# AN EVALUATION AND ESTIMATION OF RISK FACTORS ASSOCIATED WITH CHOLERA : A CASE STUDY OF REGISTERED PATIENTS IN RAYMOND MHLABA LOCAL MUNICIPALITY, SOUTH AFRICA

**University of Fort Hare**

*Together in Excellence*

## A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN BIOSTATISTICS & EPIDEMIOLOGY

BY

OSUJI  GEORGELEEN O.

(201315879)

DEPARTMENT OF STATISTICS

FACULTY OF SCIENCE AND AGRICULTURE

UNIVERSITY OF FORT HARE

ALICE, 5700  -  SOUTH  AFRICA

SUPERVISOR: DR  J  NDEGE

CO-SUPERVISOR: PROF Y QIN

2017

# DECLARATION

I, the undersigned, declare that this dissertation submitted to the University of Fort Hare for the award of degree of Masters of Science in Biostatistics and Epidemiology in the Department of Statistics, Faculty of Science and Agriculture, is my own original work and that I have not previously, in it's entirely or in part, submitted this dissertation and the content therein to this University or any other institution for degree award or any other purpose.

Signature/date  _____

Osuji Georgeleen O

Signature/date  ____              _____

Dr  J  Ndege (Supervisor)

Signature/date  _____

Prof Y Qin (Co-Supervisor)

# DEDICATION

This dissertation is dedicated to my beloved son and husband

# ACKNOWLEDGEMENTS

# ABSTRACT

**Background:** Cholera is an acute infectious disease of the small intestine caused by the bacterium called *Vibrio cholerae,* which has two serogroups01and 0139which is also known as choleragenic *V. cholerae*. This disease is characterized by profuse watery diarrhoea and severe dehydration which can lead to death of both adult and children if treatment is not promptly given. Cholera is spread through ingestion of *V. cholera* contaminated water and food. Cholera has displayed global presence more than seven times and caused tremendous disaster to humankind.

**Method:** This was a retrospective study among patients with cholera within the period of ten years (2005 to 2015) and the total number of patients was 106. The target population for this study were patients at Raymond Mhlaba Local Municipality who attended Victoria hospital and were diagnosed with *Vibrio cholerae* species with respect to sources of water and non-water sources during the mentioned period. A multivariate Logistic regression was used to determine the risk factors of cholera and comparison was made in the treatment of cholera outcomes for factors which were statistically significant at $P < 0.05$.

**Results:** The median age was 24.5 (IQR: 7.0-44.8) for all respondents with cholera. Patients within the age range of 26-40 and 41-55 were found to have a higher risk of cholera (2.20, 95% CI: 1.51, 4.22) and (1.13, 95% CI: 0.61, 2.01) respectively. The risk of cholera was considerably higher among the black race (2.51, 95% CI: 1.52, 4.31) compared to the coloured (1.33, 95% CI: 0.75, 3.713). Patients who used source of water supply from carrier/Tanker and Dam/River had higher increased risk of contracting cholera (1.71, 95% CI: 0.92, 3.62)  and  (2.61, 95% CI: 1.38, 4.25) respectively  compared to patients that used other sources of water.

Home, party and restaurant as places patients had eaten 24 hours earlier before the onset of cholera were associated with increased risk of severe cholera. Patients who shared toilet facilities

had increased risk of cholera (0.91, 95% CI: 0.47, 1.62) compared to the ones who used private toilet. Those patients who did not practice hand washing had an increased risk of contracting cholera (1.45, 95% CI: 0.88, 2.12) compared to the ones who washed their hands.

When Logistic regression was carried out, the following risk factors were found to be statistically significant in causing cholera at 5% significance level; Age ( 26-40), gender, level of education, marital status, sources of water supply, place eaten in the last 24 hours before onset of cholera, type of toilet used and hand washing.

**Conclusion:** Improvement in level of education, sources of water supply, place of last eaten before cholera sickness, toilet facilities, hand washing practices are key risk factors for cholera disease and hospitalization among patients in Raymond Mhlaba local Municipality, Eastern Cape. The strong association between water and sanitation highlights the need for a more thorough assessment of potential waterborne exposures and the risk faced by family members suffering from cholera infection cases and may warrant renewed research regarding the use of targeted chemoprophylaxis in endemic rural settings.

**KEY WORDS : Cholera, Risk factors, Odds ratio, Logistic regression**

## ACRONYMS

| | |
|---|---|
| CFR | Case fatality rate |
| CI | Confidence Interval |
| CSS | Cross Sectional Study |
| CTX | Cholera Toxin |
| GIS | Geographic Information System |
| GLMs | Generalized linear models |
| NHRD | the National Health Research Database |
| OCV | Oral Cholera Vaccines |
| OR | Odds Ratio |
| ROC | Receiver Operating Characteristics |
| SAS | Statistical Analysis System |
| SPSS | Statistical package for Social Science |
| STATA | Statistics Data |
| TCBS | Thiosulphate Citrate Bile Salts |
| TCP | Toxin Co regulated Pilus |
| UN | United Nations |
| VIF | Variance Inflation Factor |
| VPI | Vibrio Pathogenicity Island |
| WASH | Water Sanitation and Hygiene |
| WHO | World Health Organisation |

## LIST OF FIGURES AND APPENDIX

**LIST OF TABLES**

**Table of Contents**

# CHAPTER ONE

## 1.0  Introduction

Globally, cholera continues to be a threat to public health and a key indicator to lack of economic and social development (Dick et al., 2012; Harris et al., 2012), which is mainly discovered in developing and underdeveloped countries. It is one of the most recent re-emerging infectious communicable and waterborne diseases.

Cholera is an acute infectious disease of the small intestine caused by a bacterium called *Vibrio cholerae,* which has two serogroups 01 and 0139 (Harris et al., 2012) which is also known as choleragenic *V. cholerae*. This disease is characterized by profuse watery diarrhoea and severe dehydration which can lead to death of both adult and children if treatment is not promptly given. It is believed that it is spread through contaminated water and food since its first prevalence in Ganges, India in 1817. Cholera has displayed a global presence more than seven times and caused tremendous disaster to humankind (Xu, Kan, & Wang, 2015).

South Africa is a water-scarce country and the demands on this resource are growing as the economy expands and its population increases. For the country to continue to develop economically, while meeting the wide-ranging needs for water, urgent steps must be taken to protect the quality of the resource (Manager, Umfolozi, War, & Hospital, 2005). The United Nations (UN) set a goal in their Millennium Declaration to reduce the amount of people without safe drinking water by half by the year 2015 in South Africa (The Nations & Declaration, 2014). Safe drinking water for human consumption should be free from pathogens such as bacteria, viruses and protozoan parasites, meet the standard guidelines for taste, odour, appearance and chemical concentrations, and must be available in adequate quantities for domestic purposes. However, inadequate sanitation and persistent faecal contamination of water sources is responsible for a large percentage of developing countries not having access to clean and  safe drinking water and hence the continued outbreak of

cholera diseases. However the pattern of cholera epidemic in South Africa has demonstrated that rural communities are still at risk of cholera. Most patients of the cholera epidemic were rural dwellers with little or no access to basic and primary health care services (Cottle, 2002).

## 1.1 Background of the Study

Cholera infects 3-5million people every year around the world and one hundred thousand to one hundred and twenty thousand people die of the infectious disease, according to world health organisation estimates. The existence of the disease was first noticed around 470-400 B.C in ancient Greece during the time of Lord Buddha and Hippocrates and in India in Sushrute-Samhita around B.C.
(Hays, 2005; http:www.choleraandthethames.co.uk/cholera-in-london/origins-of-cholera).

It was first recorded in 1563 in an Indian medical report but in more modern terms, the story of the disease begun in 1817 when it spread from its ancient homeland of the Ganges Delta in India to other part of the world (http:www.choleraandthethames.co.uk/cholera-in-london/origins-of-cholera/). The infection is now largely confined to developing countries in the tropics and subtropics. It is endemic in Africa, parts of Asia, the Middle East, and South and Central America. In order to promote the prediction and early warning of cholera breakouts, many studies have investigated the regional and environmental factors for cholera, mostly concentrated on Southern Asian countries such as Bangladesh (Xu et al. 2015; Jutla et al. 2013; Bouma & Pascual 2001), India (Kanungo et al. 2010). In Latin America countries such as Mexico (Borroto & Martinez-Piedra 2000)and Peru (Gil et al. 2004), and in some African countries including Zimbabwe, Democratic Republic of Congo (DRC), Kenya (Mendelsohn & Dawson 2008; Fleming et al. 2007; Zuckerman et al. 2007).

As early as in 1971, South Africa was considered to be at risk of cholera due to its hot, humid summers, seaports, overcrowded communities with low standard of environmental sanitations and scanty, restricted and unprotected water supplies in certain township areas

(Lamond & Kinyanjui 2012; Connor et al. 2011). A study conducted in Lebowa (Former Transvaal province in the North Eastern part of South Africa ) by the Department of Health, Welfare and Pensions (Quick et al. 1996; Igbinosa et al. 2010; Wahed et al. 2013) to determine the mode of transmission of cholera found that consumption of open river water was positively associated with an increased risk of contracting the disease, and that cholera outbreaks were associated with rainfall pattern and temperature.

According to the world health Organisation(February et al. 2014), simple preventive measures such as safe disposal of human excreta, particularly those of babies and persons infected with diarrhoea causing bacteria, hand-washing after defecation and handling infant faeces before feeding and preparing foods, and maintaining drinking water free from faecal contamination both at homes and  in hospitals were enough to curb the spread of cholera in rural communities. Although the risk factors for cholera in rural areas of Eastern Cape have been known to public health professionals working in the province, the spread of cholera and waterborne diseases remains an endemic health problem affecting the rural population during heavy winters and prolonged rainy seasons (Igbinosa et al. 2010). The rural population of the province lack access to safe water as well as basic health and primary health care services. Illiteracy, poverty and unemployment affect almost half of the rural population of the province. Many rural households do not know how to prepare oral re-hydration solutions at home. Drinking water is consumed without being boiled by rural households due to lack of knowledge and/or socio-economic status of the rural dwellers. The extent and content of coverage of health education on environmental sanitation and personal hygiene is grossly inadequate. As a result, the population may be vulnerable to communicable diseases such as cholera in heavy winters and prolonged rainy seasons.

The species *V. cholera* comprises both pathogenic and non-pathogenic strains. *Vibrio cholerae* O1 and O139 are the only serotypes, known till date as, responsible for the disease defined clinically and epidemiologically as cholera (López-Gigosos et al. 2011; Tamang et al.

2005). Cholera cases are confirmed through the isolation of *Vibrio cholera* O1 or O139 from stools in any patient with diarrhoea (WHO 2004). In endemic areas, outbreaks usually occur when war or civil unrest disrupts public sanitation services. Natural disasters like earthquake, tsunami, volcanic eruptions, landslides and floods also contribute to outbreak by disrupting the normal balance of nature (Qadri et al. 2005). This creates many health problems, food and water supplies can become contaminated by parasites and bacteria when essential systems like those for water and sewage are destroyed. In newly affected areas, outbreaks may occur during any season and affect all ages equally. The organism normally lives in aquatic environments along the coast. People acquire its infection by consuming contaminated water, seafood, or other foods. Once infected, they excrete the bacteria in stool (Adagbada et al. 2012). Thus, the infection can spread rapidly, particularly in areas where human waste is untreated.
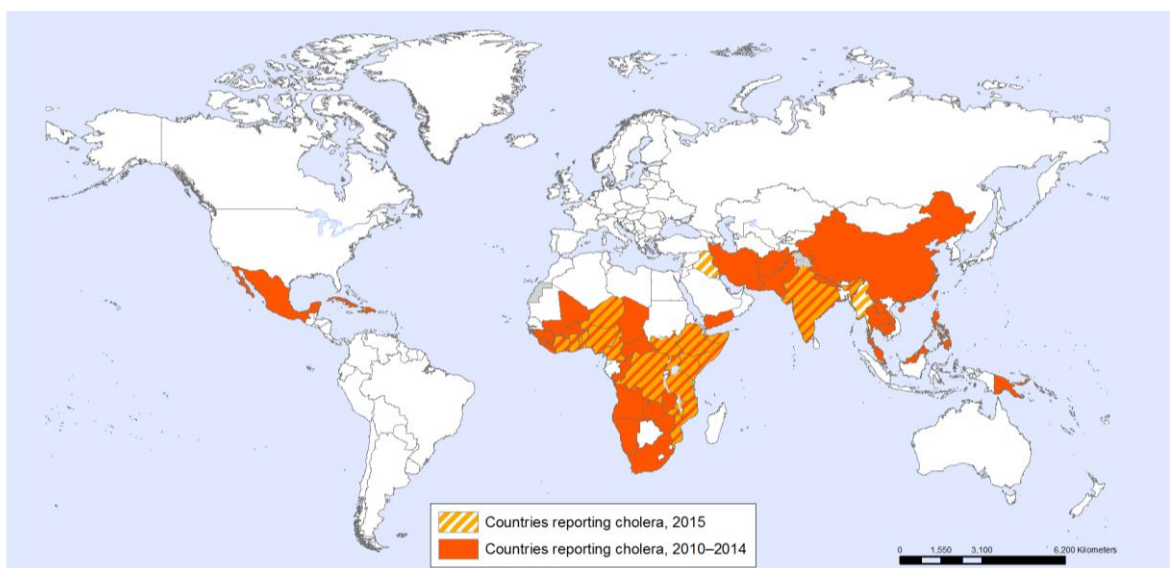


Fig 1.1: WHO map of countries reporting cholera epidemic from 2010 to 2015

(Sources: http://www.who.int/gho/epidemic_diseases/cholera/epidemics/en/; accessed March 1, 2017)

## 1.2 Research Problem Statement

It is recognized that *V. cholera* is a component of coastal and estuarine microbial ecosystems, with copepod species of zooplankton that comprise the aquatic fauna of rivers,

bays, estuaries and the open ocean serving as hosts for the bacterium (Constantin de Magny et al. 2008). The growth and abundance of the organisms in coastal waters is influenced by changes in environmental factors including temperature, salinity, nutrient availability, sea surface height, and rainfall (Cash et al. 2010; Emch et al. 2008; Lobitz et al. 2000). However, About 75% of people infected with *V. cholera* do not develop any symptoms, although the bacteria are present in their faeces for 7-14 days after infection and are shed back into the environment, potentially infecting other people. Among people who develop symptoms, 80% have mild or moderate symptoms, while around 20% develop acute watery diarrhoea with severe dehydration (Mintz & Tauxe 2013). In severe infections, more than one quart of water and salts is lost per hour. The stool looks grey and has flecks of mucus in it- termed "rice water stools". Within hours, dehydration can become severe, causing intense thirst, muscle cramps, and weakness. Very little urine is produced and the eyes may become sunken, and the skin on the fingers may become much wrinkled. If dehydration is not treated, loss of water and salts can lead to kidney failure, shock, coma, and death. In people who survive, symptoms usually subside in 3 to 6 days. Most people are free of the bacteria in two weeks. The bacteria remain in a few people indefinitely without causing symptoms.

## 1.3 Aims and Objectives of the Study

### Aim of the Study

The aim of this study is to ascertain the transmission and control of cholera outcomes in registered patients with respect to their sources of water among the inhabitants of Raymond Mhlaba Local Municipality, Eastern Cape, South Africa.

### Objectives of the Study

The specific objectives of the study are as follows:

1. To identify and analyse the effect of the risk factors of cholera in Raymond Mhlaba Local Municipality.

2. To use Multiple Logistic Regression to model the risk factors associated with cholera.

3. To check Multicollinearity and non-linearity between the dependent and independent variables associated with cholera.

## 1.5 Organisation of the dissertation

The study report is structured into five chapters. Chapter one focuses on the introduction, background to the study, research problem statement, objectives of the study and its significance. Chapter two reviewed literature on cholera, causes and symptoms as well as, the works and studies already carried out on cholera. The effects and pre-disposing risk factors to cholera disease and some of the methods used in the control of cholera were as well articulated in this chapter. Chapter three emphasizes on the methodology used in the study. Chapter four, deals with data analyses and interpretation of the results. Finally, chapter five discusses the conclusions and recommendations made for the control of cholera.

## 1.4 Significance of the study

An estimated 30-40% of South Africa's population had been without adequate water supply services prior to 1994 (Mackintosh et al., 2004). However, the Millennium Development Goal of halving, "by the year 2015 the proportion of people who are unable to reach or to afford safe drinking water" in 2005 has been achieved by the South African government through responsible agencies like the Department for Water Affairs and Forestry (www.dwaf.gov.za/WFGD/documents/WfGDv6). Nonetheless, in a bid to avail the populace with sufficient water supply taking into cognizance its quality and quantity, water has been recycled through treatment of various wastewaters and discharging the final effluents into environmental water sheds (Leong et al., 2008). Although the final effluents from wastewater treatment plants discharged into the environment are treated at various levels to make it safe and reusable, reports are emerging of pathogens surviving these treatment processes and

worse still, some species not known to cause disease in humans and animals previously, are emerging with pathogenic traits resulting in environmental distress syndrome culminated by the changing environmental conditions characterized by industrialization, globalization, human and animal migration as well as other stressor factors of biotic and abiotic nature (Boyd et al., 2000).

As South Africa aims at improving the quality of water supplied to all  people living in the country, such measures as having a comprehensive evaluation of all its wastewater final effluents and checking water sheds as potential sources of pathogenic and emerging pathogenic Vibrio species needs to be prioritized as this study anticipates to address.

# CHAPTER TWO

## LITERATURE   REVIEW

**2.0 Introduction**

The severity and extent of cholera is recognized worldwide. It is listed as one of three internationally quarantinable diseases by the World Health Organization, along with plague and yellow fever (World Health Organization; UN-Water 2014). It is the first disease that received organized modern public health surveillance and reporting of worldwide incidence. Cholera is one of the most researched communicable diseases, yet it still has devastating effects on local communities (Collins; 2003). Most of the new cases reported since 2000 are from eastern and Southern Africa. One recent study estimates there are 1.3–4.0 million cholera cases and 21 000–143 000 deaths per year. As of 2012, there were 69 cholera endemic countries with annual cholera incidences ranging from 10 to 2600 cases per 10 000 persons (Ali et al., 2015).

The cholera epidemic in Haiti following the catastrophic earthquake in 2010 provides a pertinent example, where745, 588 cases and 8972 deaths were reported to Ministry of Health as at 10 August 2015. Before the year 1994, an estimation of 30%-40% of South Africa's population was without water adequate water supply service (Mackintosh et al., 2004). Although, the millennium development goal of having the proportion of people who are unable to afford or reach safe drinking water in 2005 has been achieved by the South African government through responsible agencies like the Department for water affairs and Forestry (www.dwaf.gov.za/WFGD/documents/WfGDv6). South Africa experienced major cholera outbreaks during 1980 to 1984 and further outbreaks have occurred since August 2000. From August 2000 to February 2002, the disease infected at least 114 000 people in KwaZulu-Natal (more than 70% of the total cases reported in the country) and claimed at least 259 lives in the province (Cottle& Deedat 2002). Initial reports of the cholera outbreak in 2000 came from the largely rural and impoverished communities near Empangeni town. The

8

source of the epidemic was traced to the Umhlathuze River, in the northern part of the KwaZulu-Natal Province. Subsequent outbreaks have occurred in the Eastern Cape and Mpumalanga provinces. Hot and humid climatic conditions and socio-economic conditions are risk factors, with most cases reported during the 'hot' months of the year,

Aryal*et al.* (2012) reported that tap water/piped water was found to be the most common source of drinking water in the urban areas, whereas a tube well or borehole was common source of drinking water in the rural areas(Aryal et al. 2012). People are more likely to drink untreated water as water supply system in the rural areas do not have provisions of water treatment facilities. During the summer, there is huge scarcity of water in South Africa, while in winter, the availability of water increases with increase in rainfall but is severely contaminated with excreted organisms due to surface water runoff (Pokhrel & Viraraghavan 2004).

## 2.1 Cholera outbreaks in South Africa

In 2001, 106,389 people were infected with Cholera and as a result 232 died of the epidemic in Kwazulu-Natal, Northern Province, Gauteng and Mpumalanga (www.health24.com/Medical/Cholera/Recent-cholera-outbreaks-in-SA-20120721). In 2003, South Africa suffered cholera outbreak where 3901 cases were reported in KwaZulu-Natal, Eastern Cape and Mpumalanga with 45 deaths in total.

In Mpumalanga's Nkomazi region, which borders Mozambique 1,773 cases were reported in 2004 as well as 29 deaths. During the same period in 2004 in Eastern Cape Province, 738 people were diagnosed with cholera which resulted in four deaths. Also in North West Province,260 cases occurred where two people died (www.health24.com/Medical/Cholera/Recent-cholera-outbreaks-in-SA-20120721).

South Africa had seven bouts of cholera epidemics in the 1980s usually between the months of June to October (www.health24.com/Medical/Cholera/Recent-cholera-outbreaks-in-SA-20120721).

However, in the recent past, several outbreaks of cholera in the Eastern Cape Province resulting in morbidities and mortalities have been reported. In the consecutive years of 2002, 2003, 2004, 2007 and 2013, there were reported cases of cholera outbreaks in King SabataDalindyebo Local Municipality (Umtata) under the OR Tambo District Municipality, Ntabankulu and Qumbu both suburban towns in the same municipality, reported that thirteen people died with many hospitalized from cholera (Health Systems Trust, 2002, 2003, 2004). Prior to the outbreak in 2004, there was also a reported outbreak of waterborne diarrhoea in the municipality which was thought to be cholera although it was later found out to have been misdiagnosed (Health Systems Trust, 2004).

In 2007, the Eastern Cape Health Department in Queenstown reported an apparent cholera outbreak at Ilinga Township which was associated with sewerage flooding in the township (News 24 archives, 2007). Similarly, on the 24th of January 2014, 36 people in Fort Beaufort and its surrounding communities have been hit by an outbreak of diarrhoea and were hospitalized. This was reported by in the SABC news. Residents attributed the poor quality of municipal supplied water as the source of the problem
 (SABC; www.SABCNewsOnline/posts/10152142053906543).


## 2.2 Previous related research

## 2.2.1 Risk factors for sustained cholera transmission, Juba County in southern Sudan 2014

The study by Thomas el al. (2015) identified the risk factors for the 2014 cholera disease outbreak in Juba County as associated with eating away from home and during commuting, poorly treated drinking water and the failed oral cholera vaccination administered in the County in the month of April, 2014.

The age range of the study population was between 2 to 69 years and the sample size was 134. A team of 19 trained research assistants administered a semi structured

questionnaire and conducted the environmental assessments to evaluate the use of safe drinking water, improved food and personal hygiene, sanitation facilities and oral cholera vaccination. The statistical package used was Epi info (centres for diseases control and prevention, Atlanta Georgia USA) matched unadjusted and adjusted odd ratios were calculated using bivariate and multivariate models respectively to identify risk factors for cholera.

It came out clearly that eating and travelling outside the home were risk factors for becoming sick and that getting oral cholera vaccination and treating drinking water at home provides protection against the illness. For prevention and control, it was recommended that global oral cholera vaccine stockpiles be enhanced and also to educate people about the risk factors of cholera.

### 2.2.2 Sub-Sahara Africa Cholera Epidemic: South Africa and Tanzania

This was a cross sectional, comparative and descriptive study among households of Kwazulu Natal in South Africa conducted by Hoquemonjur AKM of university of Pretoria in 2003.Thirty communities were selected by systematic random selection and were divided into two groups; cholera patients (group1) and non-cholera patients (group2). A pre-test questionnaire was administered by trained personnel and data was collected during the months of November to December 2001. A total number of 1420 households from both groups constituted the study population. The statistical tool employed for data analysis included the two-sample test on mean and proportions, specificity test, sensitivity test, odd ratio, Pearson's chi-square tests of association, binary logistics regression analysis and ROC (receiver of characteristics ) analysis. The factors seen to influenced protection against diarrhoea disease includes boiling water prior to consumption (OR=0.41, 95% CI 0.19-0.90) and use of disinfectant (JIK) on surfaces within the homes (OR=0.45, 95% CI, 0.19-0.94). Similarly, the use of untreated river or dam water was significantly associated with diarrhoea

disease (OR=2.92,95% CI,1.06-7.80).Thus, conclusively, there was significant difference between two groups of household with regards to knowledge and practise of good hygiene, basic provision of safe water and effective utilization of sanitary facilities.

The study by Camilo et al (2001) identified and described risk factors of the 1997 cholera epidemic in rural area (Ifakara) in southern Tanzania. The research was conducted as a prospective hospital-based, matched case-control study design to identify cholera risk factors (bathing in the river, residing 10 minutes' walking distance from water source and eating dried fish) with analysis based on the first 180 cases and 360 matched controls.

The age range of the study population was from 5 years and above and the duration of the study was from June 23 to December 31 1997. Patient hospitalized with acute onset of watery diarrhoea and the place of cholera outbreak was first noticed around Lumemo River. Data on possible risk factors were documented by a standardized questionnaire and were double-entered into FoxPro database (Microsoft Corp). Statistics analysis was performed with STATA statistical software (Stata C.orp; 1997). Mcneman's chi-square test was used for the univariate analysis and multivariate conditional logistics regression. The level of significance were variables associated with cholera(bathing in the river, residing 10 minutes' walking distance from water source and eating dried fish)  at or below 5 %.

Previous related studies likewise applied similar statistical methods; univariate analysis, bivariate analysis, multivariate conditional logistic regression and logistics regression analysis. Questionnaires were administered to patients or guardians/parents for data collection. Different software packages such as SPSS, STATA and Epi info were used for data analysis.

Therefore multiple logistic regression analysis (bivariate and multivariate) is a dominant statistical method used to show the determinants of risk factor associated with disease outbreaks.

## 2.3  Vibrio cholera

*Vibrio cholera* is a gram-negative, nonspore, curved rod-shaped bacterium that is part of the vibronaceae family whose motility depends on a single polar flagellum. Vibrio is sensitive to low PH and die rapidly in solutions of PH below six. In aerated solutions they reach higher population densities. However, they are quite unbiased of alkaline conditions and also grow anaerobic ally. The nutritional requirement of *Vibrio cholerae* is simple and fresh isolates are prototrophic. In favourable conditions they grow rapidly with a generation time of less than half an hour. The growth rate of Vibrio cholerae depends on nutrient availability. Nevertheless, the growth of the pathogen follows a logistic dose-response curve (codeco, 2001; Jensen et al, 2006).

*Vibriocholerae* is huge and the species are very diverse. The sero-groups are divided into 206 (Yamai et al., 1997; Shimada et al., 1994) which are known as 01, 02, 03, etc. Only 01 and 039 sero-group is pathogenic (WHO, 1996; Alexander, 2008). There is existence of both toxigenic and non-toxigenic strains. Non-toxigenic strains can obtain toxicity through a temperate bacteriophage (Albert et al, 1998). Sero-group 01 is divided into three serotypes, namely Ogawa, Inaba and Hijokima (Reidl and Klose, 2002). The sero-group from this strains are divided into two biotypes, E1 Tor and Classical (Koch, 1884). All the members of the sero-groups include these two biotypes.

It is only the toxigenic strain of sero-group 01 and 0139 that causes cholera. Some non-01 strains causes diarrhoea but are not epidemic or endemic (Faruque et al, 1998). Occasionally the non-01 strains are isolated from cases of diarrhoea and they are ubiquitous in estuarine environments (Kaper et al, 1995). By bacteriophages, horizontal gene transmission and other aims are responsible for creating pathogenic *Vibrio cholera* by carrying genes involved in the colony of human and production of cholera toxin (Islam et al.,1997).

There are two most critical genes that are clustered together (genetic material) namely, the Vibrio pathogenicity island (VPI) conceals a number of genes, including those needed to express the toxin co regulated pilus (TCP), which is essential for colonization of the intestines. The lysogenic bacteriophage CTX, which harbours the genes for cholera toxin, uses the TCP to enter only those strains of *Vibrio cholera* capable of colonizing humans. It then lysogenizes conquered cells by amalgamating its whole genome into bacterium. Hence, a relationship between Vibrio cholera and CTX- ϕ showing TCP is mandatory to cause a cholera epidemic. Reviews suggest that the new variant strains detected lately in various parts of Asia and Africa cause more severe cholera with higher rates of fatality (WHO, 2010).

### 2.3.1  Ecology of *Vibrio cholera*

*Vibrio cholera* is a heterotrophic bacterium, which attaches itself to a wide variety of aquatic organisms, mainly plankton. It lives in two different habitats which is in human intestines and aquatic ecosystems. *Vibrio cholera* cannot synthesize its own food and responsible for mineralizing organic matter which forms an important component of aquatic food webs and nutrient cycles (Biddanda and Cotner, 2002).

The resulting ecological changes and human activity directly affect the bacterium persistence and spread in aquatic environments. Vibrio cholera reacts to environment conditions by decreasing or increasing its rate of metabolism, and therefore can enter a state of dormancy. In cholera endemic areas, outbreaks start when people get infected with the pathogen from the environment. The outbreaks may be accelerated by faecal contamination (Franco, 1997).

Cholera is an acute epidemic infection disease caused by the bacterium *Vibrio cholera*e which has short incubation period from less than one day to five days.  The probability of getting sick upon contact with a contaminated person depends on the pathogen density and interactions of the pathogen with the immune system (Cash et al, 1974).The three

major epidemiologic patterns in cholera are: neo-epidemic (newly invaded, cholera-receptive areas), heavily endemic and in developed countries with good sanitation, occasional limited outbreaks (WHO 2009). These patterns depend hugely on environmental factors.

## 2.4  Causes and Symptoms of Cholera

The bacterium, *Vibrio cholerae*, is the cause of cholera which is a severe water-borne infectious disease (Ryan 2004; WHO, 2010). The *Vibrio cholera* sero-group is divided into about 200 groups which only serotypes 01 and 0139 contains pathogenic members (Alexandra, 2008). The short incubation period cholera has is less than one day to five days (Obeng, 2015). However, the effect of the deadly disease are as a result of a potent toxin called CTX which the bacteria produces in the small intestine. CTX binds to the intestine walls, where it impedes with the normal flow of sodium and chloride. This causes the body to secrete huge amount of water, leading to diarrhoea and a rapid loss of fluids and salt (electrolytes).

The main source of cholera infection is contaminated water supplies, although uncooked fruits, raw shellfish, vegetables and other foods can also harbour *Vibrio cholera*. *Vibrio cholera* survives in humans, animals and the environment. In humans, it may cause illness and is continually shed through the stool (Desmarchelier, 1997; Greenough, 1999). In environment, cholera causing bacteria occurs in coastal waters where they get attached to copepods (tiny crustaceans). The bacteria then moves with their hosts, spreading worldwide as the copepods follow their source of food, which is some types of algae and plankton that grow vigorously when water temperatures rise. Algae growth is further intensified by the area in agricultural runoff and in sewage. The ability of the bacteria to live inside and connect to aquatic organisms enables them to survive in harsher environments (codeco, 2001).

Most people who get exposed cholera causing bacteria (*Vibrio cholerae*) sometimes do not become ill immediately or even know they have been infected, yet they shed the bacteria into the environment. Thus, contaminating fomites, household items and cause infection and risk to the public. On the other hand, some patients show symptoms as soon as they are infected. Infected persons often show mild or severe symptoms while some are asymptomatic (Akor, 2007).Only about one in 20 infected individual develop severe diarrhoea along with vomiting which result in dehydration. The following are the signs and symptoms of cholera;

    i.   Diarrhoea described as "rice water" in nature with pale milky appearance is one of the primary symptoms of cholera. The symptom usually start suddenly, half a day to five days after ingestion of bacteria. An untreated person with cholera may produce 10 to 20 litres of diarrhoea a day. Half of affected individual with severe cholera dies if the severe diarrhoea is untreated.

    ii.   Nausea and Vomiting, this happens in early stages of cholera, vomiting may persist for hours at a time.

    iii.  Dehydration; this maybe from mild to severe depending on how much body fluid have been lost. Severe dehydration indicates loss of 10% or more of total body weight.

Dehydration signs and symptoms includes extreme thirst, irritability, sunken eyes, lethargy, dry mucous membranes (dry mouth, nose, throat and eyelids), little or no urine output, dry and loss of skin elasticity, low blood pressure, an irregular heartbeat(arrhythmia),shock and muscle cramps

(http://www.mayoclinic.org/diseasesconditions/cholera/ca). Children in general develop the same symptoms as adult but they are particularly susceptible to low blood sugar (hypoglycemia) due to loss of fluid, which may cause: extreme drowsiness, convulsion and coma (Sack and Chaignat, 2006).

## 2.5   Risk factors associated with Cholera

Everybody is susceptible to cholera with the exception of babies who derive immunity from nursing mothers who have previous had cholera. But still some factors can make one more vulnerable to the disease. The following are some of the risk factors:

- **Contaminated water**: A number of environmental surveys carried out shows that drinking or use of contaminated water especially water from the rivers in rural areas are infected by *Vibrio cholerae* and clean  water that is collected in dirty containers that are not often covered increases the risk of cholera

  (http://www.kznhealth.gov.za/cholerareview.pdf)

- **Poor sanitary or unhygienic conditions:** Cholera outbreaks are mostly found where there are poor hygienic conditions (WHO, 2010; Emch, 2008). Such conditions are more common to refugee camps, natural disasters, war-torn areas, impoverished countries and areas devastated by famine. All these are suitable conditions for Vibrio cholera to thrive and the residents in such areas are at higher risk of contracting the disease.

- **Raw or uncooked food:** Eating uncooked or raw contaminated food like shellfish from water is known to harbour Vibrio cholera and it hugely increases the risk of cholera.

- **Household Exposure**: Living in the same household with people infected with cholera is a greater risk of contracting the disease.

- **Type O Blood**: people with type O blood group are more vulnerable to get cholera compared to people with other blood group types (Sack, Nair, Siddique, 2004).

- **Achlorhydria or hypochlorhydria**: These are the people with reduced or non-existent stomach acid. Vibrio cholera cannot survive in an acidic environment and often ordinary stomach acid serves as a first-line of defence against infection though people with low level of stomach acid for example, older adults, children and people

who take proton pump inhibitors or H-2 blockers lack this protection that makes them be at greater risk of cholera.

- **Malnutrition**: Malnourished children are more likely to develop severe cases of cholera if they are infected with the disease (WHO, 2010).

- **Gender**: Women are often more at risk of contracting cholera than men because they tend to be responsible for caring for those who are sick at home, and may not be aware of the necessary precautions to prevent transmission.

- **Cystic fibrosis**: People who are more resistant to cholera are those who are not affected by cystic fibrosis. Genetic deficiency of cystic fibrosis trans-membrane conductance regulator convey proteins, intercedes with bacteria binding to the gastro intestinal walls therefore reducing the risk of infection
  (http://www.informationhealthcare.com/cholera).

- Age: It is reported that children under the age of 5 years have high chance of getting cholera easily compared to adult. Worldwide 11 million cholera cases occur every year among children under the age of 5 years (Black et al, 2010).

## 2.6  Cholera transmission

Cholera transmission is closely linked to inadequate access to clean water, sanitation facilities, seasonal bloom of bacteria and contaminated food (Sack et al. 2004). Only about 20 per cent of those infected develop acute watery diarrhoea (AWD), and of these, between 10–20 per cent develop severe watery diarrhoea with vomiting. Contamination source may include excreta from cholera sufferers who shed the bacteria into the environment and these organisms find their way into waterways, groundwater or drinking water supplies. Drinking contaminated water and eating any foods washed in the contaminated water could cause a person to contract the infection. If treatment is not prompt and adequate, the loss of large amounts of fluid and salts through diarrhoea and vomiting can lead to severe dehydration and

death within hours (Lamond& Kinyanjui, 2012). The case fatality rate (CFR) if untreated may reach 30–50 per cent.

Since cholera is usually transmitted through faeces, contaminated water, hands or food, it remain an ever-present risk in many countries. New outbreaks can occur sporadically where water supply, sanitation, food safety, and hygiene are inadequate. The greatest risk occurs in over-populated communities, displaced populations and refugee settings, which are characterized by poor sanitation, unsafe drinking water and increased person- to-person contact, because the incubation period is short (two hours to five days), the number of cases can rise very rapidly (Cottle& Deedat, 2002).

Cholera may be transmitted through the following ways:

i. *Contaminated water and/or food*: Although seafood has been blamed in the past, this is a less common problem than with raw or undercooked food.

ii. *Person-to-person transmission*: Is the most common means of infection, mainly through direct contact with contaminated hands.

iii. *Corpses of cholera patients are highly infectious:* Through body fluids – physical contact during funeral ceremonies is also a major medium.

iv. *Cholera treatment centres:* Can serve as sources of contamination if hygiene or sanitation and isolation measures are inadequate.

## 2.7  Clinical diagnosis

The most recommended test for diagnosing cholera or the gold standard is the culture method. Patient stool sample is taken using a sterile cotton bud and placed on a plate containing thiosulphate citrate biles salts (TCBS) agar a medium selected to isolate the bacteria from diarrhoea. While incubating *Vibrio cholera* shows yellow clumps which is then analysed to

identify the exact strain of cholera. This diagnosis enables cholera to be differentiated from other protozoal, bacterial or viral that causes dysentery.

In situation where cholera is endemic, quick immune chromatographic dipstick testing is often available. It includes placing a dipstick strip into a stool sample and then reading the display lines. If two lines appears on the dipstick, cholera is confirmed but if its only one line it is ruled out. For the test to make a diagnosis it takes between 2 and 15 minutes. Another clinical diagnosis is by testing blood for antibodies against *Vibrio cholerae*. In spite of the fact that over 100 serogroups of *Vibrio cholerae* have been discovered only two are responsible for cholera epidemics, which are serogroup 01 and serogroup 0139.

## 2.8 Control and treatment of Cholera

Since the etiological agent, *Vibrio cholerae*, is transmitted via the faecal oral route, improving water quality, sanitation, and hygiene (WASH) is the cornerstone of a cholera control strategy (Clemens, 2011). However, improving WASH in low income countries requires sustained health effort and financial resources over many years, while oral cholera vaccines (OCVs) can produce an immediate impact for residents residing in endemic or epidemic areas (Wierzba et al.2015). Two OCVs that contain killed cholera Vibrio whole cells have been prequalified by World Health Organisation and are available for purchase by United Nations (UN) agencies (WHO, 2011). WHO recommends periodic mass vaccination campaigns with OCVs targeted at pre-school and school-aged children in endemic regions as pre-emptive and reactive vaccination strategies for the control of cholera epidemic (WHO, 2011).  However, OCVs have seldom been used for the control of cholera in either endemic or outbreak situations (Clemens, 2011). Policy makers require evidence of the feasibility and impact of vaccination programs before OCVs are widely used. Mass oral cholera vaccinations have been shown to be feasible in developing country settings such as Sudanese refugee camps in Uganda (Legros et al. 1999).

## 2.9 Economic and Socio-cultural Effects of Cholera

The existence of cholera in any nation has a negative effect on the individual and the country's prosperity due to its influence on economic and social decisions. Mortality as a result of cholera has a huge effect on national economies. The Gross Domestic Productivity (GDP) of a nation can decrease due to cholera infection of skilled workers, where either a worker is sacked by the employer or resigns as a result of the stigma due to health concerns or the worker may miss many days of productive work due to sickness. There may be costs associated with labour substitution, depending on the value of the activities from where the substituting labour is drawn (UNAIDS/WHO, 2004).

Cholera results in political, social and economic costs (Griffith et al, 2006). It obstructs economic development through various ways which includes: population growth, quality of life, workers productivity, savings and investment, premature mortality, medical cost and fertility.

Cholera causes a lot of health problems in many developing countries. It affected 3-5 million people worldwide (Pruss-ustun et al, 2008) and every year from 2010, a total number of 100,000 to 130,000 people have died from cholera (WHO, 2010). Cholera outbreaks continue to increase in numbers; in Haiti for example, cholera outbreak claimed more than 400, 000 lives in 2010 (WHO, 2010). In 2011, there were 589,854 total outbreaks including 7816 deaths which were reported to WHO from 58 countries (WHO, 2011). More numerous cases were unaccounted for due to fear of trade and travel sanctions and limitations in surveillance systems. Cholera continues to be a huge problem in Africa and various Asian countries (Albert et al., 1998).In 2002, Lanata et al. calculated, using the fraction cases of diarrhoea estimated to be caused by cholera, that globally 11 million cholera cases occur every year among children under the age of 5 years (Black et al, 2010).

From the year 2004 to 2008, more than 830,000 cases of cholera were reported to World Health Organisation (WHO), representing an increase of 24% in the number of reported cases within a period of five years. At the time of epidemics, mortality rate can be as high as 20% and in poor resource settings it ranges from 5% to 10% if there is unavailability of appropriate rehydration therapy. However, mortality rate can be less than 1% if infected persons are given proper and quick treatment. Someone who is not treated may produce 10 - 20 litres of diarrhoea per day with deadly outcome (WHO, 2010).

In South Africa, during 2001 cholera epidemic spread through the north eastern and eastern parts of South Africa. The most affected province was kwazulu natal and was the first case confirmed on 14[th of] august, 2000. *Vibrio cholera* E1 Tor ogawa was isolated and by April 5th the epidemic had brought about 82,275 cases to cholera treatment centres and caused 171 deaths. (http://www.kznhealth.gov.za/cholerareview.pdf).The highest mortality rate in South Africa was from 1981 to 1982 where 218 people were killed of cholera (http://www.hst.org.za/news/south-africa).

The outbreak of cholera causes suffering to humans, panic, disrupt the economic and social structure and can also hinder development in the affected communities. Some countries place restrictions on travels to affected countries (countries with cholera outbreaks) as preventive measures. In the same vein, import ban of certain food items are other preventive measures that have been practiced across geographical divide. In 1991, Peru experienced cholera outbreak and it was estimated that about $770 million was lost following the drastic fall in tourism and food export (www.who.int/topics/cholera/impact/en/). The economic cost of cholera epidemic may be direct and/or indirect costs, however is most cases they go in tandem. The direct costs accrue as a result of treatment and medication as well as all the ancillary household expenditures including transportation to the hospital, and physician charges. There is also the cost of community outreach, preventive prophylaxis administered to the contacts of sufferers and education programs to the community about the disease. The

indirect costs may include waiting and travelling time, production and income losses and treatment time associated with cholera.

In countries with cholera epidemic, adverse effects were recorded on the economies of such countries as man hour was lost due to morbidity. Impoverished families face hardships paying for medications and hospital stays. Productivity losses can be high, and this has been the case in the sub-Sahara African region which have been affected by cholera (http://www.globalization101.org/cholera).

# CHAPTER THREE
## RESEARCH   METHODOLOGY

### 3.0   Introduction

This chapter discusses about the research design, study population, sample design and sampling procedure as well as the proposed methods for data analysis.

### 3.1   Research design

The research design for this study was a quantitative and cross-sectional. The choice of the design survey was considered appropriate because it allows the verification of the risk factors associated with cholera. It also permits the assessment of the statistical significance of the risk factors using the methodology described below. The description through a multiple logistic regression model was preferred because the dependent variable is dichotomous (cholera infection through water sources and cholera infection through other non-water sources) and the independent variables are either continuous or categorical. The use of survey therefore was considered to be more appropriate in terms of resources, time and the overall objective of the study

The aspects of this design are briefly described as follows:

i.  **Quantitative design:** The quantitative designs are employed to determine the relationship among the risk factors of cholera and its prevalence in the area. This study compares outcomes of study variables and attempts to identify predictors of the differential outcomes among randomly selected study participants using objective research methods. A pre-designed and validated data collection instrument (questionnaire) will be used to obtain the study data, and data analysis is performed to make comparisons, assess correlations, and test statistical significance.

ii. **Non-experimental design:** In this study there is no new intervention introduced by the researcher, and all exposure, intervention, treatment, control and outcome variables to be used in the study are captured as they existed or happened during the time frame selected for the study.

iii. **Cross-sectional design:** In this study, information on emerging of cholera in relation to potential risk factors (exposures) are collected at one point in time. The study attempts to investigate associations between emerging and other risk factors among the study participants in the comparison groups. Therefore, the study design was chosen in order to measure the prevalence of a disease and the exposure status in a population at a particular point in time.

### 3.2  Ethical Consideration

In this study, only the individual patient information and clinical data was obtained during routine medical care provision and recorded in the hospital registers which were extracted. The extracted data pool includes only anonymous data without any personal identifiers. No additional information beyond what has already been gathered during the medical care of the study participants was collected. All data was held in confidentially and would not be used or shared outside the scope of the study.

The final version of the study protocol of ethical clearance was obtained from University of Fort Hare's Research Ethics Committee (UREC) prior to commencement of the study. Permission was secured also from the head of the selected health facilities for conducting the study prior to the commencement of the study and collection of data.

### 3.3 Informed Consent

A written permission was requested for from Goven Mbeki research department of University of Fort Hare, Alice campus to conduct the study in the area. The Hospitals

concerned were also informed and authorization and permission obtained before carrying out the study.

## 3.4 Study area



**Figure 3.1: Map of Raymond Mhlaba Local Municipality (Source: AfriGIS (Pty) Ltd. Google)**

Raymond Mhlaba Local Municipality (32.7901° S, 26.8329° E) was selected for this study. There are many towns and locations under this municipality, most of which are considered to be rural and peri-urban. There were cases of diarrhoea in the municipality. Outbreaks of diarrhoea diseases mostly occur as the result of low availability of drinking water and poor sanitation. Thus, this study focused on assessing the major factors predisposing inhabitants to cholera and waterborne diseases in this study area.

### 3.4.1  Study Population, Size and Frame

The target population for this study were patients diagnosed with *Vibrio cholerae* infection and those presenting clinical diarrhoea as out patients or in-patients within the health facilities of Raymond Mhlaba Local Municipality health facilities. Consequently, data source

were hospital and clinic registry records where patients with cholera may have been documented.

The sample size of 106 was selected using "The Epi Info 7" and the sampling frame was the list of all cholera patients enrolled (within the selected time frame for the study, which is the period of 2005 to 2015) in the selected health facilities at which the study was conducted. In this cross-sectional study, comparison was made in the treatment of cholera outcomes with environmental factors and risk factors. The Epi Info 7 was used to detect odd ratio of cholera exposed group and unexposed at 95% confidence interval.

### 3.4.2  Data collection

Data collection tool was a questionnaire (data guide) pre-designed to capture the relevant patients associated characteristics with cholera which were extracted from hospital records. There was no direct interaction with study participants. The questionnaire was used to extract data from the records contained in the hospital records for the cholera patients who were selected using simple random sampling technique from file records in the registers that were available at the eligible hospital where the study was carried out. The questionnaire was designed to capture baseline characteristics such as social and demographic characteristics of the selected patient's cholera status, medical care received, the outcomes and relevant time-lines with respect to starting and termination of medical care and/or follow-ups.

### 3.4.3  Data analysis

Cholera infection cases (patients) were characterised as water and non-water sources, these patients were further stratified based on gender; male and female. Afterwards, comparison was done using chi-squared tests for categorical variables. Socio-demographic and physio-clinical characteristics of patients included age, gender, nationality, race, level of education, marital status, body weight, height and body temperature. The questionnaire

similarly explored patients clinical data; patients complains, number of days with cholera, frequency of stool and vomiting per day, patients' stool characteristics, patients dehydration status, self-medication drug used before visiting clinic, laboratory tests, intervention provided by the physician, rehydration and antibiotics therapy used for patients, antibiotics prescribed. Other variables of interest included management of cholera diagnosis, sources of water supply, what was eaten in the last 24 hours, and the food eaten before having cholera and the patient's alcohol history. Others were mothers' breast feeding and vitamin A deficiency. Severe malnutrition was also assessed by the medical personnel.

The collected data was entered into an excel sheet and subsequently imported to SPSS statistical software for analysis. Data checking was conducted for any errors, missing or outliers and implausible results. Exploratory data analysis using tables and charts was done to further understand the data and detect errors and strange values. Descriptive statistics was used to summarize the data using tables and charts. Odds ratio was used to measure the association between cholera treatment outcomes among patients. Comparison of treatment outcomes was performed for risk factors of cholera for predictors and covariates groups using Analysis of Variance (ANOVA) to see if the different between the two groups are significant. Log-Logistic distribution was used as a reliability model for cholera treatment outcomes to look at associations between individual patient demographic and environmental factors in relation with their risk factors.

**3.5 Statistical Methods**

Poisson regression was used to assess the risk ratios (RR) at 95% confidence intervals (95% CI) for cholera risk factors with robust variance estimates to compensate for variance over-estimation. Variables with risk factors more than 5% missing data were excluded from analysis. A linear trend test was performed for ordinal variables with $\geq 4$ strata. Univariate risk factors associated with cholera with a p-value 0.10 and an RR of 0.9 or .1.1 were

variables for the multivariate model. We excluded risk factors with an RR between 0.9 and 1.1 were removed due to weak likelihood association, which might not be statistically significant solely because of large sample size. Collinearity among multivariate variables was assessed using variance inflation factors (VIF), with a VIF of $\geq 10$ indicating collinearity. If collinear variables were found, only the predictor judged to be more biologically plausible was considered for the multivariate model.

Multivariate regression model was built by adding and testing candidate predictors individually, in order of effect size. Continuous predictors were retained if the Wald test was significant (*p-value*< 0.05). Retention of categorical predictors was also dependent on a significant Wald test for at least one stratum. Since regression with robust standard errors does not provide log likelihoods, we could not perform likelihood ratio tests to compare models. After inclusion in the model, risk factors were not re-evaluated in subsequent model building steps. Analyses used two-sided significance levels and were performed with Stata/IC 11.2 (StataCorp LP, College Station, TX) and SPSS IBM, USA (SPSS Version 20).

### 3.5.1 Sensitivity and Specificity

The result of a diagnostic test is said to be positive if it states that the disease is present and negative if it states that the disease is absent. The accuracy of diagnostic tests is often assessed with two conditional probabilities: Given that a patient has a disease, the probability the diagnostic test is positive is called the sensitivity. Given that the patient does not have the disease, the probability the test is negative is called the specificity. Let $X$ denote the true state of a person, with categories 1 = cholera from water, 2 = cholera from non-water sources, and let $Y$ = outcome of diagnostic test, with categories 1 = positive, 2 = negative. Then we have:

$$sensitivity = P(Y = 1|X = 1) \text{ and } specificity = P(Y = 2|X = 2)$$

The higher the sensitivity and specificity, the better the diagnostic test.

### 3.5.2 Odds ratio

If an event "A" has probability $p\ (A)$ of occurring, then the odds is defined as:

$$Odds(A) = \frac{p(A)}{1-p(A)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.1)$$

This implies that:

$$p(A) = \frac{Odds(A)}{1+Odds(A)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.2)$$

Now, suppose that $X$ denotes the event that an individual is exposed to a risk of having a disease and that $D$ denotes the event that the individual has the disease. We denote the complementary events as $\bar{X}$ and $\bar{D}$. The odds of an individual contracting the cholera disease through water:

$$Odds(D/X) = \frac{p(D/X)}{1-p(D/X)}$$

And the odds of an individual not contracting the cholera disease through water shows that he is not exposed:

$$Odds(D/\bar{X}) = \frac{p(D/\bar{X})}{1-p(D/\bar{X})}$$

The odds ratio is:

$$OR = \frac{Odds(D/X)}{Odds(D/\bar{X})} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.3)$$

is a measure of the influence of exposure on subsequent disease.

Now consider the following contingency table of joint and marginal probabilities:

Table 3.1: Contingency table of joint and marginal probabilities

|  | $\bar{D}$ | $D$ |  |
|---|---|---|---|
| $\bar{X}$ | $\pi_{00}$ | $\pi_{01}$ | $\pi_{0.}$ |
| $X$ | $\pi_{10}$ | $\pi_{11}$ | $\pi_{1.}$ |
|  | $\pi_{.0}$ | $\pi_{.1}$ | $1$ |

In this notation,

$$p(D/X) = \frac{\pi_{11}}{\pi_{10} + \pi_{11}}$$

$$p(D/\bar{X}) = \frac{\pi_{01}}{\pi_{00} + \pi_{01}}$$

So that:

$$Odds(D/X) = \frac{\pi_{11}}{\pi_{10}} \qquad (3.4)$$

$$Odds(D/\bar{X}) = \frac{\pi_{01}}{\pi_{00}} \qquad (3.5)$$

The odds ratio is:

$$OR = \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}} \qquad (3.6)$$

In equation (3.4) it shows those who had cholera through water and (3.5) shows those who did not have cholera through water.

Therefore odd ratio equals to those that had cholera disease through water over those who did not have cholera through water.

### 3.5.3 Generalised linear models (GLMs)

The logistic regression model is a broad class model known as Generalized Linear Models (GLMs). GLMs likewise include linear regression, ANOVA, Poisson regression, etc. A General Linear Model specifies the relationship between a dependent (or response) variable Y, and a set of predictor variables, the X"s, so that we calculate fitted values

$$Y = \quad \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

In the equation $\beta_0$ is the regression coefficient for the intercept and the $\beta_i$'s values are the regression coefficients for variables $i = 1$ through k computed from the data. As an example we could predict cholera as a function of water and non-water, where water is a dummy variable

There are three components to a Generalized Linear Model (Mccullaghe, 1989)

1. **Random Component:** The *random component* of a Generalized Linear Model identifies the dependent/ response variable (cholera disease through water) and selects a probability distribution for it. Denote the observations on response variable by $(Y_1, Y_2, \ldots, Y_n)$.

2. **Systematic Component:** The *systematic component* of a GLM specifies the independent/explanatory variables which is the cholera disease that is not gotten through water such as poor sanitary condition, exposed food, income level etc. ($X_1$, $X_2$, $X_3$, ........, $X_n$ respectively). These enter linearly as predictors on the right-hand side of the model equation. That is, the systematic component specifies the variables that are the $\{x_j\}$ in the expression:

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (3.7)$$

This linear combination of the explanatory variables is called the *linear predictor*.

1. **Link Function:** Let us denote the expected value of Y, the mean of its probability distribution is given by

$$\mu = E(Y) \qquad (3.8)$$

The third component of a GLM, the *link function*, specifies a function $g(.)$ that relates $\mu$ to the linear predictor as:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (3.9)$$

g ($\mu$) is the dependent variable (cholera gotten through water) and $\beta_1 x_1 + \cdots + \beta_k x_k$ $are\ the\ preditors$

The link function $g(.)$ connects the random and systematic components.

## 3.6     Logistic regression model

### 3.6.1     Introduction

Logistic regression is a form of statistical modelling that is usually appropriate for categorical outcome variables. It relates the relationships between a categorical response variable and a set of explanatory variables which can be categorical or continuous. Given the nature of the problem at hand and of data used in this work, multiple logistic regression model is used. In general, this model is employed to model the outcomes of a categorical dependent variable. For categorical variables, it is inappropriate to use linear regression model because the response values are not measured on a ratio scale and the error terms are not normally distributed. In addition, the linear regression model can generate as predicted values any real number ranging from negative to positive infinity, whereas a categorical variable can only take on a limited number of discrete values within a specified range.

The crucial limitation of linear regression is that it cannot deal with dependent variables that are dichotomous (binary) and categorical. Many interesting variables in business and medical world are dichotomous. For example, cholera can be gotten through water or non-water(food/poor sanitation) consumers make a decision to buy or not buy, a product may pass or fail quality control; there are good or poor credit risks; an employee may be promoted or not, etc. A range of regression techniques have been developed for analysing data with categorical dependent variables. These techniques include logistic regression analysis. Logistic regression determines the impact of multiple independent variables (poor sanitary conditions, raw exposed food etc.) presented simultaneously to predict membership of one or other of the two dependent variable (water) categories. The logistic regression is the most popular multivariable method used in health science (Tetrault et al., 2008).

## 3.7  Binary Logistic regression with single independent variable

This estimates the probability that a characteristic is present (e.g. estimate probability of "success") given the values of explanatory variables, in this case a single independent variable; $\pi = Pr\ (Y = 1|X = x)$.

Many categorical response variables have only two categories. Let us denote a binary response variable (water) by $Yi$ and its two possible outcomes by 1 ("yes") and 0 ("no"). The distribution of $Y$ is specified by probabilities:

$P(Yi = 1) = \pi$ of 'Yes' and $P(Yi = 0) = (1 - \pi)$ of 'No'. Its mean is $E(Y) = \pi$.

X= ( $X_1,\ X_2,\ X_3,\ \ldots\ldots\ldots,\ X_k$ ) be a set of explanatory variables which can be continuous, discrete or a combination. $X_i$ is the observed value of the explanatory variables for observation $i$. but in this case we focus only on a single variable x which may be food or poor sanitary condition. For $k$ independent observations, the number of 'yes' has the binomial distribution specified by the index $n$ and parameter $\pi$. The formula was given in equation (3.2). Each binary observation is a binomial variate with $n = 1$. Although Generalized Linear Models can have multiple explanatory variables, for simplicity we introduce them using a single $x$.

As the value of $x$ changes, and $\pi$ is replaced by $\pi(x)$ when we want to describe its dependence on that value. Relationships between $\pi(x)$and $x$ are usually nonlinear rather than linear. In the logistic regression model, the random component for the (Yes, No) outcomes has a *binomial distribution.* The link function is the logit function$\ln\left[\frac{\pi}{1-\pi}\right] of\ \pi$, which is defined as the log of odds of success and symbolized by "logit ($\pi$)." Logistic regression models are often called *logit models*. Whereas $\pi$ is restricted to the range [0, 1] the logit can be any real number.

**The model:**

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.10)$$

By introducing exponential on both sides in equation **3.10**, will have:

$$\frac{\pi(x)}{1-\pi(x)} = e^{\beta_0 + \beta_1 x_1}$$

$$\pi(x) = e^{\beta_0 + \beta_1 x_1} - \pi(x)e^{\beta_0 + \beta_1 x_1}$$

$$\pi(x) + \pi(x)e^{\beta_0 + \beta_1 x_1} = e^{\beta_0 + \beta_1 x_1}$$

$$\pi(x)(1 + e^{\beta_0 + \beta_1 x_1}) = e^{\beta_0 + \beta_1 x_1}$$

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1}}{(1 + e^{\beta_0 + \beta_1 x_1})} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.11)$$

### 3.8.1  Interpretation of Parameter Estimates:

▪ $exp(\beta_0)$ = the odds that the characteristic is present in an observation $i$ when $X_i = 0$, .

▪ $exp(\beta_1)$ = for every unit increase in $X_{i1}$, the odds that the characteristic is present is multiplied by $exp(\beta_1)$. This is similar to simple linear regression but instead of additive change it is a multiplicative change in rate. This is an estimated *odds ratio*.

$exp(\beta\ 0 + \beta\ 1\ (x\ i1 + 1))exp(\beta\ 0 + \beta\ 1\ x\ i1) = exp(\beta\ 1\ )\ exp(\beta0 + \beta1(xi1+1))exp(\beta0 + \beta1xi1) = exp(\beta$

1)

In general, the logistic model stipulates that the effect of a covariate on the chance of "yes" is linear on the log-odds scale, or multiplicative on the odds scale.

▪ If $\beta_j > 0$, then $exp(\beta_j) > 1$, and the odds increase.

▪ If $\beta_j < 0$, then $exp(\beta_j) < 1$, and the odds decrease.

### 3.7.1 The logistic curve

Logistic regression fits a logistic curve to the relationship between *x* and *y*. Logistic curve has an S-shaped or sigmoid curve. A logistic curve starts with slow, linear growth, followed by exponential growth, which eventually assumes a stable rate.
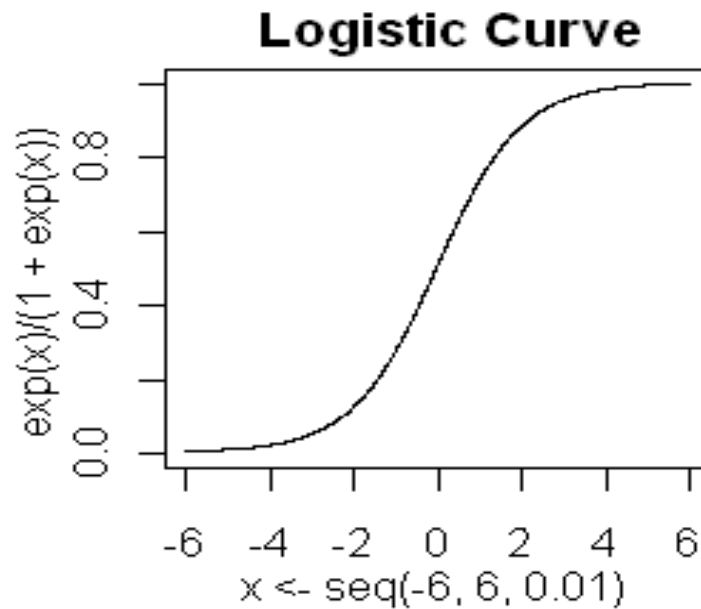


*Figure 3.2: Logistic regression curve*

### 3.7.2  Assumptions of logistic regression

2.  Logistic regression does not assume a linear relationship between the dependent and independent variables.

3.  The dependent variable must be dichotomous (2 categories).

4.  The independent variables are not normally distributed, nor linearly related, nor of equal variance within each group.

5.  The categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups.

6. Larger samples are needed than for linear regression because maximum likelihood coefficients are large sample estimates.

## 3.8 Multiple logistic regression

**The model**

Let us consider the general logistic regression model with multiple explanatory variables. Denote the $k$ predictors for a binary response $Y$ by $X_1, X_2, \dots, X_k$. We use $\pi(x)$ to represent the probability that $Y = 1$ for success and $1 - \pi(x)$ to represent the probability that $Y = 0$. These probabilities are written in the following form:

$$\pi(x) = P(Y = 1 / X_1, X_2, \dots, X_k)$$

$$1 - \pi(x) = P(Y = 0 / X_1, X_2, \dots, X_k)$$

$$logit(\pi(x)) = \ln \frac{P(Y = 1 / X_1, X_2, \dots, X_k)}{P(Y = 0 / X_1, X_2, \dots, X_k)}$$

$$\ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_k + \varepsilon$$

$$\therefore \quad \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + \varepsilon$$

$$\pi(x) = P(Y = 1 / X_1, X_2, \dots, X_k) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_j + \varepsilon}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_j + \varepsilon}}$$

The parameter $\beta_j$ refers to the effect of $X_j$ on the log odds that $Y = 1$, controlling the other predictor variables. For example, $\exp(\beta_j)$ is the multiplicative effect on the odds of a one-unit increase in $X_j$, at fixed levels of the other predictor variables.

### 3.8.1　The parameter estimations

The goal of logistic regression is to estimate the $K + 1$ unknown parameters $\beta = (\beta_0, \beta_1, .., \beta_k)$ . This is done with maximum likelihood estimation which entails finding the set of parameters for which the probability of the observed data is greatest. The maximum likelihood equation is derived from the binomial distribution of the dependent variable. For a set of observations in the data $(x_i; y_i)$ , the contribution to the likelihood function is $\pi(x_i)$ where $y_i = 1$ and $1 - \pi(x_i)$ where $y_i = 0$ . The following equation results for the contribution (call it $\varphi(x_i)$) to the likelihood function for the observation $(x_i; y_i)$:

$$\varphi(x_i) = \pi(x_i)^{y_t}[1 - \pi(x_i)]^{1-y_t}$$

The observations are assumed to be independent of each other so we can multiply their likelihood contributions to obtain the complete likelihood function. The result is given as:

$$l(\beta) = \prod_{i=1}^{k} \varphi(x_i) = \prod_{i=1}^{k} \pi(x_i)^{y_t}[1 - \pi(x_i)]^{1-y_t}$$

Generally the basis of multiple regressions is to learn more about the relationship between several independent/explanatory or predictor variables and a response/dependent variable. A multiple regression enables the simultaneous testing and modelling of multiple independent/explanatory variables.

In this study we can also apply multiple regression by taking the risk factors and water been the dependent variable, with poor sanitary condition, exposed food and others being the independent/predictor or explanatory variables.

### 3.8.2 Evaluation of a logistic regression model

Evaluations of logistic regression model include the overall evaluations, statistical test of individual predictors, goodness-of-fit statistics, and validations of predicted probabilities. Each one of these is illustrated next for the logistic regression model.

### 3.8.3 Overall model evaluations: The likelihood ratio test

A logistic model is said to provide a better fit to the data if it demonstrates an improvement over the intercept only model (also called the null model, which has no predictors). The likelihood ratio test for overall significance of the coefficients for the independent variables in the model is used. The test is based on the "G" statistic under the null hypothesis:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

and G statistic is calculated as:

$$G = \chi^2 = (-2 \ln likelihood\ of\ null\ model) - (-2 \ln likelihood\ of\ model\ with\ the\ variables)$$

$$G = -2ln \frac{likelihood\ of\ null\ model}{\log likelihood\ of\ model\ with\ the\ variables}$$

The distribution of "G" is a chi-square with k degree-of-freedom, where K is the number of covariates in the logistic regression equation. This is a measure of how well all of the independent variables affect the response variable (Bewick et al. 2005).

If the *p-value* for the overall model fit statistic is less than the conventional 0.05, then reject $H_0$ at $\alpha = 0.05$ and the conclusion will be that there is evidence that at least one of the independent variables contributes to the prediction of the outcome.

### 3.9 Statistical significance of individual regression coefficients: Wald test and confidence Interval

#### Wald statistic

To assess the significance of the logistic regression coefficients, the Wald statistic is used (Afifi et al., 2004; Bewick et al., 2005).The Wald statistic is calculated as:

$$W_j = \frac{\hat{\beta}^2_j}{\left[SE(\hat{\beta}_j)\right]^2}$$

Where $\hat{\beta}_j$ represents the estimated coefficient of $\beta$ and $SE(\hat{\beta}_j)$ denote the standard error. Under the null hypothesis $H_0: \beta_j = 0$, the quantity follows a chi-square distribution with one degree of freedom. If the estimated value of the slope is small and its estimated variability is large, then we cannot conclude that the slope is significantly different from zero and vice-versa (Afifi et al., 2004).

### 3.10 Confidence Interval

Odds ratio with 95% confidence interval (CI) can be used to test for the contribution of individual predictors (Katz, 1999). However, unlike the p value, the 95% CI does not report a measure's statistical significance. The 95% Confidence Interval is used to estimate the precision of the Odds Ratio (OR). A large Confidence Interval indicates a low level of precision of the Odds Ratio, whereas a small Confidence Interval indicates a higher precision of the Odds Ratio. This is computed as follows:

A 95% Confidence interval for $\hat{\beta}_j$, is given by: $\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)$

A 95% CI for log Odds Ratio $= \ln(OR) \pm 1.96 \times \{SE \ln(OR)\}$

Where the sample mean is log odds ratio, and the standard error of the log odds ratio.

$$A\ 95\%\ CI\ for\ Odds\ Ratio = e^{\ln(OR) \pm 1.96 \times \{SE \ln(OR)\}}$$

(Morris & Gardner, 1988).

### 3.11 Goodness-of-fit statistics: Hosmer-Lemeshow test

The Hosmer-Lemeshow test helps to examine whether the observed proportions of events are similar to the predicted probabilities of occurrence in subgroups of the model population. The Hosmer-Lemeshow test is assessed by dividing the predicted probabilities into deciles (10 groups based on percentile ranks) and then computing a Pearson Chi-square that compares the predicted to the observed frequencies in a 2-by-12 table.

The value of the test statistics is:

$$H = \sum_{p=1}^{12} \frac{(O_p - E_p)^2}{E_p}$$

In the above formula, $O_p$ and $E_p$ denote the observed events, and expected events for the $p^{th}$ risk deciles group respectively. The test statistic asymptotically follows a chi-square distribution with $8$ (number of groups minus two) degrees of freedom. Small values (with large value closer to 1) indicate a good fit to the data, therefore, good overall model fit. Large values (with p-value $< 0.05$) indicate a poor fit to the data.

### 3.12 Graphing prediction accuracy: Receiver Operating Characteristic (ROC) curve

One primary goal of performing logistic regression is to generate an equation that can reliably classify observations into one of two outcomes. The degree to which predictions agree with the data may be shown graphically by a receiver operating characteristic (ROC) curve. According to Hosmer & Lemeshow (2000), the ROC curve is a plot of sensitivity versus 1-specificity. Sensitivity is defined as the proportion of observations correctly classified as an event. Specificity is defined as the proportion of observations

correctly classified as non-event. Hence, 1-specificity is the proportion of observations misclassified as an event; which is also called the false positive fraction.

ROC (Receiver Operating Characteristic) analysis is used as a method for evaluation and comparison of classifiers (Ferri et al., 2002). The ROC gives complete description of classification accuracy as given by the area under the ROC curve. The model with a larger area below the ROC curve is considered as better model. Alternatively, the one with the greatest height on the ROC curve at a desirable probability cut-off should be chosen. In other words, the best model is the one associated with the greatest sensitivity and the lowest 1-specificity.

# CHAPTER FOUR

# DATA ANALYSIS AND RESULTS

## 4.0    Introduction

This chapter discusses data analysis results and interpretation. The data used for this study captured the following independent variables; age, gender, nationality, race, level of education, marital status, body weight, height and body temperature. Similarly, patients complains, number of days cholera symptom manifested, frequency of stooling and vomiting per day, patients' stool characteristics, patients dehydration status, self-medication and drugs used prior to clinic visitation, intervention provided, rehydration and antibiotics therapy. Other variables of interest included sources of water supply and the type of food eaten 24 hours prior to disease onset. The dependent variable of this study is the cholera disease acquired through consumption of *Vibrio cholerae* contaminated water.

## 4.1    Description of the study population

The median age of study population was 24.5 (IQR: 7.0-44.7) for all respondents with cholera and, the gender distribution were 50% females and 50%  males most of whom were South Africa citizens (75.5%); more than two-thirds of the patients were of black race (63.2%), indicating that the study area was a predominantly black community area. As far as education was concerned, about 75.5% of the respondents had at least matric level of education indicating a well-informed population. While 71% of the population were single, only 11% accounted for the married respondents. This large proportion of single mothers could have a great impact on the spread of cholera if they are not well-educated.

The diagnostic tests for patients stool and blood in search of *Vibrio cholerae* is shown in Figure 4.1. About 31.6% of the respondents had routine laboratory test through stool examination, 67.1% had stool culture for *Vibrio cholerae* and only 1.3% of the patients had

routine blood examination (Figure 4.1). This means that for most respondents, *Vibrio cholera* diagnosis was done through stool culture.
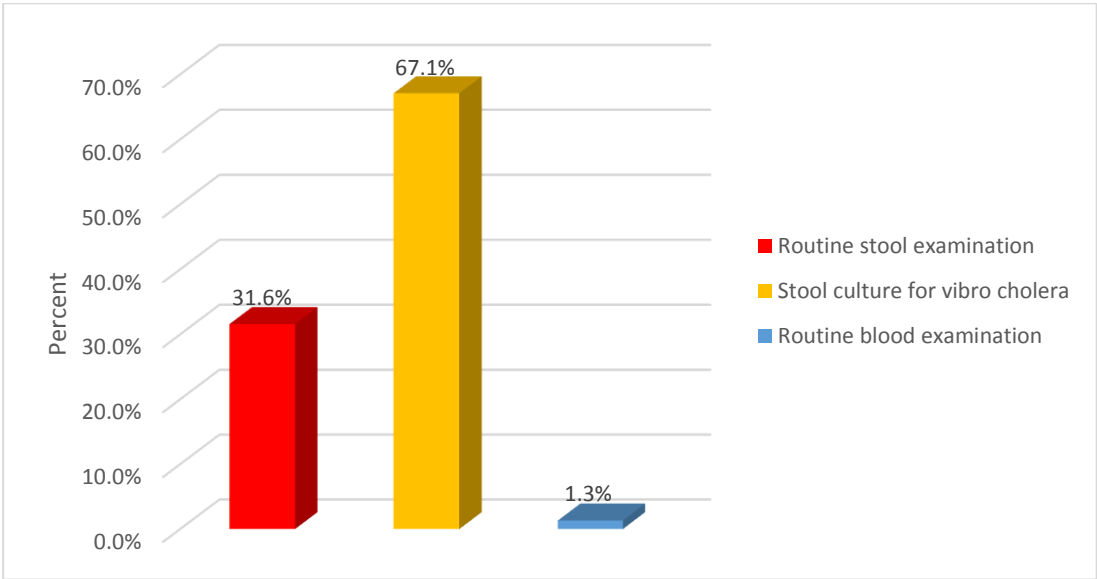


Figure 4.1: Diagnostic tests for stool and blood samples for *Vibrio cholerae*

Considering the age distribution of the study respondents, the median age is between 45 years for upper age range and 7 years for lower age range. There is no significant difference between male and female respondent with respect to cholera cases. Seventy six percent of the respondents were South Africans while 24% were foreign nationals. There were more cases of cholera among the blacks (63.2%) compared to other races. There was a high level of cholera cases among respondents with matric and lower levels of education (about 70%). There was high prevalence cases of cholera among single females (67%). In terms of water sources, a higher proportion of respondents used dam/river water (51.9%) and tap water (43%). Sixty eight percent of the respondents had history of using alcohol.

**Table 4.1: Descriptive statistics of the study participants**

| | | Frequency | Percent | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| **Age** | | Mean = 27.71 Median =24.5 (IQR:7-44.7) | | 1.861 | 28.47 | 36.46 |
| **Gender** | Male | 53 | 50.0 | 4.9 | 39.6 | 59.4 |
| | Female | 53 | 50.0 | 4.9 | 40.6 | 60.4 |
| **Nationality** | South African | 80 | 75.5 | 4.3 | 67.0 | 84.0 |
| | Other Nationals | 26 | 24.5 | 4.3 | 16.0 | 33.0 |
| **Race** | White | 3 | 2.8 | 1.6 | .0 | 6.6 |
| | Coloured | 36 | 34.0 | 4.6 | 24.5 | 43.4 |
| | Black | 67 | 63.2 | 4.7 | 53.8 | 72.6 |
| **Level of Education** | Grade 12 | 26 | 24.5 | 4.1 | 16.0 | 33.0 |
| | Matric | 48 | 45.3 | 4.9 | 35.8 | 54.7 |
| | Graduate | 23 | 21.7 | 4.1 | 14.2 | 30.2 |
| | Post graduate | 9 | 8.5 | 2.7 | 3.8 | 14.2 |
| **Marital status** | Single | 71 | 67.0 | 4.4 | 58.5 | 75.5 |
| | Married | 11 | 10.4 | 2.9 | 4.7 | 16.0 |
| | Divorced | 10 | 9.4 | 2.9 | 3.8 | 15.1 |
| | Widowed | 8 | 7.5 | 2.6 | 2.8 | 13.2 |
| | Separate | 6 | 5.7 | 2.2 | 1.9 | 10.4 |
| **Weight** | | Mean = 62.78 Median =72.50 | | 2.88 | 64.69 | 75.58 |
| **Height** | | Mean = 1.57 Median =1.69 | | 0.028 | 1.51 | 1.62 |
| **Temperature** | | Mean = 37.88 Median =37.90 | | 0.089 | 37.69 | 38.05 |
| **Water sources** | Tap | 45 | 42.5 | 4.8 | 34.0 | 51.9 |
| | Carrier/Tanker | 4 | 3.8 | 1.8 | .9 | 7.5 |
| | Dam/River | 55 | 51.9 | 4.8 | 41.5 | 61.3 |
| | Rain water | 2 | 1.9 | 1.3 | .0 | 4.7 |
| **Alcohol history** | Yes | 72 | 67.9 | 4.6 | 59.4 | 76.4 |
| | No | 34 | 32.1 | 4.6 | 23.6 | 40.6 |

## 4.2  Association between Water and Non-water Source Cases of Cholera in Socio-Demographics

From the tables 4.2 below, it indicated that age was not significantly associated with cholera cases among the respondents (p-value>0.05). However, it was found that race and water sources are both significantly associated with cholera (p-values = 0.007 and 0.005) respectively. The black respondents are seen to be the more vulnerable to cholera cases followed by coloured respondents in both water and non-water as the sources of cholera cases. As far as the water sources are concerned, Dam/river was found to be more significantly associated with cholera cases, followed by those respondents who use tap water and lastly those who use rain water.

Subsequently, respondent's gender, nationality, level of education, marital status, weight, height, temperature and alcohol history were not significantly related to cholera.

**Table 4.2:** **Association between Water and Non-water sources of Cholera Cases in socio-demographic characteristics**

| | | | sources | | P-value |
|---|---|---|---|---|---|
| | | | **Water** | **Non-water** | |
| Age | | Count | 65 | 42 | .208 |
| | | % | 60.7% | 39.3% | |
| Gender | Male | Count | 35 | 19 | |
| | | % | 32.7% | 17.8% | |
| | Female | Count | 30 | 23 | .384 |
| | | % | 28.0% | 21.5% | |
| Nationality | South African | Count | 50 | 31 | |
| | | % | 46.7% | 29.0% | |
| | Other national | Count | 15 | 11 | .714 |
| | | % | 14.0% | 10.3% | |
| Race | White | Count | 0 | 3 | |
| | | % | 0.0% | 2.8% | |
| | Coloured | Count | 21 | 15 | .007* |
| | | % | 19.6% | 14.0% | |
| | Black | Count | 44 | 24 | |
| | | % | 41.1% | 22.4% | |
| Level of Education | Grade 12 | Count | 17 | 9 | |
| | | % | 15.9% | 8.4% | |
| | Matric | Count | 28 | 20 | |
| | | % | 26.2% | 18.7% | |
| | Graduate | Count | 13 | 10 | .987 |
| | | % | 12.1% | 9.3% | |
| | Post grad | Count | 7 | 3 | |
| | | % | 6.5% | 2.8% | |
| Marital status | Single | Count | 44 | 28 | |
| | | % | 41.1% | 26.2% | |
| | Married | Count | 7 | 4 | |
| | | % | 6.5% | 3.7% | |
| | Divorced | Count | 5 | 5 | .827 |
| | | % | 4.7% | 4.7% | |
| | Widowed | Count | 6 | 2 | |
| | | % | 5.6% | 1.9% | |
| | Separate | Count | 3 | 3 | |
| | | % | 2.8% | 2.8% | |
| Weight | | Count | 65 | 42 | .502 |
| | | % | 60.7% | 39.3% | |
| Height | | Count | 64 | 42 | .682 |
| | | % | 60.4% | 39.6% | |
| Temperature | | Count | 65 | 42 | .155 |
| | | % | 60.7% | 39.3% | |
| Water sources | Tap | Count | 28 | 17 | |
| | | % | 26.2% | 15.9% | |
| | Carrier/Tanker | Count | 4 | 0 | |
| | | % | 3.7% | 0.0% | |
| | Dam/River | Count | 31 | 24 | .005* |
| | | % | 29.0% | 22.4% | |
| | Rain water | Count | 2 | 1 | |
| | | % | 1.9% | 0.9% | |
| Alcohol history | Yes | Count | 45 | 27 | |
| | | % | 42.1% | 25.2% | |
| | No | Count | 20 | 15 | .594 |
| | | % | 18.7% | 14.0% | |

*Statistically significant at P<0.05

## 4.3 Multicollinearity between the independent variables

If one variable is a perfect linear function of another in the model, standard errors become infinite and the solution to the model becomes indeterminate. To the extent that one independent is a near but not perfect linear function of another independent, the problem of multicollinearity will occur in logistic regression. As the independent variables increase in correlation with each other, the standard errors of the logit (effect) coefficients will become inflated. Multicollinearity does not change the estimates of the coefficients, only their reliability. To avoid the misleading results, we have used the Variance Inflation Factor (VIF) to check for multicollinearity between the independent variables. According to Robert M. (2007), the Variance Inflation Factor (VIF) and tolerance are both widely used measures of the degree of multicollinearity of the $i$th independent variable with the other independent variables in a regression model. Neter et al. (1989) stated that a maximum VIF value in excess of 10 is often taken as an indication that multicollinearity may be unduly influencing the estimates. Also, Hair et al. (1995) suggested that a VIF of less than 10 are indicative of inconsequential collinearity. Marquardt (1970) uses a VIF greater than 10 as a guideline for serious multicollinearity.

According to Kennedy (1992), a VIF greater than 10 indicates harmful collinearity. When the VIF reaches these threshold levels, researchers may feel compelled to reduce the collinearity by eliminating one or more variables from their analysis; combining two or more independent variables into a single index; resorting to a biased regression technique that can reduce the variance of the estimated regression coefficients because the VIF exceeds a threshold value (Belsley et al., 1980).

**Table 4.3: Checking for Multicollinearity by Variance Inflation Factor Coefficients[a]**

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | -4.126 | 2.237 | | -1.844 | .068 | | |
| Age | .004 | .005 | .180 | .862 | .391 | .212 | 4.717 |
| gender | .057 | .097 | .058 | .581 | .562 | .931 | 1.075 |
| nationality | .058 | .115 | .051 | .511 | .611 | .906 | 1.103 |
| Race | -.215 | .095 | -.239 | -2.252 | .027 | .818 | 1.223 |
| education | .069 | .063 | .125 | 1.106 | .271 | .717 | 1.395 |
| marital | -.056 | .057 | -.142 | -.984 | .328 | .442 | 2.263 |
| weight | -.007 | .005 | -.387 | -1.481 | .142 | .135 | 7.424 |
| height | .415 | .351 | .218 | 1.183 | .240 | .270 | 3.708 |
| temp | .141 | .061 | .280 | 2.328 | .022 | .635 | 1.574 |
| Water source | .049 | .049 | .100 | .994 | .323 | .906 | 1.103 |
| alcohol history | .079 | .154 | .076 | .517 | .607 | .427 | 2.340 |

a. Dependent Variable: sources

The tables above (tables 4.3) showed that we have no problem of multicollinearity among our independent variables, since in all cases, the VIF <10. Thus we may proceed with all our independent variables to fit the multiple logistic regression model.

## 4.4 Non-linearity between dependent and independent variables, and non-normality errors

The logistic regression does not assume a linear relationship between the dependents and the independents variables and the error terms are not assumed to be normally distributed. The following table is on the linearity between the dependent variable (sources of cholera) and the independent variables (age, gender, level of education, marital status, sources of water supply, place eaten in the last 24hrs and toilet characterisation).

The following are the Q-Q plots of standardized residuals

A



B



C



D

E



F



G



H
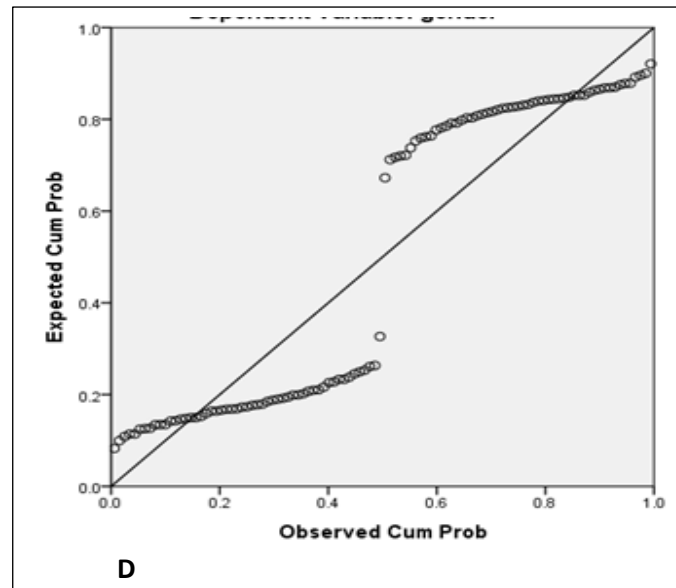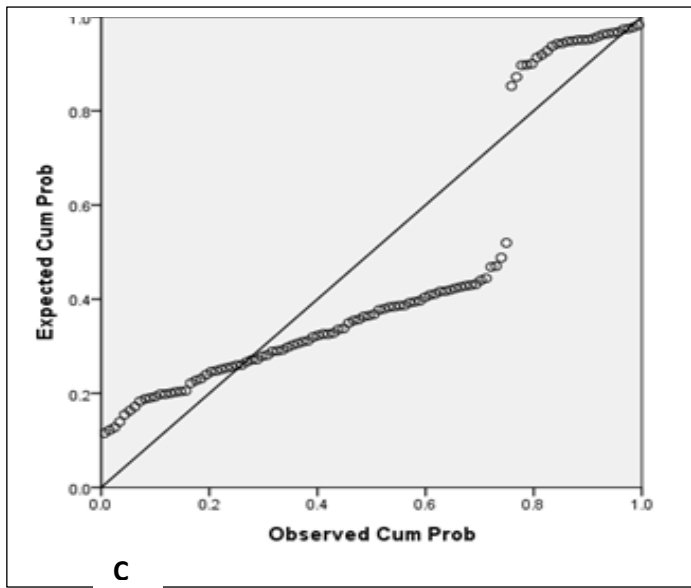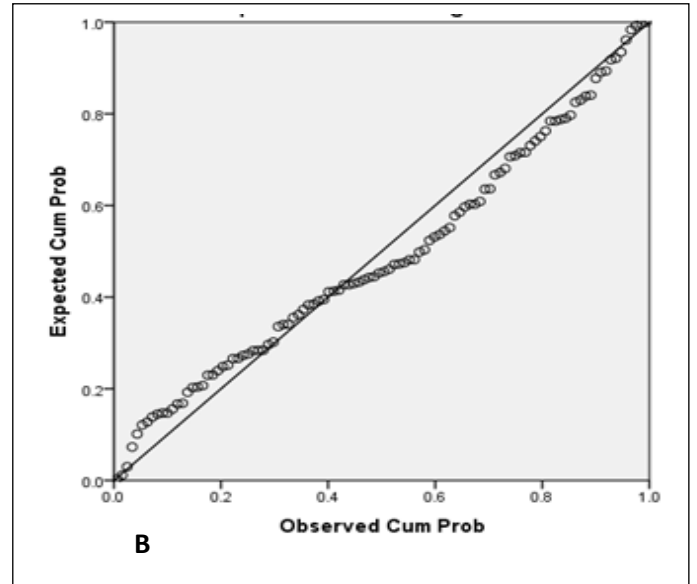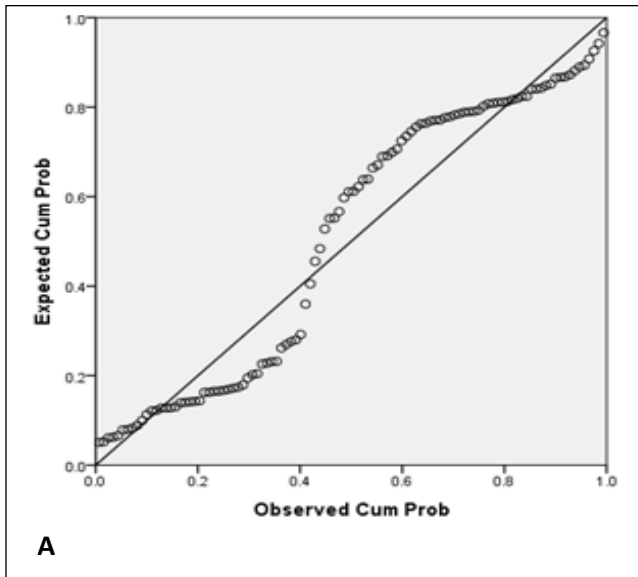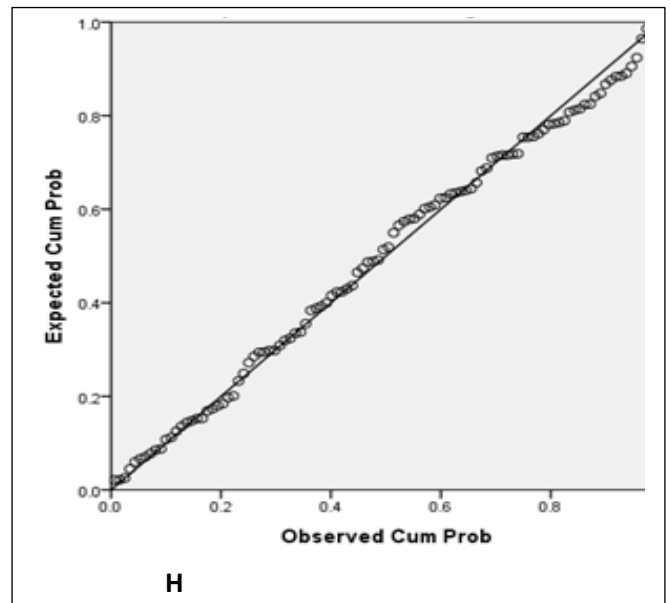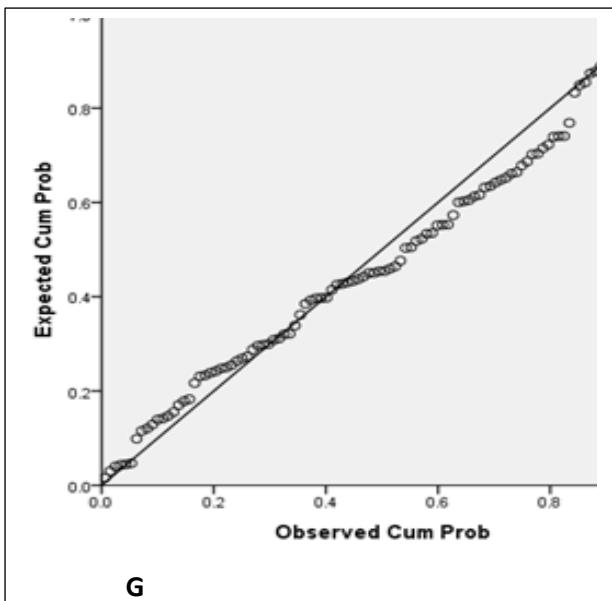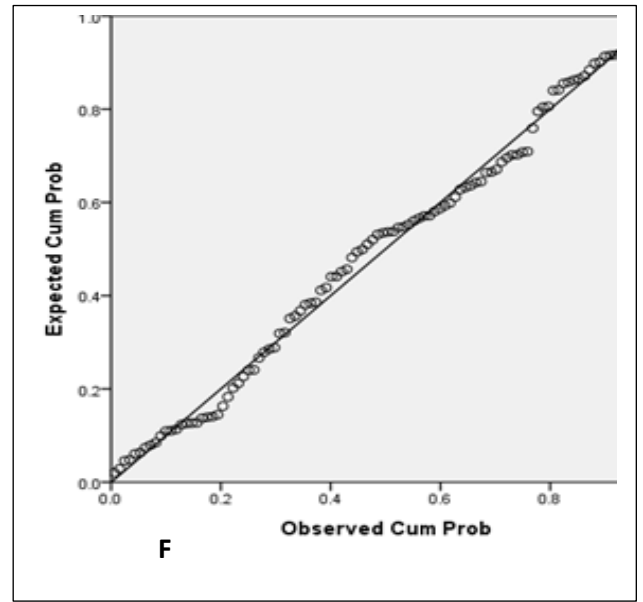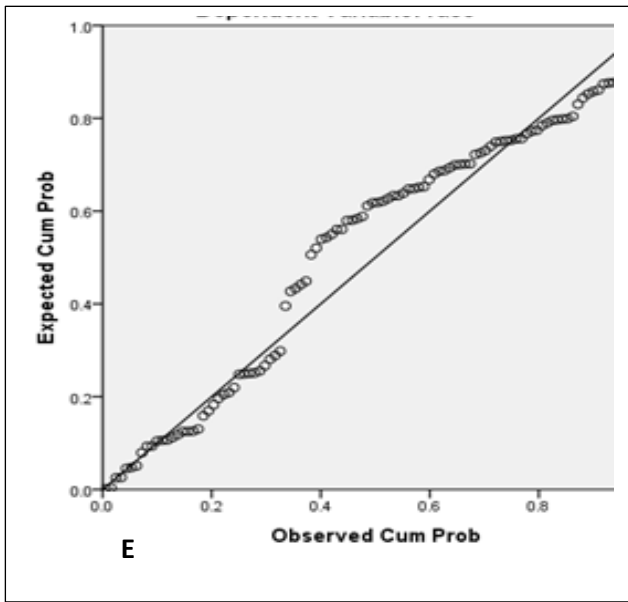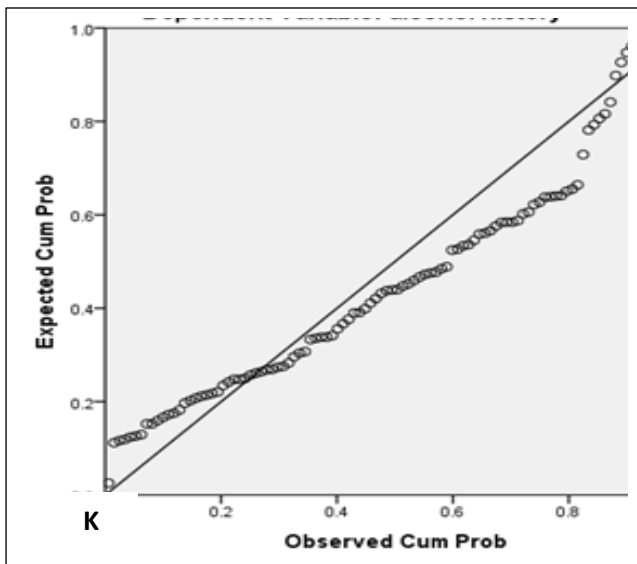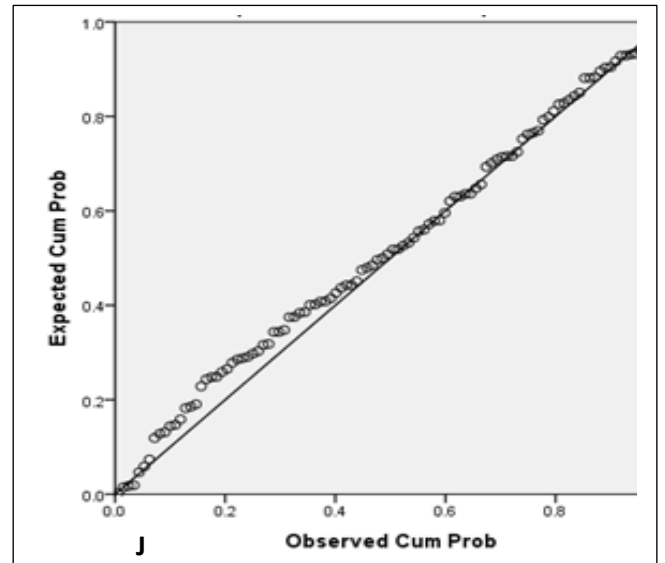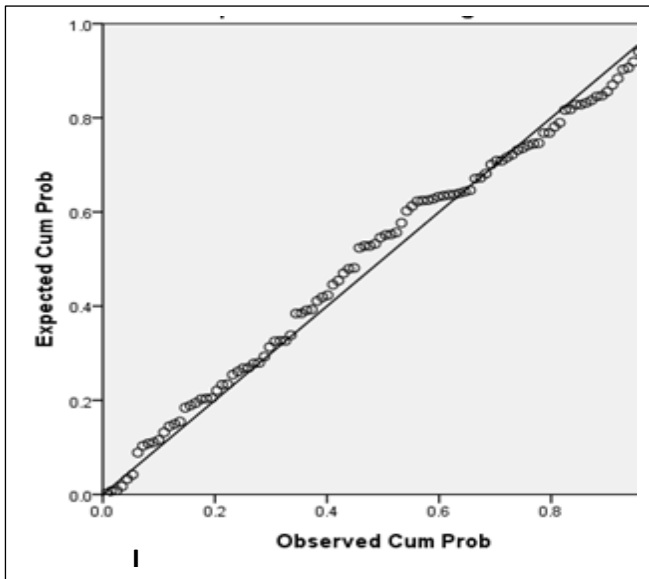
A= Dependent variable - Water; B= Independent variable - Age; C= Independent variable - Nationality; D = Independent variable - Gender; E = Independent variable - Race; F = Independent variable - Education; G = Independent variable - Marital Statues; H = Independent variable - Weight; I = Independent variable - Height; J = Independent variable - Temperature; K = Independent variable - Alcohol history

Figure 4.2: Q-Q PLOT of regression standardized Residuals of dependent and independent variables

It is observable from above Q- Q Plots of regression standardized residuals that errors are not normally distributed. Thus the logistic regression may be used instead of ordinary least squares regression.

## 4.5 Multiple logistic regression model fitting for risk factors

The data on variables associated with risk factors were all considered and no instances of collinearity among the assessed variables were found. In the bivariate model analysis, cholera risk increased with age and lower levels of education (P <0.05), when adjusted for a priori confounders and the other predictors in the final model (Table 4.4). Patients within the age range of 26-40 and 41-55 were found to have a significantly higher risk of cholera (2.20, 95% CI: 1.51, 4.22) and (1.13, 95% CI: 0.61, 2.01) compared to other age groups. In the case of race, the risk of cholera was two times higher among the blacks compared to the coloured (2.51, 95% CI: 1.52, 4.31). In terms of education, those with Metric and Grade 12 levels of education had a significantly higher risk of contracting cholera (2.50, 95% CI: 1.21, 4.29) and (1.43, 95% CI: 0.75, 3.51) compared to patients with higher levels of educational background, perhaps because they were better informed about how to deal with the disease and also to avoid its infection. Patients who were married and those who were divorced had an increased risk of cholera (1.71, 95% CI: 0.91, 3.56) and (2.51, 95% CI: 1.53, 4.25) respectively compared to those with other marital status, probably because they had more children and insufficient income to take care of their family health needs. Those whose source of water supply was from carrier/tank had less risk of cholera infection (1.71, 95% CI: 0.92, 3.62) compared to patients whose source of water supply was from Dam/River (2.61, 95% CI: 1.38, 4.25). This may be due to the fact that carrier/tank water was better treated and therefore safer compared to Dam/River water which is often dirty, untreated and contains many impurities. Party and Restaurant as eating places in the last 24 hours ( 1.33, 95% CI: 1.05, 3.88) and ( 1.12, 95% CI: 0.99, 1.82) posed higher risk of contracting cholera compared to meals eaten in the last 24 hours at Friend's house or Home. This may possibly be because some of the meals cooked in these Restaurants and in Parties are not hygienic and could influence spread of cholera. Patients who shared toilet facilities were at a higher risk of contracting cholera (0.91, 95% CI: 0.47, 1.62) compared to those who had private toilets. Those patients who did

not practice hand washing before eating or eating unwashed food were at an increased risk of contracting cholera (1.45, 95% CI: 0.88, 2.12) compared to those who practised hand washing and washing of food before eating.

In the Multivariate analyses carried out to identify the risk factors associated with cholera (Table 4.4) the following observations can be made. From the binary logistic regression model, cholera was significantly associated with age within the group of 26-40 years old, female patients, patients with lowest level of education, those having their water supply through Dam/River, patients that last ate food at home and party in the last 24 hours, patients who share toilet facilities and those who were not practicing hand washing. Multivariate analyses were also carried out to identify the risk factors not significantly associated with cholera. In the binary logistic regression model, cholera was not significantly associated with Nationality of the patients, race, marital status, patients with higher level of education and alcohol history.

The odds of cholera was 2.11 times higher in patients of age group 26-40 years of having cholera compared to patients in the age group above 55 years and those less than 26 years (AOR = 2.11, 95% CI (1.59, 4.21) and the odds of cholera was 1.53 times higher in female patients (AOR= 1.53, 95% CI (1.02, 2.54) compared to male patients. The odds of cholera was 1.55 times higher in patients with Grade 12 level of education (AOR= 1.55, 95% CI (0.98, 3.12) compared to patients who had more than Grade 12 education and the odds ratio of cholera was 2.11 higher among the patients who used Dam/River as the source of water supply (AOR= 2.11, 95% CI: 1.41, 4.25) compared to other patients that use other sources of water supply.

The multiple regression model for the risk factors:

$$Y \text{ (1=water, 0=non-water)} = \beta_0 + \beta_{1\,age} + \beta_{2\,gender} + \beta_{3\,education} + \beta_{4\,marital+}$$

$$+\beta_{5\,water\,source} + \beta_{6\,eaten} + \beta_{7\,toilet} + \beta_{8\,hand\,wash} + e$$

**Table 4.4: Bivariate and Multivariate analysis of risk factors of Cholera among**

| Variable | Odds Ratio (95% CI) | |
|---|---|---|
| | Bivariate (COR) | Multivariate (AOR) |
| **Age** | | |
| 0-10 | 1.11 (0.53, 2.27) | 1.15 (0.35, 2.44) |
| 11-25 | 1.55 (0.91, 3.77) | 1.60 (0.81, 3.38) |
| 26-40 | 2.20 (1.51, 4.22)* | 2.11 (1.59, 4.21)* |
| 41-55 | 1.13 (0.61, 2.01)* | 1.14(0.61, 2.14) |
| >55[+] | 1.00 | 1.00 |
| **Gender** | | |
| Male[+] | 1.00 | 1.00 |
| Female | 1.49 (0.78, 2.13) | 1.53 (1.02, 2.54)* |
| **Nationality** | | |
| South Africa | 1.38 (1.05, 2.08) | 1.13 (0.89, 1.88) |
| Other countries[+] | 1.00 | 1.00 |
| **Race** | | |
| White[+] | 1.00 | 1.00 |
| Coloured | 1.33 (0.75, 3.73) | 1.72 (1.66, 3.79) |
| Black | 2.51 (1.52, 4.31)* | 2.47( 1.82, 4.31) |
| **Level of Education** | | |
| Grade 12 | 1.43 (0.75, 3.51)* | 1.55 (0.98, 3.12)* |
| Matric | 2.50 (1.21,4.29)* | 2.29 ( 1.21, 4.10) |
| Graduate | 1.22 (0.76, 2.05) | 1.28 (0.82, 2.01) |
| Post graduate[+] | 1.00 | 1.00 |
| **Marital status** | | |
| Single | 1.11 (0.61, 2.48) | 1.11 (0.55, 2.44) |
| Married | 1.71 (0.91, 3.56)* | 1.52 (0.88, 3.61)* |
| Divorced | 2.51 (1.53, 4.25)* | 2.43 (1.42, 4.11) |
| Widowed | 1.12 (0.62,1.87) | 1.33 (0.72, 2.23) |
| Separated[+] | 1.00 | 1.00 |
| **Sources of water supply** | | |
| Tap[+] | 1.00 | 1.00 |
| Carrier/Tanker | 1.71 (0.92, 3.62)* | 1.55 (0.83, 3.64) |
| Dam/River | 2.61 (1.38, 4.25)* | 2.11 (1.41, 4.25)* |
| Rain water | 1.11 (0.66, 2.03) | 1.25 (0.99, 2.33) |
| **Place eaten in the last 24hrs** | | |
| Friend's house[+] | 1.00 | 1.00 |
| Home | 1.05 (1.45, 4.02)* | 1.31 (0.75, 3.45)* |
| Party | 1.33 (1.05, 3.88)* | 2.12 (1.35, 4.02)* |
| Restaurant | 1.12 (0.99,1.82)* | 1.14 (0.61, 2.14) |
| **Alcohol history** | | |
| Yes | 1.13 (0.88, 1.89) | 1.42 (0.95, 2.43) |
| No[+] | 1.00 | 1.00 |
| **Toilet characteristics** | | |
| **Privately use by only you[+]** | 1.00 | 1.00 |
| **Shared toilet** | 0.91 (0.47, 1.62)* | 1.11 (0.68, 2.12)* |
| **Hand washing practices** | | |
| **Yes[+]** | 1.00 | 1.00 |
| **No** | 1.45 (0.88, 2.12)* | 0.85 (0.51, 1.42)* |

*Statistically significant at P<0.05, +Reference group

Plotting sensitivity versus (1–specificity) over all possible cut-points is shown in the Figure below. The area under the ROC curve for the full model is 0.843 this is considered reasonable discrimination.
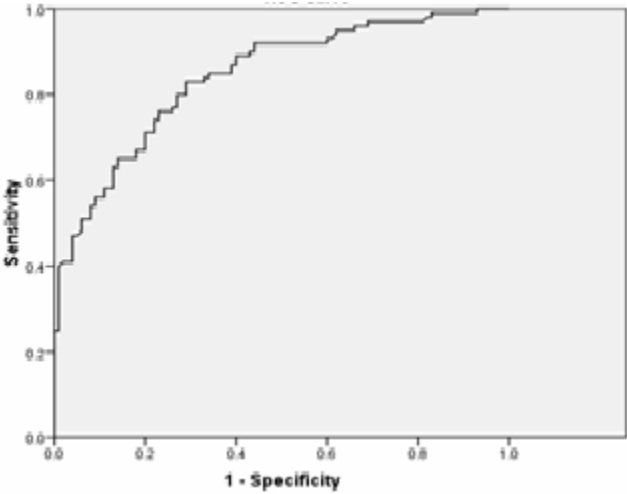


Figure 4.3:   Receiver Operating Characteristic (ROC) curve for the reduced model

Comparing the two models (bivariate and multivariate model), area under the ROC curve has become a particularly important measure for evaluating models' performance because it is the average sensitivity over all possible specificities. The larger the area, the better the model performs (Bradley, 1997).

We conclude that the multivariate model (which has the area under the ROC curve of 0.891 and its overall explanatory strength is 81.4%) fits better the data than the bivariate model with all predictor variables (which has the area under the ROC curve of 0.801 and its overall explanatory strength is 76.7%).

# CHAPTER FIVE

## CONCLUSION AND RECOMMENDATIONS

This study investigated risk factors associated with cholera disease in patients registered in a public hospital in the Raymond Mhlaba Local Municipality in the Eastern Cape Province of South Africa. Several factors emerged from both bivariate and multivariate logistic regression analysis that could be significantly associated with cholera from water and other non-water sources in the Local Municipality.

Some findings were common to both the water and non-water sources settings, while others were limited to one setting or the other. Age group between 26-40 years was strongly associated with cholera risk in both settings, with a two-fold increase in terms of risk compared to those of other age groups. This is in agreement with a study done in Kerssa district Eastern Ethiopia (Bezatu et al. 2013) and Vietnam (Rohmawati et al. 2012). This might be due to the fact that most of people in this age group are more exposed to a variety of factors that cause cholera compared to the other age groups.

Other socio-demographic factors also had strong impact as risk factors causing cholera from water and non-water sources settings: increasing level of education was associated with a decrease in the risk of cholera infection, probably because higher education results in better information about how to take care of oneself and family to avoid cholera infection and other risk factors; married and divorced patients were observed to have a higher risk of cholera infection. This could possibly be due to economic difficulties if they had more children than they could comfortably take care of especially if they were not economically empowered. The strong association between water sources and sanitation variables as a major cause of cholera is an indication of the importance of water in cholera disease transmission and management.

The increased risk associated with eating with a family member who has cholera in the last 24hours has also been found in studies of non-water cholera (Sobel et al. 2004). In our

study, the increased risk is likely due to shared primary exposures as well as genetic/familial susceptibility (Rahman et al. 2009) and secondary person-to-person transmission through environmental contamination (Giebultowicz et al. 2011).

However, since cholera infection confers natural immunity (Ali et al. 2011), it is unlikely that an individual would contribute more than one cholera case to our study. Nonetheless, we cannot rule out the possibility that a patient classified in this study as having non-water cholera might have had cholera in the past. This possible misclassification might have led to over or underestimation of associations. We were also unable to assess family clustering of cholera cases. Though clustering could lead to violations of underlying independent observation assumptions. In addition, antibiotic use prior to hospitalization could not be assessed. This could have skewed the data, since antibiotic treatment is highly efficacious. Despite these limitations, systematic sampling, and expert laboratory diagnosis of V. cholera in the hospitals where the patients visited, careful and thorough extraction of recorded data from the hospital records and accurate modelling and in-depth statistical analysis are strengths of this study, as is the fact that our reference group is comprised of hospital patients with other causes of cholera.

In conclusion, we report that increasing age, gender, level of education, marital status, sources of water supply, place of last eating 24 hours before cholera sickness, toilet facilities and hand washing practices are key correlates of risk factors associated with cholera infection among patients in Raymond Mhlaba Local Municipality in the Eastern Cape, South Africa. The strong association between water and sanitation amenities highlights the need for a more thorough assessment of potential waterborne exposures and treatment of water sources to alleviate risks of contracting cholera. Continued attention should be directed to promotion of breastfeeding for young children as a source of clean nutrition, female education, securing viable livelihoods, and provision of safe water sources. Finally, the risk faced by family

members of cholera cases may warrant renewed research regarding the use of targeted chemoprophylaxis in endemic rural settings.

**Further Research**

The study was only carried out only in one Municipality in the Eastern Cape, South Africa, called Raymond Mhlaba Local Municipality. Cholera is still a very dangerous disease in many rural and peri-urban areas in the country where people live in congested communities and Municipalities with very little water and lack of proper toilets. More studies need to be done in many communities on a large scale to eradicate this disease. This will equip administrators, managers and community leaders with tailor made solutions to alleviate this disease in their communities.

The sample size used in this study was rather limited due to financial constraints and time. Further studies should explore the possibility of not only covering more municipalities but also using large sample sizes as this will allow for more variables to be examined and proper conclusions and decision made to help the communities.

# References

Adagbada, A.O. et al., 2012. Cholera epidemiology in nigeria: An overview. *Pan African Medical Journal*, 12(1), pp.1–12.

Adeneye AK, 2010. *Risk factor associated with cholera outbreak* in Bauchi and Gombe State in north east Nigeria

Afifi, A., Clark, V. A., & May, S., (2004). *Computer- Aided Multivariate Analysis*. Fourth Edition, Champman and Hall/CRC.

Akor AN (2007).*Handbook on* Management *of Tropical Disease in Developing Countries.A Guide to the* Diagnosis *and Treatment of Common Diseases in the Tropics.*Bamise Printing Press Lokoja, Nigeria

Alexander K. and Kirscher (2008). Rapid Growth of Planktonic Vibrio cholerae non 01 and non 0139 Strains in Large Alkaline in Austria: Dissolve Organic Carbon Quality. Applied and Environmental Microbiology; 74:2004-2015

Ali, M. et al., 2011. Natural cholera infection-derived immunity in an endemic setting.*Journal of Infectious Diseases*, 204(6), pp.912–918.

Ali M, Nelson AR, Lopez AL, Sack DA. Updated *global burden of cholera in endemic countries*.PLoSNegl Trop Dis 2015; 9:e0003832.

Aryal, K. et al., 2012. Environmental Burden of Diarrhoeal Diseases due to Unsafe Water Supply and Poor Sanitation Coverage in Nepal. *J Nepal Health Res Counc*, 10(21), pp.125–129.

Assessment, R. et al., Economic Survey In The Eastern Cape Chapter Six : Providing Water & Sanitation.

Belsley, D. A., Kuh, E. &Welsch, R. E. (1980).*Regression Diagnostics*: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Bewick, V., Cheek, L., & Ball, J. (2005).*Statistics review 14: Logistic regression*. Critical Care

(London, England), 9(1), 112-118. http://dx.doi.org/ 10.1186/cc3045.

Bezatu, M., Berhane, Y. & Worku, A., 2013. *Prevalence of cholera and associated risk factors among children under-five years of age in Eastern Ethiopia*: A cross-sectional study. *Open Journal of Preventive Medicine*, 3(7), pp.446–453.

Biddanda BA, Cotner JB (2002). *Love handles in aquatic ecosystems: Role of dissolved organic carbon drawdown, resuspended sediments and terrigenous inputs in the carbon balance of Lake Michigan*. Ecosystems.Forthcoming.

Black RE, Morris SS, Bryce J.*Where and why are 10 million children dying every year*? Lancet. 2003; 361(9376):2226–2234. [PubMed]

Borroto, R.J. & Martinez-Piedra, R., 2000. *Geographical patterns of cholera* in Mexico, 1991-1996. *International Journal of Epidemiology*, 29(4), pp.764–772. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=lhh&AN=20002016391&site=e host-live%5Cnhttp://ije.oxfordjournals.org/content/29/4/764.full.pdf.

Bouma, M.J. & Pascual, M., 2001. *Seasonal and interannual cycles of endemic cholera in Bengal* 1891-1940 in relation to climate and geography. *Hydrobiologia*, 460, pp.147–156.

Bradley, A. P. (1997). *The use of the area under the roc curve in the evaluation of machine learning algorithms.* Pattern Recognition. Jul; 30(7): pp.1145-59. 72 .

Brian van Wyk, 2012*. Research design and methods Part I.*

Camilo, et al., 2001.*Cholera outbreak* in southern Tanzania risk factors and patterns of transmission

Cash RA, Music SI, Libonati JP, Snyder MJJ, Wenzel RP, Hornick RB (1974*). Response of man to infection with Vibrio cholerae.*I. Clinical, serological, and bacteriologic responses to known inoculum. *J. Infect. Dis.* **129**, 45–52

Cash, B.A. et al., 2010. Disentangling the impact of ENSO and Indian ocean variability on the regional climate of Bangladesh*: Implications for cholera risk.* Journal of Climate,23(10), pp.2817–2831.

Codeço TC, Lele S, Pascual M, Bouma M, KO, A.I, (2008). *A stochastic model for ecological*

*systems with strong nonlinear response to environmental drivers: application to two water-borne diseases. J R. Soc. Interface 5 (19), 247–252.*

Collins, A.E., 2003. Vulnerability to coastal cholera ecology. *Social Science and Medicine*, 57(8), pp.1397–1407.

Connor, K.A.O. et al., 2011. *Risk Factors Early.* , 17(11), pp.2010–2012.

Constantin de Magny, G. et al., 2008. Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), pp.17676–17681.

Cottle, E. & Deedat, H., 2002. *The Cholera Outbreak : A 2000-2002 case study of the source of the outbreak in the Madlebe Tribal Authority areas uThungulu Region, KwaZulu-Natal, South Africa*,

Emch, M. et al., 2008. Local environmental predictors of cholera in Bangladesh and Vietnam. *American Journal of Tropical Medicine and Hygiene*, 78(5), pp.823–832.

Faruque SM, Albert MJ, Mekalanos JJ. *Epidemiology, genetics and ecology of toxigenic Vibrio cholerae. Microbiology and Molecular Biology Reviews.1998; 62:1301–1314.*

February, R., Cholera, S. & During, H., 2014. Fact sheet N°107 Reviewed February 2014. , (February).

Ferri C., Flach P., Hernandez-Orallo J. (2002). *Learning Decision Trees Using the Area under the ROC Curve.*Nineteenth International Conference on Machine Learning (ICML 2002); Morgan Kaufmann; pp. 46-139.

Fleming, G., Merwe, M. Van Der & McFerren, G., 2007. Fuzzy expert systems and GIS for cholera health risk prediction in southern Africa. *Environmental Modelling & Software*, 22(4), pp.442–448.

Giebultowicz, S. et al., 2011. A comparison of spatial and social clustering of cholera in Matlab, Bangladesh. *Health and Place*, 17(2), pp.490–497.

Gil, A.I. et al., 2004. Occurrence and distribution of Vibrio cholerae in the coastal

environment of Peru. *Environmental Microbiology*, 6(7), pp.699–706.

Griffith DC, Kelly-Hope LA, Miller MA. Review of reported cholera outbreaks worldwide, 1995–2005. *Am J Trop Med Hyg*2006; 75: 973-7

Hair, J.F Jr., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). *Multivariate data Analysis*, 3rd edition .Macmillan.New York.

Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. Cholera. Lancet 2012;379:2466–76.

Hays, JN (2005). "*Epidemics and pandemics: their impacts on human history*". *P.347.ISBN 1-85109-658-2*

HoquemonjurA, 2003. *An investigation into risk factors associated with the cholera epidemic in kwazulu- natal during 2000*

Hosmer, D. W., &Lemeshow, S. (2000). *Applied logistic regression* (2nd Ed.). New York, NY: John Wiley & Sons Inc. http://www.choleraandthethames.co.uk/cholera-in-london/origins-of-cholera

http://www.dwaf.gov.za/WFGD/documents/WfGDv6http://www.health24.com/Medical/Cholera/Recent-cholera-outbreaks/Recent-cholera-outbreaks-in-SA-20120721

http://www.hst.org.za/publications/cholera-outbreak-2000-2002-case-study-source-outbreak-madlebe-tribal-authority-areas-ut

http://www.hst.org.za/sites/default/files/chap18_03.pdf

http://www.mayoclinic.org/diseasesconditions/cholera/basics/causes/con-     20031469

http://www.informahealthcare.com/cholerahttp://www.kznhealth.gov.za/cholerareview.pdf

http://www.hst.org.za/news/south-africa-asks-

http://www.merriam-webster.com/dictionary/epidemiology  help-cholera-outbreak-thousands-are-ill-and-59-are-dead

http://www.who.int/topics/epidemiology/en/

https://www.SABCNewsonline/posts/10152142053906543

http://www.who.int/topics/cholera/impact/en/

http://www.globalization101.org/cholera

Igbinosa, E.O., Philosophy, D.O.F. & Science, F.O.F., 2010. *Surveillance Of Invasive Vibrio Species In Discharged Aqueous Effluents Of Wastewater Treatment Plants In The Eastern Cape Province Of South Africa.*

Isham V, Medley G, eds (1996)., *'Models for Infectious Human diseases*, Cambridge University Press, Cambridge, UK.

Jutla, A. et al., 2013. A water marker monitored by satellites to predict seasonal endemic cholera. *Remote Sens Lett*, 4(8), pp.822–831.

Kaper JB, Morris JG, Jr, Levine MM (1995. Cholera.*ClinMicrobiol Rev. 1995;* 8:48–86.

Kanungo, S. et al., 2010. Cholera in India: An analysis of reports, 1997-2006. *Bulletin of the World Health Organization*, 88(3), pp.185–191.

Katz, M. H., (1999). *Multivariable analysis: A practical guide for clinicians*. Cambridge: Cambridge University Pres, London.

Kennedy, P. (1992). *A guide to Econometrics*. Oxford: Blackwell, London.

Koch, R (1884). *An address on cholera and its bacillus*. British Medical Journal: 403–407.

Lamond, E. & Kinyanjui, J., 2012. *Cholera Outbreak Guidelines Preparedness , Prevention And Control*. , (June).

Lobitz, B. et al., 2000. Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), pp.1438–1443.

López-Gigosos, R.M. et al., 2011. Vaccination strategies to combat an infectious globe: oral cholera vaccines. *Journal of global infectious diseases*, 3(1), pp.56–62.

Manager, M. et al., 2005. *The Cholera Epidemic Of 2000 / 2001 In Kwazulu-Natal* : Implications For Health Promotion And Education Zeleke Worku. , 10(4), pp.66–74.

Mendelsohn, J. & Dawson, T., 2008. Climate and cholera in KwaZulu-Natal, South Africa: The role of environmental factors and implications for epidemic preparedness. *International Journal of Hygiene and Environmental Health*, 211(1–2), pp.156–162.

Ministère de la Santé Publiqueet de la Population (2014) Rapport journalier MSPP du 07 Janvier 2014. Port-au-Prince: Ministère de la Santé Publiqueet de la Population. Available:

http://mspp.gouv.ht/site/downloads/Rapport%20Web_07.01_Avec_Courbes_Departementales.pdf

Mintz, E.D. & Tauxe, R. V., 2013. Cholera in Africa: A closer look and a time for action. *Journal of Infectious Diseases*, 208(1), pp.4–7.

Morris, J. A., & Gardner, M. J. (1988). *Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates*. British Medical Journal (Clinical Research Ed.), 296(6632), 1313-1316.

McCullaghe, P. and Nelder, JA. (1989). Second Edition: *Generalized Linear Models*

Neter, J., Wasserman, W. & Kutner, M.H. (1989).*Applied linear regression models*. Homewood, IL: Irwin.

Nsagha et al., 2015. Assessing the risk factors of cholera epidemic in Buea Health District of cameroon.2015;15;1128,bmc public health.

Obeng ,C (2015). *Deterministic sirb cholera model with logistic approach in dormaa-ahenkro,Ghana.*

Pokhrel, D. & Viraraghavan, T., 2004. Diarrhoeal diseases in Nepal vis-a ` -vis water supply and sanitation status Damodar Pokhrel and Thiruvenkatachari Viraraghavan. *Journal of Water and Health*, 2(2), pp.71–81.

Prüss-Üstün A, Bos R, Gore F, Bartram J.*Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health*. Geneva: World Health Organization; 2008.

Qadri, F. et al., 2005. Enterotoxigenic Escherichia coli and Vibrio cholerae cholera, Bangladesh, 2004. *Emerging Infectious Diseases*, 11(7), pp.1104–1107.

Quick, R.E. et al., 1996. Using a knowledge, attitudes and practices survey to supplement findings of an outbreak investigation: cholera prevention measures during the 1991 epidemic in Peru. *International journal of epidemiology*, 25(4), pp.872–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8921469.

Rahman, K.M. et al., 2009. Familial aggregation of vibrio cholerae-associated infection in Matlab, Bangladesh. *Journal of Health, Population and Nutrition*, 27(6), pp.733–738.

Rajasekar, S., & Philominathan, P. (2013). *Research methodology,* 1–53.

Reidl J, Klose KE (June 2002). "Vibrio cholerae and cholera: out of the water and into the host". *FEMS Microbiol. Rev.* **26** (2): 125–39.

Robert, M. (2007). *A Caution Regarding Rules of Thumbfor Variance Inflation Factors* Tetrault, J. M., Sauler, M., Wells, C. K., & Concato, J. (2008). *Reporting of multivariable methods in the medical literature*. Journal of Investigative Medicine, 56.

Rohmawati, N., Panza, A. & Lertmaharit, S., 2012. Factors Associated With Cholera Among Children Under Five Years Of Age In Banten Province , Indonesia. *Journal of Health Research*, 26(1), pp.31–34.

Ryan KJ, Ray CG (editors) (2004). *Sherris Medical Microbiology (4th ed.). McGraw Hill. pp. 376–7. ISBN 0-8385-852*

Sack DA, Sack RB, Chaignat CL (August 2006). "Getting serious about cholera". *N. Engl. J. Med.* **355** (7): 649–51. doi:10.1056/NEJMp068144.PMID 16914700.

Shimada TE (1993). Outbreaks of Vibrio cholerae non-O1 in India and Bangladesh. Lancet, 341:1347.

Sobel, J. et al., 2004. Pathogen-specific risk factors and protective factors for acute choleral illness in children aged 12-59 months in Sao Paulo, Brazil. *Clin Infect Dis*, 38(11), pp.1545–1551.

Tamang, M.D. et al., 2005. An outbreak of El Tor cholera in Kavre district, Nepal.

*Kathmandu University Medical Journal*, 3 no. 2(10), pp.138–142.

The, I., Nations, U. & Declaration, M., 2015. Literature review 2.1.

Thomas, et al., 2015.*Risk factors for sustained cholera transmission*, Juba County southern Sudan 2014

Wahed, T. et al., 2013. Knowledge of, attitudes toward, and preventive practices relating to cholera and oral cholera vaccine among urban high-risk groups: findings of a cross-sectional study in Dhaka, Bangladesh. *BMC public health*, 13(1), p.242.

WHO (1999): *Cholera in 1998. Weekly Epidemiological Recort, 74:257–263*

WHO. (2004). *Communicable Disease Toolkit Sierra Leone: WHO/CDS/2004.25.*

WHO/Cholera Fact 2010.

WHO (2011). *International travel and Health 2011*. Geneva: World Health Organization

World Health Organization; UN-Water, 2014. *Investing in water and sanitation: Increase access, reducing inequalities*,

Xu, M., Kan, B. & Wang, D., 2015. Identifying environmental risk factors of cholera in a coastal area with geospatial technologies. *International Journal of Environmental Research and Public Health*, 12(1), pp.354–370.

Yamai S (1997). Distribution of serogroups of Vibrio cholerae non-O1 nonO139 with specific reference to their ability to produce cholera toxin and addition of novel serogroups. Journal of the Japanese Association for Infections Diseases, 71:1037–1045

Zuckerman, J.N., Rombo, L. & Fisch, A., 2007. The true burden and risk of cholera: implications for prevention and control. *Lancet Infectious Diseases*, 7(8), pp.521–530.

APPENDIX 1
ETHICAL CLEARANCE CERTIFICATE

University of Fort Hare
*Together in Excellence*

## ETHICAL CLEARANCE CERTIFICATE
## REC-270710-028-RA Level 01

Certificate Reference Number:     QIN101SOSU01

Project title:                    **Determinant of risk factors associated with cholera and its effects: A case study Raymond Mhlaba Local Municipality of the Eastern Cape, South Africa.**

Nature of Project:                Masters in Statistics

Principal Researcher:             Georgeleen Osuji

Supervisor:                       Prof Y QIN
Co-supervisor:                    Mr J Ndege

On behalf of the University of Fort Hare's Research Ethics Committee (UREC) I hereby give ethical approval in respect of the undertakings contained in the above-mentioned project and research instrument(s). Should any other instruments be used, these require separate authorization. The Researcher may therefore commence with the research as from the date of this certificate, using the reference number indicated above.

Please note that the UREC must be informed immediately of

- Any material change in the conditions or undertakings mentioned in the document
- Any material breaches of ethical undertakings or events that impact upon the ethical conduct of the research

The Principal Researcher must report to the UREC in the prescribed format, where applicable, annually, and at the end of the project, in respect of ethical compliance.

**Special conditions:** Research that includes children as per the official regulations of the act must take the following into account:
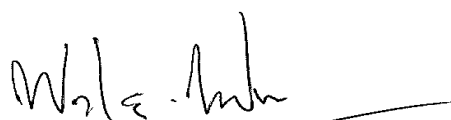
Note: The UREC is aware of the provisions of s71 of the National Health Act 61 of 2003 and that matters pertaining to obtaining the Minister's consent are under discussion and remain unresolved. Nonetheless, as was decided at a meeting between the National Health Research Ethics Committee and stakeholders on 6 June 2013, university ethics committees may continue to grant ethical clearance for research involving children without the Minister's consent, provided that the prescripts of the previous rules have been met. This certificate is granted in terms of this agreement.

The UREC retains the right to

- Withdraw or amend this Ethical Clearance Certificate if
  - o Any unethical principal or practices are revealed or suspected
  - o Relevant information has been withheld or misrepresented
  - o Regulatory changes of whatsoever nature so require
  - o The conditions contained in the Certificate have not been adhered to

- Request access to any information or data at any time during the course or after completion of the project.

- In addition to the need to comply with the highest level of ethical conduct principle investigators must report back annually as an evaluation and monitoring mechanism on the progress being made by the research. Such a report must be sent to the Dean of Research's office

The Ethics Committee wished you well in your research.

Yours sincerely

**Professor Wilson Akpan**
**Acting Dean of Research**

19 January 2017

APPENDIX 2

THE QUESTIONNAIRE

University of Fort Hare
*Together in Excellence*

SECTION A:  DEMOGRAPHIC SOCI-ECONOMIC

1. Age…
2. Gender: Male…………………… Female…………………….
3. Nationality: SA citizen……………….National of other Countries………………..
4. Race: White……………..
   Coloured…………..Black………………Others…………….
5. Level of Education:  Primary school…….High school……Post matric….
   None……...
6. Marital status: Single………. Married……….. Divorced…………
   Widowed………… Separated…………… Others…………….

SECTION B: CLINICAL DATA

7. Date admitted…………..
8. Weight…………..kg
9. Height…………cm
10. Temperature………….$^0$C
11. Patient's complaints: Fever…………. Abdominal pain…………. Vomiting………….
12. Source of cholera: Water…………. Others……………
13. Number of days with cholera…………………days
14. Frequency of stools/24hours……………
15. Frequency of vomit/day………………
16. Patient's stool characteristics: Watery stool…. Loose stool….. Blood in the
    stool……….. Mucous in stool……… Fatty stool……… Pale coloured…… extremely
    smelly………..
17. Patient's dehydration status: None…….. Mild……… Moderate……….
    Severe………..
18. Self-medication drug used before visiting clinic: Anticholinergic……. Oral
    rehydration therapy…. Antipyretics…… Probiotics…… Dioctahedralsmectite…..
    Herbal medicines….. Antibiotics…………
19. Laboratory tests ordered: Routine stool examination… Stool culture for
    vibrocholerae….. Stool culture for non-vibrocholerae…… Routine blood
    examination…….
20. Interventions ordered by treating physician: Antiemetics….. Anticholera…..
    Antipyretic… Anticholinergic….. Antiacids…… Herbal medicines……..
    Probiotics……. Dioctahedralsmectite…. Antibiotics….. Rehydration…….
21. Rehydration and Antobiotic therapies ordered for patients: ORT………. Intravenous
    fluid replacement therapy……. Vomiting therapy…… Antibiotics……

22. Antibiotics prescribed: Rifaximin…… Aztreonam…. Metronidazole…… S-G cephalosporins… T-G cephalosporins… Aminoglycosides…. Fluoroquinolous…… Others…

23. Management of cholera diagnosis: Acute infectious cholera… Acute bacterial dysentery…

24. What is your source of water supply? Tap……… Carrier/Tanker……… Dam/River………. Rain water………. Spring………. Windmill……….

25. If cholera is acute, where did you eat in the last 24hours:……………

26. What food ate before having cholera?...................

27. Alcohol history: Yes………… No……………..

APPENDIX C

NHRD APPROVAL LETTER

Eastern Cape Department of Health

| Enquiries: | Madoda Xokwe | TelNo: | 040 608 0856 |
| Date: | 27 February 2017 | Fax No: | 0436421409 |
| e-mail address: | madoda.xokwe@echealth.gov.za | | |

Dear Ms. G. Osuji

**Re: Determinant of Risk Factors Associated With Cholera and Its Effect in Nkonkobe Municipality Eastern Cape South Africa (EC_2017RP58_883)**

The Department of Health would like to inform you that your application for conducting a research on the abovementioned topic has been approved based on the following conditions:

1. During your study, you will follow the submitted protocol with ethical approval and can only deviate from it after having a written approval from the Department of Health in writing.

2. You are advised to ensure, observe and respect the rights and culture of your research participants and maintain confidentiality of their identities and shall remove or not collect any information which can be used to link the participants.

3. The Department of Health expects you to provide a progress on your study every 3 months (from date you received this letter) in writing.

4. At the end of your study, you will be expected to send a full written report with your findings and implementable recommendations to the Epidemiological Research & Surveillance Management. You may be invited to the department to come and present your research findings with your implementable recommendations.

5. Your results on the Eastern Cape will not be presented anywhere unless you have shared them with the Department of Health as indicated above.

Your compliance in this regard will be highly appreciated.

SECRETARIAT: EASTERN CAPE HEALTH RESEARCH COMMITTEE

72