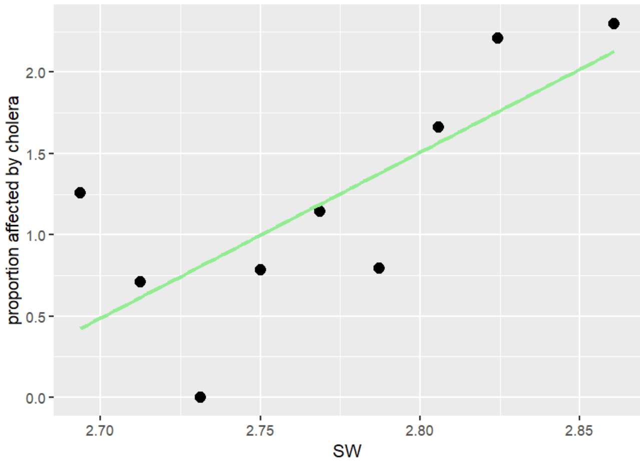

Figure 5. *The proportion of population affected by Cholera against the proportion of population using SW.*

From Figure 5, as the proportion of population using Surface Water increases, the number of Cholera cases increases.

Based on the 3 plotted scatter plots the data points in Figure 3, Figure 4 and Figure 5 indicated a linear relationship between proportion of population using least basic drinking water services, unimproved water services, surface water and proportion of population affected by Cholera. This is because most of the points do not lie far away from the line of best fit.

4.3. Parameter Estimation

$\hat{\beta_0}$, $\hat{\beta_1}$, $\hat{\beta_2}$ and $\hat{\beta_3}$ were least square estimates of β_0, β_1, β_2 and β_3 respectively and were determined using the ordinary least square method. Regression analysis was used to model the relationship between proportion of population using Least Basic Drinking Water (X_1), proportion of population using Unimproved Water services (X_2) and proportion of population using Surface Water sources (X_3). There was a need to determine the magnitude of the resulting relationship between the explanatory variables and use the outcome in making prediction based on the resulting model in Table 6.

Table 4. *The maximum , minimum, mean and standard deviation.*

Coefficients	Estimates	Std .error	t value	$Pr(>\|t\|)$
(Intercept)	-995.51	438.84	-2.269	0.0726
LBW	96.85	44.99	2.153	0.0840
UBW	-115.05	47.45	-2.425	0.0598
SW	353.75	148.73	2.378	0.0633

Having established the proportion of population using Least Basic Drinking Water (X_1), proportion of population using Unimproved Water services (X_2) and proportion of population using Surface Water sources (X_3), the study opted for a model that would provide the best fit in explaining the relationship. The fitted specific multiple linear regression model as in equation 9 was of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (9)$$

the estimates were fitted in the model to obtain

$$Y = -995.51 + 96.85X_1 - 115.05X_2 + 353.75X_3 + \varepsilon \quad (10)$$

Based on equation 10, the intercept β_0 was -995.51. This was the Y intercept; the value of Y when X_1, X_2 and X_3 were 0.

The proportion of population using Least Basic drinking water had a β_1 of 96.85. This indicated that proportion of population affected by Cholera changed by 96.85 when proportion of population using Least basic drinking water changes by a single unit.

The Proportion of population using Unimproved water had a β_2 of -115.05. This indicated that proportion of population affected by Cholera changed by -115.05 when proportion of population using Unimproved water changes by a single unit. Finally the Proportion of population using Surface water had a β_3 of 353.75 which indicated that proportion of population affected by Cholera changed by 353.75 when proportion of population using Surface water changes by a single unit.

4.4. Correlation Between Predictor Variables and Response Variable

Table 5. *Correlation Coefficient.*

Population	correlation coefficient
LBW	-0.7818763
UBW	0.7686439
SW	0.7713783

Based on Table 5 the results showed that that;

i. There exist a strong negative linear relationship between proportion of population using Least Basic drinking Water and proportion of population affected by Cholera.

ii. There exist a strong positive linear relationship between proportion of population using Unimproved water and proportion of population affected by Cholera.

iii. There exist a strong positive linear relationship between proportion of population using Surface water and proportion of population affected by Cholera.

Based on Table 5 there exist a strong positive linear relation between the proportion of population using surface water and the proportion of population affected by Cholera. This corresponds to the research done by [11] which also indicated relationship between Surface water and spread of Cholera. In addition to that there exist a strong negative linear relationship between the proportion of population using Least Basic drinking Water and the proportion of population affected by Cholera. This also corresponds to the research done by [8] which showed a statistical significance and negative effects between Least Basic drinking water services and Cholera.In conclusion there exist a strong positive linear relationship between the proportion of population using Unimproved basic water and the proportion of population affected by Cholera. This too corresponds to the findings of [4] which indicated that the danger of being affected by Cholera was linked to use of unimproved basic water sources.

4.5. Coefficient of Determination

The coefficient of determination was used to measure the proportion of variation in the response variable that can be explained by the explanatory variable.

Table 6. *Summary of Model.*

Multiple R-squared	Adjusted R-squared	Std .error	F-statistics
0.8828	0.8124	0.01229	12.55

Based on the analysis in Table 6 above, the R-squared was 0.8828 which indicated that 88.28% variations in the proportion of population affected by cholera can be explained by the proportion of population using Least basic water, Unimproved water and Surface water.

The adjusted R^2 was used in identifying the percentage variation in target field that can be explained by the inputs. The adjusted R^2 was mostly preferred to the multiple R-squared because multiple R-squared reduced the use of excess variables that are of no significance to the model. From the analysis the adjusted R^2 was found to be 0.8124, this showed that 81.24% variations in the proportion of population being affected by Cholera was explained by the proportion of population using Least basic water, Unimproved water and Surface water.

The adjusted R^2 was used to test for goodness of fit because adjusted R^2 is more accurate than the normal R^2 since the adjusted R^2 penalizes the model for using excess variables that does not have significant impact to the model.

5. Conclusion

The main aim of the study was to conduct a multiple linear regression analysis to determine factors contributing to the spread of Cholera in Kenya from 2000 to 2022. Based on the outcome the results showed that the various water sources contributed to the spread of Cholera.

The results showed that there exist linear relation between the Least Basic drinking water, Unimproved water, Surface water and the proportion of population being affected by Cholera. From the analysis, the adjusted R^2 was 0.8124, this showed that 81.24% variation in the proportion of population being affected by Cholera was explained by the proportion of population using Least basic water, Unimproved Water and Surface Water. The adjusted R^2 was used to test for goodness of fit because adjusted R^2 was more accurate than the normal R^2. This is is because adjusted R^2 penalized the model for using excess variables that did not have significant impact to the model.

Based on the analysis the proportion that used Least Basic drinking water services was the least affected followed by proportion that used Unimproved Water and finally proportion that used surface water lead to the spread of Cholera the most. In conclusion the study recommends an improved access to treated (Least basic water), increased investment in infrastructure (water storage facilities) especially in heavily populated areas such as refugee camps and finally creation of awareness to the general population on dangers of using untreated water, increase training on disease surveillance and outbreak response.

Abbreviations

WHO	World Health Organization
UNICEF	United Nations International Children's Emergency Fund
MOH	Ministry of Health
IDSR	Integrated Disease Surveillance and Response
IRC	International Rescue Committee
MLR	Multiple Linear Regression
df	Degrees of Freedom
LBW	Least Basic Water
UBW	Unimproved Basic Water
SW	Surface Water
SS	Sum of Squares
MSS	Mean Sum of Squares

Acknowledgments

We wish to express our sincere gratitute to Roseline Achieng Oburu, Celestine Muturi Bosibori and Kirinyaga University fraternity for their support.

ORCID

0009-0000-6010-2525 (Samuel Okinyi),
0000-0001-7572-7325 (Joseph Njuguna Karomo),
0009-0008-5128-8338 (Nzioka Dorcas Mbinya),
0009-0004-4197-4490 (Mohammed Kassim Thoya)

Author Contributions

Samuel Okinyi: Conceptualization, Formal Analysis, Methodology, Visualization, Writing ¨C original draft, Writing ¨C review & editing

Joseph Njuguna Karomo: Project administration, Resources, Supervision, Validation

Nzioka Dorcas Mbinya: Data curation, Software, Visualization, Writing ¨C review & editing

Mohammed Kassim Thoya: Data curation, Investigation, Visualization, Writing ¨C review & editing

Conflicts of Interest

We hereby declare no conflicts of interest.

References

[1] Almagro-Moreno, S., & Taylor, R. K. (2013). Cholera: environmental reservoirs and impact on disease transmission. Microbiology spectrum, 1(2), 10-1128. https://doi.org/10.1128/microbiolspec.oh-0003-2012

[2] Ayinde, K., Apata, E. O., & Alaba, O. O. (2012). Estimators of linear regression model and prediction under some assumptions violation. https://doi.org/10.4236/ojs.2012.25069

[3] Charnley, G. E., Kelman, I., & Murray, K. A. (2022). Drought-related cholera outbreaks in Africa and the implications for climate change: a narrative review. Pathogens and global health, 116(1), 3-12. https://doi.org/10.1080/20477724.2021.1981716

[4] 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), 156-156. https://doi.org/10.11604/pamj.2017.28.101.12806

[5] Gaffney, M. (2016). Nature, economy, and equity: Sacred water, profane markets. American Journal of Economics and Sociology, 75(5), 1064-1231. https://doi.org/10.1111/ajes.12169

[6] Girotto, C. D., Behzadian, K., Musah, A., Chen, A. S., Ali, A., & Campos, L. C. (2021). Impact of water and sanitation services on cholera outbreaks in sub-Saharan Africa.

[7] Hickey, G. L. Checking model assumptions with regression diagnostics.

[8] 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. https://doi.org/10.2166/wh.2013.238

[9] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

[10] Mutonga, D., Langat, D., Mwangi, D., Tonui, J., Njeru, M., Abade, A., ... & Dahlke, M. (2013). National surveillance data on the epidemiology of cholera in Kenya, 1997-2010. The Journal of infectious diseases, 208 (suppl-1), S55-S61. https://doi.org/10.1093/infdis/jit201

[11] Ntajal, J., Falkenberg, T., Kistemann, T., & Evers, M. (2020). Influences of land-use dynamics and surface water systems interactions on water-related infectious diseases? A systematic review. Water, 12(3), 631. https://doi.org/10.3390/w12030631

[12] Ogisma, L., Li, T., Xiao, H., O'Donnell, F., & Molnar, J. (2021). Analysis of community level factors contributing to cholera infection and water testing access in the Northern Corridor of Haiti. Water Environment Research, 93(10), 1819-1828. https://doi.org/10.1002/wer.1591

[13] Oyugi, E. O., Boru, W., Obonyo, M., Githuku, J., Onyango, D., Wandeba, A., ... & Gura, Z. (2017). An outbreak of cholera in western Kenya, 2015: a case control study. The Pan African Medical Journal, 28 (Suppl 1). https://doi.org/10.11604/pamj.supp.2017.28.1.9477

[14] Rebaudet, S., Sudre, B., Faucher, B., & Piarroux, R. (2013). Environmental determinants of cholera outbreaks in inland Africa: a systematic review of main transmission foci and propagation routes. The Journal of infectious diseases, 208 (suppl-1), S46-S54. https://doi.org/10.1093/infdis/jit195

[15] Shapiro, R. L., Otieno, M. R., Adcock, P. M., Phillips-Howard, P. A., Hawley, W. A., Kumar, L., ... & Slutsker, L. (1999). Transmission of epidemic Vibrio cholerae O1 in rural western Kenya associated with drinking water from Lake Victoria: an environmental reservoir for cholera? The American journal of tropical medicine and hygiene, 60(2), 271-276.

[16] Shultz, A., Omollo, J. O., Burke, H., Qassim, M., Ochieng, J. B., Weinberg, M., ... & Breiman, R. F. (2009). Cholera outbreak in Kenyan refugee camp: risk factors for illness and importance of sanitation. American Journal of Tropical Medicine and Hygiene, 80(4), 640.

[17] Sulayman, F. A. T. I. M. A. (2014). Modelling and analysis of the spread of cholera disease in Nigeria with environmental control. Research thesis.