MODELLING COUNT DATA WITH EXCESS ZEROS: AN EMPIRICAL APPLICATION TO TYPHOID FEVER CASES IN OYO STATE, NIGERIA

BY

ADEPOJU-OLAJUWON, FATIMAH AYOBAMI

B.Sc. STATISTICS (UNIVERSITY OIF ILORIN)

MATRIC NUMBER: 188567

A DISSERTATION SUBMITTED TO

THE DEPARTMENT OF EPIDEMIOLOGY AND MEDICAL STATISTICS, FACULTY OF PUBLIC HEALTH, COLLEGE OF MEDICINE, UNIVERSITY OF IBADAN

IN PARTIAL FULFILMENT OF

THE REQUIREMENT FOR THE AWARD OF

MASTERS OF SCIENCE DEGREE IN MEDICAL STATISTICS

DECEMBER, 2016

CERTIFICATION

We certify that this project work was carried out under our supervision by Adepoju-Olajuwon, Fatimah.A of the Department of Epidemiology and Medical Statistics, Faculty of Public Health, College of Medicine, University of Ibadan.

.....OBYusuf..... 11/01/2019

Supervisor

Dr. (Mrs.) O.B. Yusuf

B.sc (Ibadan); M.Sc. (Ibadan); Ph.D. (Ibadan)

Department of Epidemiology and Medical Statistics,

Faculty of Public Health,

University of Ibadan

.....11/01/2019

Co-Supervisor

Dr. R.F. Afolabi

B.Sc. (Ilorin); M.Sc. (Ilorin); Ph.D. (Ilorin)

Department of Epidemiology and Medical Statistics,

Faculty of Public Health,

University of Ibadan

DEDICATION

To Almighty Allah (SWT) for His infinite mercy on me.....

.....To the Olajuwon Family

.....To my husband; Adepoju Sulaiman for his understanding and emotional support

.....and my Amiable daughter; (Adepoju Mujahidah) for her co-operation and endurance.

ACKNOWLEDGEMENT

All praise is due to Almighty Allah for giving me strength and sound mind in carrying out this research.

My heartfelt gratitude goes to my supervisors Dr. (Mrs.) O.B Yusuf and Dr. R.F Afolabi, for their homely and motivating guidance towards the success of this dissertation.

Also, to all the lecturers in the department of Epidemiology and Medical Statistics, I say a big "thank you" for making impact in my life.

The completion of this dissertation would have been impossible without the assistance of many individuals;

First, I am grateful to Mrs. Akinpelu KP and Kayode Fowobaje for their assistance towards my analysis. To my co-supervisee; Olamiposi Owoeye and Oluwaseun Babatunde for their endless advice and support towards the achievement of this dissertation. And to my course mates, Mrs. Olabisi Akinsanya for assisting in the extraction of data, Mr Olaniyi Olutola, Abiodun Adegbenro, Adekunmbi Ogunmokun and Emeka Nwimo for their helping hands in one way or the other and to others Wasiu Lamidi, Samuel Aghoghovia and Gabriel Ajayi, It's a nice experience been colleagues with them.

My special thanks to Dr. and Mrs. Temisanren for their moral support and accommodation during the course of this program. I am deeply indebted to Alh. Rasheedah Olawore for been there in taking care of my daughter, with support from Miss Suliah. Adepoju and Alh .Aduke Alimi (Junior).

I would never forget to appreciate my mother (Mrs. Ashiyah Olajuwon) for her prayers and words of encouragement at all times, if I have the opportunity of a reborn, I would still love to be her daughter. To my father (Engr. S.S Olajuwon) and all my siblings Halimah, Maryam, Bilkis, Zainab, and Jubril, even with the distance, I could still feel the affection of a family.

Also, my unending appreciation goes to all the Adepoju family and Mrs Rukayyat (Mum Malik) for the encouragement and understanding at all times. I appreciate being a member of this wonderful family.

Finally, I appreciate my husband Adepoju Sulaiman and my daughter for their unconditional love, endurance, patience, understanding and encouragement.

ACKNOWLEDGEMENT

All praise is due to Almighty Allah for giving me strength and sound mind in carrying out this research.

My heartfelt gratitude goes to my supervisors Dr. (Mrs.) O.B Yusuf and Dr. R.F Afolabi, for their homely and motivating guidance towards the success of this dissertation.

Also, to all the lecturers in the department of Epidemiology and Medical Statistics, I say a big "thank you" for making impact in my life.

The completion of this dissertation would have been impossible without the assistance of many individuals;

First, I am grateful to Mrs. Akinpelu KP and Kayode Fowobaje for their assistance towards my analysis. To my co-supervisee; Olamiposi Owoeye and Oluwaseun Babatunde for their endless advice and support towards the achievement of this dissertation. And to my course mates, Mrs. Olabisi Akinsanya for assisting in the extraction of data, Mr Olaniyi Olutola, Abiodun Adegbenro, Adekunmbi Ogunmokun and Emeka Nwimo for their helping hands in one way or the other and to others Wasiu Lamidi, Samuel Aghoghovia and Gabriel Ajayi, It's a nice experience been colleagues with them.

My special thanks to Dr. and Mrs. Temisanren for their moral support and accommodation during the course of this program. I am deeply indebted to Alh. Rasheedah Olawore for been there in taking care of my daughter, with support from Miss Suliah. Adepoju and Alh .Aduke Alimi (Junior).

I would never forget to appreciate my mother (Mrs. Ashiyah Olajuwon) for her prayers and words of encouragement at all times. if I have the opportunity of a reborn, I would still love to be her daughter. To my father (Engr. S.S Olajuwon) and all my siblings Halimah, Maryam, Bilkis, Zainab, and Jubril, even with the distance, I could still feel the affection of a family.

Also, my unending appreciation goes to all the Adepoju family and Mrs Rukayyat (Mum Malik) for the encouragement and understanding at all times. I appreciate being a member of this wonderful family.

Finally, I appreciate my husband Adepoju Sulaiman and my daughter for their unconditional love, endurance, patience, understanding and encouragement.

LIST OF ABBREVIATIONS

REID-Re-emerging Infectious Disease

IDSR-Integrated Disease Surveillance and Response

LGA-Local Government Area

ZIP-Zero inflated Poisson

ZINB-Zero inflated Negative Binomial

ZIGP-Zero inflated Generalized Poisson

ZAP-Zero altered Poisson

AIC-Akaike Information Criteria

WHO-World Health Organization

CDC-Centre for Disease Control and Prevention

UNICEF- United Nations Children's Fund

FMoH- Federal Ministry of Health

WASH-Water, Sanitation and hygiene

WSSCC- Water supply and sanitation collaborative council

MDG- Millennium development Goals

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Poisson and negative binomial regression are most popularly used to model count data but with the limitation of not accounting for the excess zeros in the data which may subsequently lead to biased estimates. Hence, this study modeled the occurrence of typhoid fever in Oyo state while accounting for the excess zeros and over-dispersion in the dataset.

A longitudinal surveillance data on typhoid fever cases was obtained from the Integrated Disease Surveillance and Response (IDSR) of Oyo state Ministry of health from 2011 to 2014. The number of reported cases of typhoid in the state was the outcome variable, while month of reporting, year of reporting, and local government areas (LGA) were the explanatory variables. The presence of over-dispersion in the data was investigated using the mean and variance. Zero-inflated models such as zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated generalized Poisson (ZIGP), and zero-altered Poisson (ZAP or hurdle) were fitted to the data. The Akaike information criteria (AIC) and the -2logL were used to select the best model among the four. Descriptive statistics, incidence rate ratios, as well as 95% CI were determined.

The total number of typhoid fever cases reported in the state was 2,970 (Mean=3.46, SD=3.89). There was 34.7% increase of typhoid fever incidence between 2011 and 2012 and decline of 89.0% between 2012 and 2014. About 11% of typhoid cases were reported in February and March while the lowest cases of the disease were reported in October (3.4%). The risk of typhoid was highest in Surulere LGA by 30.7% (IRR=4.307, 95% CI=0.892, 2.028), followed by Oyo west and Oyo East LGA by 89.0% and 56.2% (IRR=3.890, 95% CI=0.799, 1.918); (IRR=3.562, 95% CI=0.743, 1.798) respectively. In addition, the risk of typhoid was lowest in Atisbo LGA by 54.0% (IRR= 1.540, 95% CI= 0.003, 0.861). Ibarapa East LGA had 61.0% reduced risk of typhoid fever (IRR= 0.39, 95% CI=-1.309, -0.576). The AIC of the models were 51290.47, 30733.61, 51290.47, 51285.73, for the ZIP, ZINB, ZIGP, and ZAP models respectively, thus indicating that ZINB had the least AIC value.

The occurrence of typhoid fever was influenced by season (i.e. Month of reporting) and LGA of reporting. The zero inflated negative binomial (ZINB) was found to be the best regression model to estimate the factors that influences the number of typhoid cases in Oyo state in the presence of over-dispersion. The model is recommended for researchers with similar data.

Key words: Typhoid fever, count data, zero-inflated models, over-dispersion.

Word count: 410

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

Modeling count data poses a lot of challenge in many areas of interest including public health, epidemiology, sociology, psychology, engineering, agriculture and others. Count data is one which the observations can take only the non-negative integer values {0, 1, 2, 3, 4...} where these integers arise from counting rather than ranking. The Poisson, binomial and negative binomial distributions are commonly used to represent the distribution of count data. They are also referred to as the number of occurrences of an event of interest, examples of such event includes; number of accidents on a major road per year, number of products produced by a manufacturing company per month, number of suicide attempts, number of heart attacks, and number of medical cases of a re-emerging disease and so on.

Usually, the Poisson model is assumed for analysing or approximating the distribution of the count observations. However, presence of dispersions in the observed counts is underestimated using the Poisson model. The Poisson models are violated when the range of count values is limited or when over-dispersion is present. Over-dispersion occurs because a single Poisson parameter is often insufficient to describe the population. In many cases, it can be suspected that population heterogeneity which has not been accounted for is causing this over-dispersion. This population heterogeneity is unobserved, in other words, the population consists of several subpopulations, but the subpopulation membership is not observed in the sample. (Böhning et al, 1999).

Over-dispersion is the presence of greater variability (statistical dispersion) in a dataset than would be expected based on a given statistical model, which occurs when the response variance is greater than the mean, and the presence of this may cause standard errors of the estimates to be deflated or underestimated. Over-dispersion is often encountered when fitting very simple parametric models such as those based on the Poisson distribution. Extra zeros in the data also poses a lot of difficulty in analyzing count data, and this led to the development of zero-inflated count models by Lambert (1992) to provide a method of accounting for excessive zero counts.

Further, theory suggests that the excess zeros are generated from the count values and that the excess zeros can be modeled independently.

Zero-inflated models are statistical models based on a zero-inflated probability distribution, i.e. a distribution that allows for frequent zero valued observations. These models are designed to accommodate excess zeros in count data; they are also referred to as added zero models (Heilbron 1994). Zero-inflated models have been developed for the Poisson model (ZIP) (Lambert 1992), the negative binomial model (ZINB) (Hinde, and Demetrio, 2001), the generalized Poisson model (ZIGP) and zero-altered Poisson (ZAP or Hurdle) models. A zero-inflated model assumes that the zero observations have two different origins: "structural" and "sampling". The sampling zeros are due to the usual Poisson (or negative binomial) distribution, which assumes that those zero observations happened by chance and the structural zeros are observed due to some specific structure in the data. For example, if a count of high-risk sexual behaviors is the outcome, some participants may score zero because they do not have a sexual partner; these are the structural zeros since they cannot exhibit unprotected sexual behavior. Other participants have sexual partners but score zero because they have eliminated their high-risk behavior. That is, their risk behavior is assumed to be on a Poisson or negative binomial distribution that includes both zeros (the "sampling zeros") and non-zero counts (Mei-Chen et al, 2011).

The basic assumption of the Poisson regression model is that the conditional variance of the outcome is equal to the conditional mean, but in practice, the distribution of counts, such as incidents of substance use or other risk behaviors, often has a much larger than expected number of observed zeros than assumed by Poisson distribution, called "zero-inflated". For instance, many patients may already be abstaining or not having unprotected sexual occasions. The negative binomial regression can be written as an extension of Poisson regression and it enables the model to have greater flexibility in modeling the relationship between the conditional variance and the conditional mean compared to the Poisson model. Also, an often encountered characteristic of count data is that the number of zeros in the sample can exceed the number of zeros predicted by either Poisson or negative binomial model, and this is of interest because zero counts frequently have special status.

In general, Hurdle models and zero-inflated models are used for modeling count data with a multitude of zeros. The hurdle model (Mullahy, 1986) is a two component model in which one component models the probability of zero counts and the other component uses an abridged Poisson/negative binomial distribution that modifies an ordinary distribution by conditioning on a positive outcome. The zero-inflated model has a distribution that is a mixture of a binary distribution that is degenerate at zero and an ordinary count distribution such as Poisson or negative binomial. The Hurdle model considers the zeros to be completely separate from the non-zeros. The zero-inflated model is similar to `the Hurdle model; however, it permits some of the zeros can be analyzed along with the non-zeros.

1.2 PROBLEM STATEMENT

The National Institute of health reported in 2012 that re-emerging infectious diseases (REIDs) are diseases that once were major health problems globally or in a particular country and then declined dramatically, but are again becoming health problems for a significant proportion of the population. The re-emergence of some diseases can be explained by the failure to immunize enough individuals, which results in a greater proportion of susceptible individuals in a population and an increased reservoir of the infectious agent. Infectious diseases have a devastating influence nationally and globally, though several approaches can assuage distress due to these diseases (NIH, 2012). Increases in the number of individuals with conceded immune systems (due to the stress of famine, war, crowding or disease) also explain increases in the rate of emerging and re-emerging diseases.

Developed countries have guidelines that help protect the general public from infectious diseases. Public health measures typically involve eradicating the pathogen from its route of transmission. Those measures include ensuring a safe water supply, efficiently handling sewage treatment and disposal, and instigating food-safety, animal-control, and vaccination programs. Many pathogens that cause gastro intestinal diseases (like those that cause typhoid fever) are transmitted via water. For example, cholera has repeatedly reemerged over more than two centuries in association with global travel, changing seasons, war, natural disasters, and conditions that lead to inadequate sanitation, poverty, and social disruption (Morens and Fauci, 2013). Disease and microbial ecology research has identified multiple modes of transmission for

both vector-borne and waterborne pathogens that depend on environmental, climatic, infrastructural, and sociocultural conditions (Wright et al. 2004). Many pathogens move about the environment via human feces (Curtis et al. 2000), and both humans and animals can act as hosts. Exposure to fecal pathogens occurs in both private (e.g., domestic living spaces, private yards, and fields) and public spaces (e.g., workplaces, transportation hubs, markets; Kosek et al. 2003) and is most often linked to poverty, poor education, and underdevelopment.

The prevalence of emerging and re-emerging waterborne diseases are largely caused by unsafe water, inadequate sanitation and poor hygiene among human population, and is majorly related to the dearth of potable water in most parts of the developing countries (USAID, 2005). Lack of proper preparation by most developing countries on population upsurge has a major influence on increasing demand to access potable water in their cities. Hence, the populace often results to sourcing water from questionable water points. During the dry season, urban dwellers utilize water collected from alternative sources such as boreholes, wells and streams for domestic uses. Available statistics indicate that the inhabitants of Oyo state in Nigeria suffer mainly from diarrhea, gastro-enteritis, malaria, measles, tuberculosis, cholera and typhoid fever, in that order. (Oguntoke et al. 2009). Drawing from the observations made by Snow (1894) to that of Swerdlow (1992), it is clear that the environmental factor "water" and the problem of water borne diseases are linked up via two different mechanisms; through the supply of contaminated water and or lack of water for personal hygiene. Oyo state people lack adequate water supply hence, they can be susceptible to waterborne diseases.

The occurrence of flooding after heavy rainfall also potentially facilitates the transmission of waterborne and vector-borne diseases. Flooding is associated with an increased risk of infection, however this risk is low unless there is significant population displacement or water sources are compromised. The major risk factor for outbreaks associated with flooding is the contamination of drinking water facilities. The risk of outbreaks can be minimized if the risk is well recognized and disaster-response addresses the provision of unclean water as a priority. There is an increased risk of infection of water-borne diseases contracted through direct contact with polluted waters (WHO, 2016).

It is important to model the presence or absence of water-borne disease due to the excess zeros involved. Common methods of data analysis involving counts and series of zeros typically are

the Poisson and negative binomial techniques. However, these approaches to modeling may not be appropriate when observation include large number of zeros (Lambert 1992), such as the number of reported cases of re-emerging water-borne diseases in the 33 LGAs of Oyo state. The method of estimation used for analyzing the number of cases infectious diseases may however pose a problem when extra zeros in the dataset are not accounted for which may result in biased parameter estimates and wrong inferences.

1.3 JUSTIFICATION

Waterborne diseases are among the most emerging and re-emerging infectious diseases throughout the world. They are mostly endemic with a worldwide distribution and they have a heterogeneous etiology (Onyango and Angienda, 2010). Many of the existing studies have focused on re-emerging infectious diseases and waterborne diseases separately, but little or no published research have used waterborne disease (Typhoid fever) as a form of re-emerging infectious disease (REIDs). Data on these re-emerging infectious diseases are counts which usually includes lots of zeros. In view of the excess zeros that usually exist in count data, and to overcome the limitation of not taking into account these extra zeros, the zero-inflated count models provide a way for modeling excess zeros as well as allowing for over-dispersion. The excess zeros are a form of over dispersion, failure to account for them constitutes a model misspecification, which results in biased standard errors.

1.4 AIM AND SPECIFIC OBJECTIVES

1.4.1 Aim:

To model the annual occurrence of typhoid fever as a re-emerging waterborne disease in all the Local Government areas (LGA) of Oyo state.

1.4.2 Specific Objectives

1. To examine the pattern of occurrence of typhoid fever cases over the years (2011-2014) in all the LGAs of Oyo state.
2. To compare the performance of the ZIP, ZINB, ZIGP and the ZAP models in the analysis of typhoid fever cases in the presence of over-dispersion

3. To explore the effects of season, year of reporting and geographical location on typhoid fever occurrence.

3. To explore the effects of season, year of reporting and geographical location on typhoid fever occurrence.

CHAPTER TWO

LITERATURE REVIEW

2.1 RE-EMERGING DISEASES

The re-emergence of water-borne pathogens is often related to changes in the environment, the pathogen, or the host. For instance, the phenomenon of malaria re-emergence is frequently related to changes that affect the spread of the mosquito vectors (Tyagi, 2004), the development of drug-resistant parasites (Sharma, 1996) or human migration (Marques, 1987). The epidemiology of water-borne diseases is often associated with other events affecting the populations that are exposed to the infectious organisms. Different countries and continents of the world experiences easy spread of these parasites. For instance, the spread of the parasite to the new destinations of the hosts can be influenced by moving human who are nurturing the infections.

Study by Louis MW (2008) revealed that factors related to the emergence of infectious diseases are numerous and such as overpopulation, disruption due to war, mass migration of populations due to national or artificial disasters, inadequate food and water supplies, environmental changes. migration from rural to urban centers, changes in farming practice, social and cultural factors such as food habits and religious beliefs, human demography, pathogen changes. There are many possibilities to disseminate infectious parasites which make control efforts difficult. Increased diagnosis of parasitic food-borne diseases has been attributed to increased population of susceptible individuals, increased international travel, and improvement in diagnostic techniques and change in eating habits. Also, Chomel et al (2007) explained that the misuse of ploughed fields has resulted in the partial or permanent occupation of humans of such lands thereby sharing such locations with parasites and animals with the possibility of transmitting new infections. Displacement of wildlife through development can result in both global warming (Hoberg et al, 2008) and epidemics of wildlife disease with the resultant spread of the infections to human population.

Nigeria in particular and the entire West African region in general are not isolated from the threat of emerging and re-emerging bacterial diseases and other infectious diseases. However, the rate at which such infections are discovered and reported are lower in these region perhaps due to

poor surveillance, poor diagnostic methods and reporting, In addition, poor sanitation, hot tropical climate, poverty, poor government prioritization in disease control, climate change, floods, increased wars and civil conflicts with resultant transposition of large numbers of people and congested refugee camps, malnutrition and misuse of drugs resulting in drug resistance are common and can seriously intensify the rate of disease development and spread.

In addition, Ngongeh and Chiejina (2014) identified a number of notifiable diseases (EIDs) in Nigeria that are escalating even though it may not be as long as those reported for some developed countries like America. This includes drug resistant malaria, African trypanosomosis, chikungunya, schistosomosis, babesiosis, cryptosporidiosis, Lyme disease, leishmaniiasis, filariasis, giardiasis, onchocerciasis, loiasis; yellow fever, dirofilariasis and hookworm (human helminthiasis). Various parts of the world have recognized and recorded a good number of emerging and re-emerging diseases/parasites and are either helminth, entomologically borne in origin. Some members of these classes of parasites are also common in Nigeria. These include: lyme boreliosis, cryptosporidiosis, malaria and yellow fever. However, many other emerging parasites and disease conditions which are not on this list but are also tagged as emerging and reemerging and occurring not necessarily only in Nigeria or the entire West African region but in other regions of the world have been listed under the relevant subheadings namely helminth, protozoa/rickettsia and entomological/ vector borne emerging and re-emerging animal parasitic diseases.

The state of emerging and re-emerging diseases in Nigeria is quite similar to the universal situation as emerging and re-emerging diseases are equally being reported in Nigeria in the same way as in other countries of the world, for example cryptosporidiosis which is emerging in Nigeria has also been reported as an emergent disease in the United States of America. Therefore, emerging and re-emerging infections are a global occurrence which is not surprising because some of the predisposing factors such as climate change, rapid infrastructural development and attendant loss of parasite habitat are unlimited and thus globally common. Another factor that is especially important in the re-emergence of diseases is the acquired resistance of pathogens to antimicrobial medications such as antibiotics. Both bacteria and viruses can change over time and develop resistance to these drugs, so that drugs that were effective in controlling disease in the past are no longer useful.

poor surveillance, poor diagnostic methods and reporting, In addition, poor sanitation, hot tropical climate, poverty, poor government prioritization in disease control, climate change, floods, increased wars and civil conflicts with resultant transposition of large numbers of people and congested refugee camps, malnutrition and misuse of drugs resulting in drug resistance are common and can seriously intensify the rate of disease development and spread.

In addition, Ngongeh and Chiejina (2014) identified a number of notifiable diseases (EIDs) in Nigeria that are escalating even though it may not be as long as those reported for some developed countries like America. This includes drug resistant malaria, African trypanosomosis, chikungunya, schistosomosis, babesiosis, cryptosporidiosis, Lyme disease, leishmaniiasis, filariasis, giardiasis, onchocerciasis, loiasis; yellow fever, dirofilariasis and hookworm (human helminthiasis). Various parts of the world have recognized and recorded a good number of emerging and re-emerging diseases/parasites and are either helminth, entomologically borne in origin. Some members of these classes of parasites are also common in Nigeria. These include: lyme boreliosis, cryptosporidiosis, malaria and yellow fever. However, many other emerging parasites and disease conditions which are not on this list but are also tagged as emerging and reemerging and occurring not necessarily only in Nigeria or the entire West African region but in other regions of the world have been listed under the relevant subheadings namely helminth, protozoa/rickettsia and entomological/ vector borne emerging and re-emerging animal parasitic diseases.

The state of emerging and re-emerging diseases in Nigeria is quite similar to the universal situation as emerging and re-emerging diseases are equally being reported in Nigeria in the same way as in other countries of the world, for example cryptosporidiosis which is emerging in Nigeria has also been reported as an emergent disease in the United States of America. Therefore, emerging and re-emerging infections are a global occurrence which is not surprising because some of the predisposing factors such as climate change, rapid infrastructural development and attendant loss of parasite habitat are unlimited and thus globally common. Another factor that is especially important in the re-emergence of diseases is the acquired resistance of pathogens to antimicrobial medications such as antibiotics. Both bacteria and viruses can change over time and develop resistance to these drugs, so that drugs that were effective in controlling disease in the past are no longer useful.

poor surveillance, poor diagnostic methods and reporting, In addition, poor sanitation, hot tropical climate, poverty, poor government prioritization in disease control, climate change, floods, increased wars and civil conflicts with resultant transposition of large numbers of people and congested refugee camps, malnutrition and misuse of drugs resulting in drug resistance are common and can seriously intensify the rate of disease development and spread.

In addition, Ngongeh and Chiejina (2014) identified a number of notifiable diseases (EIDs) in Nigeria that are escalating even though it may not be as long as those reported for some developed countries like America. This includes drug resistant malaria, African trypanosomosis, chikungunya, schistosomosis, babesiosis, cryptosporidiosis, Lyme disease, leishmaniiasis, filariasis, giardiasis, onchocerciasis, loiasis; yellow fever, dirofilariasis and hookworm (human helminthiasis). Various parts of the world have recognized and recorded a good number of emerging and re-emerging diseases/parasites and are either helminth, entomologically borne in origin. Some members of these classes of parasites are also common in Nigeria. These include: lyme boreliosis, cryptosporidiosis, malaria and yellow fever. However, many other emerging parasites and disease conditions which are not on this list but are also tagged as emerging and reemerging and occurring not necessarily only in Nigeria or the entire West African region but in other regions of the world have been listed under the relevant subheadings namely helminth, protozoa/rickettsia and entomological/ vector borne emerging and re-emerging animal parasitic diseases.

The state of emerging and re-emerging diseases in Nigeria is quite similar to the universal situation as emerging and re-emerging diseases are equally being reported in Nigeria in the same way as in other countries of the world, for example cryptosporidiosis which is emerging in Nigeria has also been reported as an emergent disease in the United States of America. Therefore, emerging and re-emerging infections are a global occurrence which is not surprising because some of the predisposing factors such as climate change, rapid infrastructural development and attendant loss of parasite habitat are unlimited and thus globally common. Another factor that is especially important in the re-emergence of diseases is the acquired resistance of pathogens to antimicrobial medications such as antibiotics. Both bacteria and viruses can change over time and develop resistance to these drugs, so that drugs that were effective in controlling disease in the past are no longer useful.

2.1.1 WATERBORNE DISEASES

Waterborne diseases are caused by drinking polluted or dirty water. Contaminated water can cause many types of diarrheal diseases, including Cholera, and other serious ailments such as Guinea worm disease, Typhoid, and Dysentery. A diversity of microorganism, bio toxins, and toxic contaminants that symptomatically causes waterborne disease glaringly includes protozoa, viruses, bacteria, and intestinal parasites. These microorganisms lead to overwhelming illnesses such as cholera, schistosomiasis and other gastrointestinal problems. Yet other important classes of water-borne diseases are caused by metazoan parasites. Distinct examples include certain Nematodes like "roundworms". For instance, important waterborne nematodal diseases are Dracunculiasis, Schistosomiasis caused by a family of blood flukes affecting several millions of people worldwide.

However, when toxins find their way into drinking water sources and are not eliminated by water treatment processes, it can and does make people sick. This can be ascribed to lack of sanitation system, pipe breaks, leakages, ground water pollution campsites where human and wildlife use same source of water. Uncoordinated efforts of various federal, state and local agencies make Nigeria one of the countries in the world that has hazardous water supplies. Most Nigerians will usually contract a waterborne illness as a result of the low quality of drinking water.

A study by Alagiah reported that during the early 1980s, 500 million people fell victim of water related diseases. Out of which 30,000 people (half being infants) die daily and in terms of hospital burden, 50% of hospital beds are occupied by patient suffering from these diseases and 80% of man's ill-health is due to water problems in the more poverty stricken parts of developing countries (Alagiah, 1981). According to the World Health Organization (2014), 1.8 million people die annually as a result of diarrheal disease which accounted for an estimated 4.1% of the total daily global burden of disease. It was also projected that 58% of that burden or 842,000 deaths yearly is attributable to perilous water supply, sanitation and hygiene and is mostly intense in children from developing countries.

Previous studies have shown that typhoid fever ranked the highest among the water related diseases recorded between 2002 and 2008 in Nigeria, followed by cholera, hepatitis and

dracunculiasis. This incidence of waterborne diseases is as a result of inability to gain access to portable drinking water most especially peopling living in the rural areas of the country (Yusuff et al, 2014). According to another study conducted by Oguntoke et al, it was reported that typhoid fever was the most commonly reported water-borne disease, followed by bacillary dysentery. Also, about 31% waterborne diseases patients are mostly children less than nine years and 70% of people less than 29years. Similarly, those in the age group of 20-39years have the most reported cases of typhoid fever (Oguntoke et al, 2009).

Typhoid, cholera and diarrhea are water borne diseases which have claimed several million lives globally (Omole et al, 2015). Consumption of contaminated water is the major mode of contracting these diseases. UNICEF/WHO (2012) also suggests that meeting the Millennium Development Goal (MDG) in the world is likely not feasible on sanitation as more than 2.5 billion people out of an estimated 7 billion still lack access to sanitation. Nigeria has been identified as one of few countries that would not meet the MDG for water even though the world, as a whole, met this goal since 2010 (Omole, 2013).

2.1.2 TYPHOID FEVER

Typhoid fever is a type of enteric fever caused by bacterium *Salmonella typhi*, also known as *Salmonella enterica* serotype typhi, growing in the intestines and blood (CDC, 2013). This fever is spread by eating or drinking food or water contaminated with the feces of an infected person (WHO, 2008). Typhoid fever does not affect animals other than humans. It can only spread in environments where human feces or urine are able to come into contact with food or drinking water, transmission also occurs through eating raw fruit and vegetables fertilized by human excreta and through ingestion of contaminated milk and milk products. Flies may cause human infection through transfer of the infectious agents to foods. Pollution of water sources may produce epidemics of typhoid fever when large numbers of people use the same source of drinking-water (WHO, 2016). It occurs most often in children and young adults between 5 and 19 years old (WHO, 2007). However, the highest case fatality rates are reported in children <4 years of age (WHO, 2016). The disease is most commonly transmitted through poor hygiene habits and public sanitation conditions. The most prominent feature of the infection is fever which gradually rises to a high plateau. Symptoms such as diarrhea, constipation, abdominal pain

and encephalopathy may occur. Complications like intestinal perforation and gastrointestinal haemorrhage may occur in severe disease.

Typhoid fever occurs worldwide, it is predominantly endemic in many parts of the developing world, and as global travel rises, illness can and do occur around the world in span of a day (Lifshitz, 1996). Typhoid fever is endemic in Asia, Africa, Latin America, the Caribbean, and Oceania, but 80% of cases come from Bangladesh, China, India, Indonesia, Laos, Nepal, Pakistan, or Vietnam (Chau et al, 2007). Similarly, outbreaks of typhoid fever are also frequently reported from sub-Saharan Africa and countries in Southeast Asia (Muyembe et al, 2009; Baddam et al, 2012).

Furthermore, study relates that typhoid fever infects roughly 22 million people (incidence of 3.6 per 1,000 populations) and kills an estimated 200,000 people yearly (Crump et al, 2004). Also in another study, Africa has an average yearly incidence of 7.6 of typhoid fever per million travelers from 1999-2006 by country or region of departure (Lynch et al, 2009). It remains a serious problem in Nigeria (Agbakwuru et al., 2003). Most documented typhoid fever cases involve school-aged children and young adults. However, the true incidence among very young children and infants is thought to be higher in urban areas where sewage disposal is lacking or inadequate, water supplies get contaminated and thus cause the outbreaks of typhoid. The contamination of food by carrier is the second most frequent route of infection (Hornick, 1985).

2.2 INTEGRATED DISEASE SURVEILLANCE AND RESPONSE (IDSR)

Disease surveillance is one of the key components of public health which involves the collection, analysis, and interpretation of data on diseases that informs planning and implementation of health system interventions and policies. Health information that can be used to formulate evidence-based health policies that will improve the quality of life of any community might be difficult to generate without effective disease surveillance. Preparation and response to disease outbreaks are also obtained through disease surveillance dataset. Several advancement have been made to improve the capacity of countries to identify, report, prepare, and respond to occurrences of disease outbreaks over the years. However, effective disease surveillance still

poses a challenge in Nigeria and the inability to provide functional surveillance systems is intensified by high incidence rates of infectious diseases (FMoH, 2009).

The National disease surveillance system was introduced in Nigeria in 1988 following an outbreak of yellow fever. This was attributed to ineffective surveillance system in most states of the country. Several problems not unique to Nigeria but were replicated in many African countries was inundated in the Disease Surveillance and Notification System (DSN) which includes; failure to generate information for quick response, inadequate laboratory facilities for diagnosis, and presence of numerous upright surveillance programs which led to misuse of already limited resources. The World Health Organization Regional Office for Africa (WHO AFRO, 1999) proposed the adoption of the new strategy by its member States following the string of preventable outbreaks of infectious diseases in the 1990s. This strategy, named Integrated Disease Surveillance and Response (IDSR) was designed to provide a basis for a multi-tiered surveillance system which laid down activities at each level for the detection, reporting, preparedness and response to disease outbreaks.

Nigeria and other African countries adopted the IDSR strategy in 1998 at the 48th World Health Organization Regional committee meeting in Harare, Zimbabwe, through a resolution (AFRO). Implementation of the strategy commence in Nigeria in 2003 which was also adopted by the National council on Health in 2005 and approved by Federal Executive Council in 2006 (The Nigeria Academy of Science; IDSR, 2010). The IDSR tool is a comprehensive strategy for strengthening disease surveillance and response to epidemics at all levels (i.e. community, health facility, local government areas, state, and national) through rational use of resources. Its major objective was to strengthen the capacity of countries to carry out effective disease surveillance while emphasizing the need for a single integrated system which maximizes human and material resources. Consequently, the IDSR tool was adopted by 44 of the 46 African countries. Since the introduction of the IDSR strategy, disease reporting and response to epidemics have relatively improved.

The Federal Ministry of Health, Nigeria listed a number of Priority diseases and events for the IDSR system in Nigeria. These diseases classification is shown in table 1.

Table 1.1: List of Priority Diseases and Events in Nigeria

Epidemic Prone	International Health Regulations (IHR) recommended	Diseases targeted for elimination or Eradication	Other diseases of public health importance	
			Communicable	**Non-Communicable**
i. Cholera ii. Diarrhoea with blood (Shigella{ Sd1}) iii. Measles iv. Meningitis v. Viral Haemorrhagic fevers (Lassa fever) vi. Human Influenza (caused by a new subtype) vii. Yellow fever	i. SARS ii. Smallpox iii. Dengue iv. Anthrax v. SARI	i. Poliomyelitis ii. Dracunculiasis iii. Leprosy iv. Neonatal tetanus v. Lymphatic Filariasis vi. Tuberculosis	i. Diarrhoea (Children under-five) ii. Pneumonia(Children under-five) iii. HIV/AIDS iv. Malaria v. Onchocerciasis vi. STIs vii. Trypanosomiasis viii. Buruli Ulcer ix. Plaque x. Trachoma xi. Typhoid xii. Hepatitis B	xvii. Asthma xviii. Diabetes Mellitus xix. Epilepsy xx. High blood pressure xxi. Sickle cell disease xxii. Malnutrition

			xiii. Pertusis xiv. Human Rabies xv. Schistomiasis xvi. Noma	

2.3 ZERO INFLATED REGRESSION MODELS

Biomedical researches have count variables as outcomes of interest. This count data often have excessive number of zero outcomes than expected in Poisson regression. Unlike the normal distribution, the variance of a Poisson random variable depends on the mean, with the mean equal to the variance. Count data frequently depart from the Poisson distribution due to a larger frequency of extreme observations greater than the mean in the observed distribution resulting in spread (variance) called "over-dispersion". For example, services of psychiatric outpatient may report a large proportion of zero utilization of such services for many patients (Neelon et al. 2010). Frequent display of over-dispersion and excess zeros in real-life data are the main motivation for zero-inflated count models (Lambert, 1992; Greene, 1994). In addition to allowing for over-dispersion, these models provide a way of modeling the excess zeros.

However, in some situations, the multitudes of zero counts are the major source of over-dispersion, and negative binomial cannot be used to accurately model the resulting over-dispersion. In such circumstances, the zero-inflated (Poisson or negative binomial) methods of estimation can be used. Outcome variables often take the form of integers or counts, such as number of symptoms or number of risk behaviors during some defined time period which are generally not normally distributed.

Additionally, zero inflated models have a statistical advantage to standard Poisson and negative binomial models in that they model the preponderance of zeros as well as the distribution of positive counts simultaneously. Unfortunately, there is not a specific frequency of zero counts or ratio of zero to nonzero counts that can be used to determine if a particular distribution is zero-inflated. Therefore, the following four regression models, zero-inflated Poisson (ZIP) regression (Lambert, 1992; Bohning et al., 1999), zero-inflated negative binomial (ZINB) regression (Hinde and Demetrio, 2001), zero altered Poisson (ZAP or hurdle) regression (Mullahy, 1986), and zero-inflated generalized Poisson (ZIGP) regression (Famoye and Singh, 2003; Gupta et al., 2004) are commonly used to model zero-inflated count data. Other models in the literature include the two-part model (Heibron, 1994), and the semi-parametric model (Gurmu, 1997).

It has been proven that the ZIP parameter estimates can be severely biased if the non-zero counts are over-dispersed in relation to the Poisson distribution. In this case, a comprehensive form of

the negative binomial model for excess zero count data was described by Greene (1994), the Zero-Inflated Negative Binomial (ZINB), which may be more appropriate than the ZIP. Frequently, zero inflated regression models are extended to include random effects. The random-effects model accounts for the subject to subject variation as a way to directly model the correlation among the repeated measures within a subject which of course is as a result of a hierarchical study design or data collection where the observations are either clustered or repeated outcomes from individual subjects,. Additionally, Lee et al, (2006) extended the ZIP with random-effects to multi-level ZIP regression to model multilevel clustered count data, and Moghimbeigi et al, (2008) developed multi-level ZINB regression for modeling over-dispersed count data with extra zeros. Also, Mehmet (2012) confirmed in a study on "Modeling the effect of air pollutants on hospital admissions" that zero-inflated models (ZINB & ZIP) gives better results than the other models (Poisson regression, Negative binomial regression and the Generalized Poisson regression) in terms of model comparison and parameter estimation.

The generalized Poisson (GP) distribution was first introduced by Consul and Jain (1970) and subsequently studied in detail by Consul (1989). Czado et al also stated that the extended ZIGP regression model proved to be superior over GP and ZIP models and even ZIGP models with constant overall dispersion and zero-inflation parameters demonstrating the usefulness of the ZIGP. Another study conducted by Famoye and Singh (2006) on domestic violence also found that the zero-inflated generalized Poisson (ZIGP) regression is a good competitor of the ZINB that can be used to model over-dispersed data. They found that the ZIGP converges in fitting the data than the ZINB.

The Hurdle model is another zero-inflated model that fits over-dispersed data better. It is also called the zero-altered Poisson (ZAP) regression. These model is suitable to solve the excess zeros problem in the response variable when the data are censored from the right side (Saffari et al, 2012)

2.3.1 ZERO INFLATED POISSON REGRESSION (ZIP)

There are two main causes of equi-dispersion assumption violation. First, the frequency of the zero counts is more than the expected zero counts generated by the Poisson distribution. Second, the variance of observed counts data may exceed the mean due to unobserved heterogeneity. Ignoring over-dispersion and applying the standard Poisson regression for this data can cause underestimation of standard errors and p-values, therefore increasing the chance of an inflated Type 1 error. For these reasons, that is, the restrictive nature of equi-dispersion assumption in standard Poisson model, researchers have developed techniques and tests that allow detecting the over-dispersion (or under-dispersion) in the population (Mouatassim and Ezzahid, 2012)

The zero-inflated Poisson model is a random event containing excess zero-count data in unit time. For example, the number of insurance claims within a population for a certain type of risk would be zero-inflated by those people who have not taken out insurance against the risk and thus are unable to claim (Lambert, 1992). Further, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modeled independently (Long et al, 2006). When over-dispersion is a problem and the source of over-dispersion is the excess of zeroes, the Zero-inflated Poisson regression model typically shows a better model fit than the standard Poisson regression (Mouatassim and Ezzahid, 2012).

However the ZIP is not the only alternative option that allows for over-dispersion (Gurmu, 1991). The negative binomial regression model also allows for over-dispersion. Thus, the ZIP model has two parts; a Poisson count model and the logit model for predicting excess zeros. Furthermore, the ZIP model was used in several other studies such as in dental epidemiology (Böhning et al, 1999), occupational health (Lee et al, 2001) and children's growth and development (Cheung, 2002).

2.3.2 ZERO INFLATED NEGATIVE BINOMIAL REGRESSION (ZINB)

Another alternative method of analysing over-dispersed / excessive zero count outcome variables is the zero-inflated negative binomial model (Yau et al, 2003; Ridout et al, 2001). In practice, even after accounting for zero-inflation, the non-zero part of the count distribution is often over-dispersed. In this case, Greene (1994) described an extended version of the negative binomial

2.3.1 ZERO INFLATED POISSON REGRESSION (ZIP)

There are two main causes of equi-dispersion assumption violation. First, the frequency of the zero counts is more than the expected zero counts generated by the Poisson distribution. Second, the variance of observed counts data may exceed the mean due to unobserved heterogeneity. Ignoring over-dispersion and applying the standard Poisson regression for this data can cause underestimation of standard errors and p-values, therefore increasing the chance of an inflated Type 1 error. For these reasons, that is, the restrictive nature of equi-dispersion assumption in standard Poisson model, researchers have developed techniques and tests that allow detecting the over-dispersion (or under-dispersion) in the population (Mouatassim and Ezzahid, 2012)

The zero-inflated Poisson model is a random event containing excess zero-count data in unit time. For example, the number of insurance claims within a population for a certain type of risk would be zero-inflated by those people who have not taken out insurance against the risk and thus are unable to claim (Lambert, 1992). Further, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modeled independently (Long et al, 2006). When over-dispersion is a problem and the source of over-dispersion is the excess of zeroes, the Zero-inflated Poisson regression model typically shows a better model fit than the standard Poisson regression (Mouatassim and Ezzahid, 2012).

However the ZIP is not the only alternative option that allows for over-dispersion (Gurmu, 1991). The negative binomial regression model also allows for over-dispersion. Thus, the ZIP model has two parts; a Poisson count model and the logit model for predicting excess zeros. Furthermore, the ZIP model was used in several other studies such as in dental epidemiology (Böhning et al, 1999), occupational health (Lee et al, 2001) and children's growth and development (Cheung, 2002).

2.3.2 ZERO INFLATED NEGATIVE BINOMIAL REGRESSION (ZINB)

Another alternative method of analysing over-dispersed / excessive zero count outcome variables is the zero-inflated negative binomial model (Yau et al, 2003; Ridout et al, 2001). In practice, even after accounting for zero-inflation, the non-zero part of the count distribution is often over-dispersed. In this case, Greene (1994) described an extended version of the negative binomial

model for excess zero count data, the Zero-Inflated Negative Binomial (ZINB), which may be more appropriate than the ZIP. Zero-inflated Poisson regression does better when the data is not over-dispersed, i.e. when variance is not much larger than the mean. It has been established that the ZIP parameter estimates can be severely biased if the non-zero counts are over-dispersed in relation to the Poisson distribution.

The negative binomial distribution looks superficially similar to the Poisson but with a longer, fatter tail to the extent that the variance exceeds the mean. If the observed outcome is suspected to have variance larger than mean, the negative binomial distribution of the outcome is more appropriate than either the Poisson or normal distributions (Cameron and Trivedi, 1998) .The ZINB model can be extended to include random effects to directly model the correlation among the repeated measures within a subject (Yau et al, 2003).

Fang, (2008) applied the ZINB model to human micro biota sequence data, he revealed that the usefulness of the model is not limited to over-dispersed count data with excessive zeros alone, but also can be used for data with repeated measures. He also revealed that this method of estimation usually gives unbiased result compared to the Poisson model. Similarly, model for the over-dispersed count data with extra zeros was developed by Yau et al (2003). He used the single-level ZINB mixed regression to account for dependency between observations within the same cluster by including two random effects for the binary components and negative binomial. The two random effects are assumed to be independent and normally distributed for simplicity.

2.3.3 ZERO INFLATED GENERALIZED POISSON REGRESSION (ZIGP)

Zero-inflated generalized Poisson regression is a large class of regression models which contains ZIP, generalized Poisson (GP) and Poisson regression (Famoye and Singh, 2006). This class of regression models can handle over-dispersion and/or zero inflation, which count data often exhibit. A ZIGP distribution is defined analogously to a zero-inflated Poisson (ZIP) distribution (Mullahy (1986) with an additional zero-inflation parameter. The key advantage of a regression model based on the ZIGP distribution is that it allows for additional zero-inflation parameter in two ways for over-dispersion. The generalized Poisson (GP) distribution was first introduced by

Consul and Jain (1970). The distribution has one specific property, the variance is greater than, equal to or less than the mean (Famoye and Singh, 2006).

In a study on domestic violence, Famoye and Singh (2006) illustrated that in terms of the AIC statistic for the given application, every extension of the ZIGP regression model improved the model fit. Vuong tests were used to compare the non-nested models, both statistic (AIC and Vuong) selected the introduced ZIGP regression model as the model that best fits the data.

2.3.4 ZERO ALTERED POISSON REGRESSION (ZAP OR HURDLE)

The hurdle model is "a revised count model that has two generating processes not constrained to be equal (i.e. the positives and the zeros)." (Cameron and Trivedi, 1998). Mullahy (1986) has first discussed count data models. According to him, "The idea underlying the hurdle designs is that a binomial probability model rules the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the "hurdle is crossed", and the conditional distribution of the positives is governed by a truncated-at-zero count data model.

Due to the severe limitations of the Poisson distribution, other distributions can be used such as hurdle models (Boucher *et al.*, 2007). Hurdle models allow for a systematic variance in the statistical process governing observations below and above the hurdle. Explicitly, a hurdle model is mixed by a binary outcome of the count being below or above the hurdle (the selection variable), with a truncated model for outcomes above the hurdle which is why hurdle models are also sometimes called two-part models. The most essential usage of a hurdle count data model is the hurdle at zero. The hurdle at zero design can account for excess zeros. It means that this model can be used in situations where there are many zeros at the response variable. In this case, the hurdle at zero defines a probability function as the first part of the two-part models. The hurdle model is flexible and can handle both under and over-dispersion problem.

A generalized hurdle model is introduced by Gurmu (1998) for the analysis of over-dispersed or under-dispersed count data. Greene (2005) has applied the comparison between hurdle and zero-inflated models as two part-models. Gurmu and Trivedi (1996) also discussed a hurdle model to the annual number of recreational boating trips by a family. Three mixture models including a

hurdle model was applied by Dalrymple et al (2003). They claimed usefulness of the models in application to the incidence of sudden infant death syndrome (SIDS). Boucher, Denuit and Guillen (2007) compared generalized heterogeneous, zero-inflated, hurdle, and compound frequency models to the annual number of claims reported to the insurer. Saffari, Adnan and Greene (2011) argued the over-dispersion problem on count data using a right truncated Poisson regression model.

Mullahy (1986) discussed that the hurdle-at-zero model has both parts of the hurdle model based on probability functions for nonnegative integers. Some of the most popular hurdle model choices are nested models where both probability functions come from the same distribution, such as the Poisson distribution (Mullahy, 1986) or the Negative Binomial (Pohlmeier and Ulrich, 1995). However, non-nested models can also be used (Grootendorst, 1995; Gurmu, 1998; and Winkelmann, 2003).

Since models can be equivalent for certain parameter restrictions but are overlapping with suggested model by Vuong (1989), these models with a standard count distributions such as the Poisson or the Negative Binomial types do not nest. The interesting aspect of the hurdle model is to estimate the parameters by two separate steps. However, the zero-part parameters can be estimated using MLE on the first part of the likelihood function while the other parameters only use the second part, composed with only non-zero elements.

CHAPTER THREE

METHODOLOGY

3.1 STUDY AREA

The study area comprises the health facilities in all the 33 local government areas (LGAs) of Oyo State. The state is the second largest state in the South-West geopolitical zone of Nigeria with a land area of 27,148km^2 and a population of about 6million. Oyo state was carved out of the former Western state of Nigeria in 1976. Administratively, the state consists of 33 LGAs which includes: Afijio, Akinyele, Atiba, Atisbo, Egbeda, Ibadan North, Ibadan North-East, Ibadan North-West, Ibadan South-East, Ibadan South-West, Ibarapa Central, Ibarapa East, Ibarapa North, Ido, Irepo, Iseyin, Itesiwaju, Iwajowa, Kajola, Lagelu, Ogbomosho North, Ogbomosho South, Ogo-Oluwa, Olorunsogo, Oluyole, Ona-Ara, Orelope, Oriire, Oyo East, Oyo West, Saki East, Saki West and Surulere. It is bounded in the south by Ogun State and in the north by Kwara State. In the west, it is partly bounded by Ogun state and partly by the Republic of Benin, while it is bounded in the East by Osun State. The landscape consists of old hard rocks and dome shaped hills, which rise gently from about 500 meters in the southern part, reaching a height of about 1,219 meters above sea level in the northern part. Oyo state has an equatorial climate with dry and wet season and relatively high humidity. The dry season lasts from November to March while the wet season starts from April and ends in October. Average daily temperature ranges between 25°C (77°F) and 35°C (95.0°F), almost throughout the year. The vegetation pattern of Oyo State is that of rain forest in the south and guinea savanna in the north. The state operates three-tier health care systems which are primary, secondary and tertiary health centres across urban and rural areas. There are 1,648 health facilities disaggregated into 631 Primary Health centres (PHCs), 46 Secondary Health Facilities (SHFs), 5 Tertiary Health Centres (THCs) and 968 registered private health facilities (SMoH, 2008).

3.2 STUDY DESIGN

The dataset for this work was extracted from a longitudinal surveillance data of the Integrated Disease Surveillance and Response (IDSR) from Oyo State Ministry of health for the year 2011 to 2014. A secondary data analysis was conducted to achieve the study objectives.

3.3 STUDY POPULATION

The Oyo State Ministry of health's Integrated Disease Surveillance and Response (IDSR) dataset constitutes records of the 2007 to 2014 outpatients and inpatients aged 0 to 40 years and above from the 764 health facilities in 33 LGAs in the Oyo state. For the purpose of this study, all Oyo state residence are the study population while the number of people who were infected with typhoid fever from 2011 to 2014 are the target population.

3.4 DATA COLLECTION

Surveillance data on 40 diseases were collected on a monthly routine basis. The forty diseases comprise of thirty-two (32) communicable diseases and eight (8) non-communicable diseases from the 764 health facilities in the 33 LGAs across the state. Established cases of the outpatients and inpatients reviewed registries were reported to the local government disease surveillance and notification officers by the head of health facilities. The reported cases of diseases from 2007-2014 were collected through the Integrated disease surveillance case-based reporting forms which includes data on age of inpatients and outpatients (0-28days-40years and above) as well as the number of deaths. The surveillance dataset was then forwarded to the state epidemiologist at the Ministry of health. Data on the number of reported cases of waterborne disease (Typhoid fever) in all the age groups were extracted for the purpose of this study.

3.5 SAMPLE SIZE

For this study, a total of 2,970 typhoid fever cases across all age groups (0-28days to 40years and above) were extracted.

3.6 STUDY VARIABLES

3.6.1 Outcome Variable

The dependent/outcome variable for the study is the number of typhoid fever cases.

3.6.2 Explanatory Variables

The explanatory variables are:

1. Local government areas- defined by the 33 LGAs in Oyo state and labeled as LGA 1, LGA 2 ... LGA 32 and LGA 33.

2. The year of reporting defined as categorical variables labeled as Year 1 (2011), Year 2 (2012), Year 3 (2013), and Year 4 (2014).

3. Season of reporting- defined as month of reporting labeled January, February, March, April, May, June, July, August, September, October, November, and December.

3.7 DATA MANAGEMENT AND ANALYSIS

Descriptive statistics such as percentages, sample means and sample variances were performed for all independent variables. Mean and variance were determined to investigate the presence of under-dispersion or over-dispersion. The incidence rate ratios (IRR) were reported to examine the effect of the given exposure and approximate relative risk of re-emerging waterborne disease among all age groups over the years in all the 33 LGAs. All models (ZIP, ZINB, ZIGP, and ZAP) were compared and the best of the four models was selected using the Akaike information criteria (AIC) and -2logL.

3.8 UNDER-DISPERSION AND OVER-DISPERSION

In statistics, under-dispersion is the presence of smaller variation in the data than predicted, that is, when the anticipated variance is less than the mean. Conversely, over-dispersion is defined as the presence of greater variability in a dataset than expected based on a given statistical model. In other words, over-dispersion occurs when the observed variance is higher than the variance of a hypothetical model. In practice, populations are often heterogeneous (non-uniform) contrary to the assumption inherent within broadly used simple parametric models, this makes over-dispersion a very common feature in applied data analysis.

3.6.2 Explanatory Variables

The explanatory variables are:

1. Local government areas- defined by the 33 LGAs in Oyo state and labeled as LGA 1, LGA 2 ... LGA 32 and LGA 33.

2. The year of reporting defined as categorical variables labeled as Year 1 (2011), Year 2 (2012), Year 3 (2013), and Year 4 (2014).

3. Season of reporting- defined as month of reporting labeled January, February, March, April, May, June, July, August, September, October, November, and December.

3.7 DATA MANAGEMENT AND ANALYSIS

Descriptive statistics such as percentages, sample means and sample variances were performed for all independent variables. Mean and variance were determined to investigate the presence of under-dispersion or over-dispersion. The incidence rate ratios (IRR) were reported to examine the effect of the given exposure and approximate relative risk of re-emerging waterborne disease among all age groups over the years in all the 33 LGAs. All models (ZIP, ZINB, ZIGP, and ZAP) were compared and the best of the four models was selected using the Akaike information criteria (AIC) and -2logL.

3.8 UNDER-DISPERSION AND OVER-DISPERSION

In statistics, under-dispersion is the presence of smaller variation in the data than predicted, that is, when the anticipated variance is less than the mean. Conversely, over-dispersion is defined as the presence of greater variability in a dataset than expected based on a given statistical model. In other words, over-dispersion occurs when the observed variance is higher than the variance of a hypothetical model. In practice, populations are often heterogeneous (non-uniform) contrary to the assumption inherent within broadly used simple parametric models; this makes over-dispersion a very common feature in applied data analysis.

However, over-dispersion may be a result of; higher incidence of zero counts or subject heterogeneity. Over-dispersion is not an issue in ordinary regression when the population parameter is normal; this is because the normal distribution has a separate variance parameter, which is not common in count data.

3.8.1 INCIDENCE RATE RATIO

Incidence rate ratio (IRR) also sometimes called incidence density ratio is a relative difference measure used to compare the incidence rates of events occurring at any given point in time. It is most commonly used in analytic epidemiologic studies to examine causal association between a certain risk factor and an outcome.

3.9 STATISTICAL MODELS

The statistical models used in this study include:

(i) The Zero inflated Poisson (ZIP),
(ii) The Zero-inflated negative binomial (ZINB)
(iii) The Zero-inflated generalized Poisson (ZIGP)
(iv) The Zero-altered Poisson (ZAP or Hurdle)

3.9.1 ZERO-INFLATED REGRESSION MODEL

The main motivation for zero-inflated count models is that real-life data habitually show over-dispersion and redundant zeros (Lambert 1992; Greene 1994). Zero-inflated count models provide a way of modeling the excess zeros in addition to allowing for over-dispersion. In particular, for each observation, there are two possible data generation processes; the result of a Bernoulli trial determines which process is used. For observation i, Process 1 is chosen with probability φ_i and Process 2 with probability 1-φ_i . The first process generates only zero counts, whereas second process, $g(y_i|x_i)$, generates counts from either a Poisson or a negative binomial model. In general:

$$y_i \sim \begin{cases} 0 & with\ probability\ \varphi_i \\ g(y_i|x_i) & with\ probability\ 1-\varphi_i \end{cases}$$

Hence, the probability of $\{Y_{i=}y_i\}$ can be described as:

$$P(y_i|x_i) = \varphi_i + (1 - \varphi_i)\, g\,(y_i) \text{ for } y_i = 0$$

$$P(y_i|x_i) = (1 - \varphi_i)\, g\,(y_i) \text{ for } y_i \geq 1$$

Where g (y_i) follows either the Poisson or the negative binomial distribution.

When the probability φ_i is influenced by the characteristics of observation i, φ_i is a function of $z_i' \gamma$, where z_i' is the 1 x (q+1) the vector of zero-inflated covariates to be estimated, associated with the identified zero-inflated covariate vector $z_i' = (1, Z_1, \ldots, Z_q)$ and γ is the (q+1) x 1 vector of zero-inflated coefficients to be estimated. (The intercept for the zero-inflated is γ_o, while the coefficients for the q zero-inflated covariates are $\gamma_i, \gamma_{2,\ldots}\gamma_q$ and q is the number of the z covariates excluding the intercept). The parameter φ_i is the probability of zero counts from the binary process; it is frequently stated as the zero-inflation factor. This parameter φ_i is usually also characterized in terms of logistic regression model as logit $(\varphi_i) = z_i' \gamma$. The zero-inflated link function is F which is relating the product $z_i' \gamma$ (scalar) to the probability φ_i.

$$\varphi_i = F_i = F(z_i' \gamma)$$

The zero-inflated link function F can be specified as the logistic function;

$$F(z_i' \gamma) = \Lambda(z_i' \gamma) = \frac{exp(z_i' \gamma)}{1 + exp(z_i' \gamma)}$$

3.9.1.1 ZERO-INFLATED POISSON REGRESSION MODEL

The zero-inflated Poisson model (ZIP) is the first alternative method of modeling count data when applying the standard Poisson regression for such data can cause poor model fit, underestimation of standard errors and p-values, therefore increasing the chance of an inflated Type I error. This zero-inflated Poisson concerns a random event containing excess zero count data in unit time (Lambert, 1992). The zero-inflated Poisson (ZIP) model employs two components that correspond to two zero generating processes. The first process is governed by a

binary distribution that generates structural zeros. The second process is governed by a Poisson distribution that generates counts, some of which may be zero. The two model components are described as follows:

$\Pr(y_j=0)= \theta+(1-\theta)e^{-\lambda}$

$\Pr(y_j=a_i)=\frac{(1-\theta)\lambda^{a_i}e-\lambda}{a_i!}, a_i \geq 1$

Where,

y_j= outcome variable which has any non-negative integer value

λ_i = the expected Poisson count for the ith individual

θ_i = the probability of extra zeros

The mean is $(1-\theta)$, and the variance is

$$\lambda(1-\theta)(1+\lambda\theta)$$

The probability distribution of a zero-inflated Poisson random variable Y is given by;

$$\Pr(Y=y) = \begin{cases} \theta + (1-\theta)e-\lambda \ for\ y=0 \\ \frac{(1-\theta)\lambda^{y}e-\lambda}{y!} \ for\ y=1,2,\dots \end{cases}$$

The mean is given by;

$E(Y) =\mu= \lambda(1-\theta)$

$Var(Y) = \mu+\frac{\theta}{1-\theta}\mu^2$

The parameters λ and θ can be modeled as functions of linear predictors,

$h(\theta_i) = z_i'\gamma$

$g(\lambda_i) = x_i'\beta$

where h is one of the binary link functions logit or probit.

When $\theta=0$, the ZIP reduces to the classical Poisson model, otherwise, the ZIP is over-dispersed because the variance exceeds the mean

The over-dispersion is not due to the heterogeneity of the data which can be handled using the negative binomial model. Instead, it arises from the splitting of the data into the two statistical processes because of the excess zeros. According to Lambert we can model;

$\theta_i Z_i = \frac{\exp(z_i' \gamma)}{1+\exp(z_i' \gamma)}$

Z_i is a vector of covariates that defines the probability, θ_i and γ are the vector of corresponding elements of x_i and logit model can be substituted by probit specification. The parameter θ_i can also be related to λ_i.

Assume $y_i,\ldots,y_n$ are independent and θ_i is not related to λ_i. The likelihood function of (y_i) is as follows:

$$L=\prod_{y_i}[\ (\theta_i Z_i) + (1-\theta_i (z_i))\ e^{(-\lambda_i)}]\prod_{y_{i\neq 0}}[(1-\theta_i (z_i))\ \frac{\lambda i^{yi} e-\lambda i}{yi!}]$$

The log likelihood function is given as;

$$LL=\sum_{y_i=0} \log\left[e^{z_i' \gamma} + \exp\left(-e^{x_i'\beta}\right)\right) + \sum_{y_i\neq 0}[y_i\ x_i'\beta - e^{(x_i'\beta)} - \log(y_i!)] - \sum_{i=1}^{n} \log([1 + \exp(z_i' \gamma)])$$

3.9.1.2 ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL

The zero-inflated negative binomial (ZINB) is a combination distribution, analogous to ZIP distribution. The probability p for excess zeros (the distribution that takes only the value zero; 'impeccable state') and probability (1-p) for the rest of the counts (the distribution on the non-negative integers (i.e., including the value zero; 'unsatisfactory state') followed negative binomial distribution. The negative binomial distribution allows the Poisson mean to be distributed as Gamma, as a way to model over-dispersion. When modeling ungrouped count data, there are two commonly used mixture distribution for the unsatisfactory state. If the distribution for the unsatisfactory state is the Poisson, the mixture distribution is the zero-inflated Poisson (ZIP), and if the distribution for the unsatisfactory state is the negative binomial (NB), which is given by:

$$P(Y=y) = \frac{\Gamma(y+\tau)}{y!\Gamma\tau}\ (\frac{\tau}{\lambda+\tau})^{\tau}\ (\frac{\lambda}{\lambda+\tau})^{y} \qquad y=0,\ 1,\ 2,..$$

$$E(Y) = \lambda$$

$$V(Y) = \lambda + \lambda^2/\tau$$

Y is the response variable of interest, that is, number of cases of waterborne diseases in this research.

τ, is the shape parameter that quantifies the amount of over-dispersion

It can be observed that the negative binomial converges to a Poisson distribution when there is no over-dispersion (i.e. when τ approaches∞).

Thus, the zero-inflated negative binomial distribution is expressed as:

$$P(Y=y)=\begin{cases} p+(1-p)(1+\lambda/\tau)^{-\tau}, & y=0 \\ (1-p)\frac{\Gamma(y+\tau)}{y!\Gamma\tau}(1+\lambda/\tau)^{-\tau}(1+\tau/\lambda)^{-y} & \end{cases}$$

y=1, 2…

The mean and variance of ZINB is given as:

E(Y) = (1-p) λ V(Y) = (1-p) λ (1-pλ+λ/τ)

Also, the ZINB distribution approaches the ZIP distribution and the negative binomial distribution as $p\rightarrow 0\ and\ \tau\rightarrow\infty$. If both the 1/τ and p≈ 0, the ZINB distribution practically reduces to Poisson distribution.

The ZINB regression model relates p & λ to covariate matrix M and N with regression parameters β and γ as follows:

Log (λ_i) =mβ and logit (p_i) =nγ i=1, 2…h

Given the observed data, the log-likelihood function of the ZINB is given as:

LL=l($\beta, \gamma, \tau; y, n, m$)=

$$\sum_{i=1}^{h}\log\left(1+e^{n_i\gamma}-\sum_{i=1,y_i=0}^{h}log\,(e^{n_i\gamma}+(\frac{e^{m_i\beta}+\gamma}{\tau})^{-\tau}\right)$$

$$+\sum_{i=1,y_i>0}^{h}(\tau\log\left(\frac{e^{m_i\beta}+\tau}{\tau}\right)+y_i\log(1+e^{m_i\beta}\tau)$$

$$+\sum_{i=1,y_i>0}^{h}(log\Gamma(\tau)+log\Gamma(1+y_i)-log\Gamma(\tau+y_i)$$

3.9.1.3 THE ZERO-INFLATED GENERALIZED POISSON MODEL

The generalized Poisson regression (GPR) model is a natural extension of the Poisson regression model (Frome et al, 1973). Let the response variable y_i=1, 2…h be the number of typhoid fever cases. The GPR model is given by:

$$f(\mu, \alpha, y_i)=(\frac{\mu_i}{1+\alpha\mu_i})^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!}\exp\left[\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right]$$

τ, is the shape parameter that quantifies the amount of over-dispersion

It can be observed that the negative binomial converges to a Poisson distribution when there is no over-dispersion (i.e. when τ approaches∞).

Thus, the zero-inflated negative binomial distribution is expressed as:

$$P(Y=y)=\begin{cases} p+(1-p)(1+\lambda/\tau)^{-\tau}, & y=0 \\ (1-p)\frac{\Gamma(y+\tau)}{y!\Gamma\tau}(1+\lambda/\tau)^{-\tau}(1+\tau/\lambda)^{-y} & \end{cases}$$

y=1, 2…

The mean and variance of ZINB is given as:

E(Y) = (1-p) λ V(Y) = (1-p) λ (1-pλ+λ/τ)

Also, the ZINB distribution approaches the ZIP distribution and the negative binomial distribution as $p \to 0$ *and* $\tau \to \infty$. If both the 1/τ and p≈ 0, the ZINB distribution practically reduces to Poisson distribution.

The ZINB regression model relates p & λ to covariate matrix M and N with regression parameters β and γ as follows:

Log (λ_i) =mβ and logit (p_i) =nγ i=1, 2…h

Given the observed data, the log-likelihood function of the ZINB is given as:

LL=l(β, γ, τ; y, n, m)=

$$\sum_{i=1}^{h}\log\left(1+e^{n_i\gamma}-\sum_{i=1,y_i=0}^{h}log\,(e^{n_i\gamma}+(\frac{e^{m_i\beta}+\gamma}{\tau})^{-\tau}\right)$$

$$+\sum_{i=1,y_i>0}^{h}(\tau\log\left(\frac{e^{m_i\beta}+\tau}{\tau}\right)+y_i\log(1+e^{m_i\beta}\tau)$$

$$+\sum_{i=1,y_i>0}^{h}(log\Gamma(\tau)+log\Gamma(1+y_i)-log\Gamma(\tau+y_i)$$

3.9.1.3 THE ZERO-INFLATED GENERALIZED POISSON MODEL

The generalized Poisson regression (GPR) model is a natural extension of the Poisson regression model (Frome et al, 1973). Let the response variable y_i=1, 2…h be the number of typhoid fever cases. The GPR model is given by:

$$f(\mu, \alpha, y_i)=(\frac{\mu_i}{1+\alpha\mu_i})^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!}\exp\left[\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right]$$

The mean and variance is defined as:

E (y_i) = μ

Var (y_i) = $\mu_i(1 + \alpha\mu_i^2)$

More generally, the mean of y_i can also be written as E ($y_i|x_i$) = $\mu_i x_i = c_i\Lambda(x_i,\beta)$ where; $\Lambda(x_i,\beta)$ is a known link function of x_i *and* β_i which can be differentiated with respect to β, $\mu_i = \mu_i(x_i)$, $x_i$is the ith covariate matrix X defines as ($x_{i1} = 1, x_{i2}, \ldots, x_{ik}$), and c_i is a measure of exposure.

A ZIGP distribution is defined analogously to a ZIP distribution with an additional zero-inflated parameter. If y_i are independent random variable having a ZIGP distribution, the zeros are assumed to occur in two distinct states. The only occurrences at the first state are zeros referred to as 'Structural zero' which occurs with probability φ_i. The second state are called 'Sampling zero', they occur with probability (1 - φ_i) and leads to a generalized Poisson (GP) distribution with α and τ_i. Thus, the second state process leads to a component mixture distribution with probability mass function in ZIGP model defined as;

P(Y=y|x_i, v_i) = φ_i+ (1- φ_i) f (τ,α_i, 0), y_i=0

= (1- φ_i) f (τ_i,α_i,y_i), $y_i > 0$

Where f (τ_i,α_i, y_i), y_i = 0,1,2 ...is the GPR model and 0<φ_i < 1.

τ_i *and* φ_i are non-negative/ linear functions of some covariates that can be modeled via logit and link functions respectively.

The mean and variance of the *i*th observation Y_i of the ZIGP are respectively given as;

E ($y_i|x_i$) = (1- φ_i) $\tau_i(x_i)$

V ($y_i|x_i$) = (1- φ_i) [τ_i^2+$\tau_i(1 + \alpha\tau_i)^2$] - (1- φ_i)

=E ($y_i|x_i$) [$(1 + \alpha\tau_i)^2 + \varphi_i\tau_i$]

The distribution of $y_i$exhibits over-dispersion when $\varphi_i > 0$. The ZIGP model reduces to the GPR when φ_i=0 and reduces to ZIP model when α=0 (Lambert, 1992).

The log-likelihood function of the ZIGP regression model is given by;

LL= $\sum_{i=1}^{n} \log(1 + \tau_i^{-\mu}) + \sum_{y_i=0} log(\tau_i^{-\mu}$+exp [-$\tau_i$/ (1+α$\tau_i$)]) +

$$\sum_{y_i>0}\left\{y_i \log\left[\frac{\tau_i}{1+\alpha\tau_i}\right] + (y_i - 1)\log(1 + \alpha y_i) - log(y_i!) - \tau_i\frac{(1+\alpha y_i)}{1+(\alpha\tau_i)}\right\}$$

The mean and variance is defined as:

E (y_i) = μ

Var (y_i) = $\mu_i(1 + \alpha\mu_i^2)$

More generally, the mean of y_i can also be written as E ($y_i|x_i$) = $\mu_i x_i$= $c_i\Lambda(x_i,\beta)$ where; $\Lambda(x_i,\beta)$ is a known link function of x_i *and* β_i which can be differentiated with respect to β, μ_i = $\mu_i(x_i)$, $x_i$is the ith covariate matrix X defines as ($x_{i1} = 1, x_{i2}, \ldots, x_{ik}$), and c_i is a measure of exposure.

A ZIGP distribution is defined analogously to a ZIP distribution with an additional zero-inflated parameter. If y_i are independent random variable having a ZIGP distribution, the zeros are assumed to occur in two distinct states. The only occurrences at the first state are zeros referred to as 'Structural zero' which occurs with probability φ_i. The second state are called 'Sampling zero', they occur with probability (1 - φ_i) and leads to a generalized Poisson (GP) distribution with α and τ_i. Thus, the second state process leads to a component mixture distribution with probability mass function in ZIGP model defined as;

P(Y=y|x_i, v_i) = φ_i+ (1- φ_i) f (τ,α_i, 0), y_i=0

= (1- φ_i) f (τ_i,α_i,y_i), $y_i > 0$

Where f (τ_i,α_i, y_i), y_i = 0,1,2 ...is the GPR model and 0<φ_i < 1.

τ_i *and* φ_i are non-negative/ linear functions of some covariates that can be modeled via logit and link functions respectively.

The mean and variance of the *i*th observation Y_i of the ZIGP are respectively given as;

E ($y_i|x_i$) = (1- φ_i) $\tau_i(x_i)$

V ($y_i|x_i$) = (1- φ_i) [τ_i^2+$\tau_i(1 + \alpha\tau_i)^2$] - (1- φ_i)

=E ($y_i|x_i$) [$(1 + \alpha\tau_i)^2 + \varphi_i\tau_i$]

The distribution of $y_i$exhibits over-dispersion when $\varphi_i > 0$. The ZIGP model reduces to the GPR when φ_i=0 and reduces to ZIP model when α=0 (Lambert, 1992).

The log-likelihood function of the ZIGP regression model is given by;

LL= $\sum_{i=1}^{n} \log(1 + \tau_i^{-\mu}) + \sum_{y_i=0} log(\tau_i^{-\mu}$+exp [-$\tau_i$/ (1+$\alpha\tau_i$)]) +

$$\sum_{y_i>0} \left\{ y_i \log\left[\frac{\tau_i}{1+\alpha\tau_i}\right] + (y_i - 1)\log(1 + \alpha y_i) - log(y_i!) - \tau_i \frac{(1+\alpha y_i)}{1+(\alpha\tau_i)} \right\}$$

3.9.1.4 THE ZERO-ALTERED POISSON OR HURDLE MODEL

The zero-altered Poisson (ZAP) or hurdle model is a revised count model in which the process generating the zeros and the positives are not constrained to be the same (Cameron and Trivedi, 1998). Hurdle models allow for a logical modification in the statistical process governing observations below or above the hurdle. Particularly, a hurdle model is mixed by a binary outcome of the count being below or above the hurdle (the selection variable), with a truncated model for outcomes above the hurdle. That is why hurdle models are occasionally also called two-part models. The essential notion of the hurdle developments is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the "hurdle is crossed", and the conditional distribution of the positives is governed by a truncated-at-zero count data model (Mullahy, 1986). Thus, a hurdle model is flexible and can handle both under- and over-dispersion problem.

The most important usage of a hurdle count data model is the hurdle at zero which can be used in situations where there are many zeros at the response variable. Therefore, the hurdle at zero defines a probability (Pr ($Y = 0$)) as the first part of the two- part models. The hurdle model interestingly estimates parameters by two separate steps. In fact, the zero-part parameters can be estimated using MLE on the first part of the likelihood function while the other parameters only use the second part, only composed with non-zero elements. The zero-altered Poisson process has a probability mass function;

$$\Pr(Y=y|Y\neq 0)=\begin{cases} \dfrac{\lambda^{y}}{(e^{\lambda}-1)y!} & y=1,2,3\ldots \\ 0, \quad otherwise & \end{cases}$$

The log-likelihood can be written;

$$LL=\ln\left\{\prod_{i\in\Omega_o}(e^{-e^{x_i\beta_i}})\left\{\prod_{i\in\Omega_1}(e^{-e^{x_i\beta_i}})\prod_{i\in\Omega_i}\frac{e^{y_i x_i\beta_2}}{(e^{e^{x_i\beta_2}}-1)y_i}\right\}\right.$$

$$=\{\sum_{i\in\Omega_o}-e^{x_i\beta_1}+\sum_{i\in\Omega_i}\ln(1-e^{-e^{x_i\beta_i}})\}+\{\sum_{i\in\Omega_i}y_i x_i\beta_2-\sum_{i\in\Omega_i}\ln(e^{ex_i\beta_2}- \quad 1)-$$

$$\{\sum_{i\in\Omega_i}\ln(y_i!)\}$$

$$=\ln\{L_i(\beta_i)\}+\ln\{L_2(\beta_2)\}$$

Where,

3.9.1.4 THE ZERO-ALTERED POISSON OR HURDLE MODEL

The zero-altered Poisson (ZAP) or hurdle model is a revised count model in which the process generating the zeros and the positives are not constrained to be the same (Cameron and Trivedi, 1998). Hurdle models allow for a logical modification in the statistical process governing observations below or above the hurdle. Particularly, a hurdle model is mixed by a binary outcome of the count being below or above the hurdle (the selection variable), with a truncated model for outcomes above the hurdle. That is why hurdle models are occasionally also called two-part models. The essential notion of the hurdle developments is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the "hurdle is crossed", and the conditional distribution of the positives is governed by a truncated-at-zero count data model (Mullahy, 1986). Thus, a hurdle model is flexible and can handle both under- and over-dispersion problem.

The most important usage of a hurdle count data model is the hurdle at zero which can be used in situations where there are many zeros at the response variable. Therefore, the hurdle at zero defines a probability (Pr ($Y = 0$)) as the first part of the two- part models. The hurdle model interestingly estimates parameters by two separate steps. In fact, the zero-part parameters can be estimated using MLE on the first part of the likelihood function while the other parameters only use the second part, only composed with non-zero elements. The zero-altered Poisson process has a probability mass function;

$$\Pr(Y=y|Y\neq 0)=\begin{cases} \frac{\lambda^y}{(e^{\lambda}-1)y!} & y=1, 2, 3\ldots \\ 0, \quad otherwise & \end{cases}$$

The log-likelihood can be written;

$$LL=\ln\left\{\prod_{i\in\Omega_o}(e^{-e^{x_i\beta_i}})\left\{\prod_{i\in\Omega_1}(e^{-e^{x_i\beta_i}})\prod_{i\in\Omega_i}\frac{e^{y_i x_i\beta_2}}{(e^{e^{x_i\beta_2}}-1)y_i}\right\}\right.$$

$$=\{\sum_{i\in\Omega_o}-e^{x_i\beta_1}+\sum_{i\in\Omega_i}\ln(1-e^{-e^{x_i\beta_i}})\}+\{\sum_{i\in\Omega_i}y_i x_i\beta_2-\sum_{i\in\Omega_i}\ln(e^{e^{x_i\beta_2}}-1)-$$

$$\{\sum_{i\in\Omega_i}\ln(y_i!)\}$$

$$=\ln\{L_i(\beta_i)\}+\ln\{L_2(\beta_2)\}$$

Where,

ln L_i (β_i) is the log-likelihood for the binary outcome model and;

ln L_2 (β_2) is the log-likelihood for the zero-altered Poisson model.

Y is the response variable (re-emerging water-borne disease)

x_i are the independent variables (season, year and geographical locations)

3.10 MODEL SELECTION

The goodness of fit tests used for the selection of the best regression model includes;

3.10.1 -2log-likelihood Statistics

The -2log-likelihood statistics often referred to as deviance, is a quality of fit statistics for a model that is usually used for statistical hypothesis testing. It is used to compare two models using the likelihood function (L). The likelihood (L) is a function of the parameters of a statistical model given data. Likelihood is used to describe a function of a parameter for a given outcome. -2log L is used to measure agreement between the data and the fitted model, hence, the preferred model is the model that has a smaller -2logL.

3.10.2 Akaike Information Criterion (AIC)

The Akaike information criterion (AIC) was first developed by Hirotugu Akaike (1973) to measure relative quality of statistical models for a given set of data. AIC estimates the quality of a set of models relative to each of the other models for the given data. Hence, AIC provides a means for model selection.

AIC is established on information principle that offers relative estimate of lost information when the generated data process is represented by a given model. Thus, the AIC penalizes the log-likelihood with regard to the number of estimated parameters; it deals with the trade-off between the complexity and the goodness of fit of the model. The AIC is not a hypothesis test, does not have a p-value, and does not provide anything about the quality of the model in an absolute sense. Instead, the AIC focuses on the strength of evidence and gives a measure of uncertainty for each model. The AIC value in a given statistical model is defined by;

$$AIC = 2k - 2\ln(L)$$

Where,

L is the maximum likelihood function of the model,

k is the number of parameters to be estimated in the model

AIC includes a penalty that discourages over-fitting (increasing function of the number of estimated parameters) and rewards goodness of fit (as assessed by the likelihood function). (Bozdogan, 2000). Hence, the preferred model is the one with the minimum AIC value.

CHAPTER FOUR

RESULTS

4.1 Descriptive Statistics of Typhoid fever cases in Oyo state (2011-2014)

Table 4.1 below shows the descriptive statistics for typhoid fever cases in Oyo state by year. The mean number of cases in the population for the four years was 4.40 with variance of 10007.21. In addition, the mean number of cases was 3.77 (VAR= 93.11), 5.63 (VAR=308.12), 3.65 (VAR=141.29), and0.64 (VAR=9.99) for 2011, 2012, 2013, and 2014 respectively.

Table 4.1: Descriptive statistics of typhoid fever cases in Oyo state by year

Year	Mean	Variance	Min	Max	No of zero cases (%)	No of non-zero cases (%)
2011(1)`	3.77	93.11	0	98	1803(65.0%)	969(35.0%)
2012(2)	5.63	308.12	0	480	1718(62.0%)	1054(38.0%)
2013(3)	3.65	141.29	0	216	2068(74.6%)	704(25.4%)
2014(4)	0.64	9.99	0	58	2529(91.2%)	243(8.8%)

4.1.2 Descriptive Statistics of Typhoid fever cases in Oyo state by Month (2011 to 2014)

Table 4.2 below shows the descriptive statistics for typhoid fever cases in Oyo state by month. The number of typhoid fever incidence was reported highest in February with 336 cases and the lowest was in October with 101 cases.

Table 4.2: Descriptive statistics of typhoid in Oyo state by Month

Month	Mean	Variance	Min	Max	No of zero cases (%)	No of non-zero cases (%)
Jan	3.81	109.79	0	102	618(66.9%)	306(33.1%)
Feb	4.68	165.31	0	100	588(63.6%)	336(36.4%
Mar	4.09	157.05	0	216	589(63.7%)	335(36.3%)
Apr	3.28	114.58	0	119	667(72.2%)	257(27.8%)
May	4.06	122.23	0	107	638(69.0%)	286(31.0%)
Jun	4.44	131.08	0	117	636(68.8%)	288(31.2%)
Jul	3.84	128.40	0	97	654(70.8%)	270(29.2%)
Aug	4.10	129.91	0	138	643(69.6%)	281(30.4%)
Sep	2.54	126.21	0	141	736(79.7%)	188(20.3%)
Oct	1.40	110.78	0	93	823(89.1%)	101(10.9%)
Nov	2.39	94.35	0	78	752(81.4%)	172(18.6%)
Dec	2.87	149.70	0	480	774(83.8%)	150(16.2%)

4.1.3 Distribution of Typhoid fever cases in Oyo state (2011 to 2014)

Figure 4.1 below shows the distribution of typhoid fever cases across the state over the year 2011 to 2014. The dataset consists of 2,970 (26. 8%)reported cases of typhoid within the year 2011 to 2014 and 8,118 (73. 2%) non cases (i.e. zero cases).

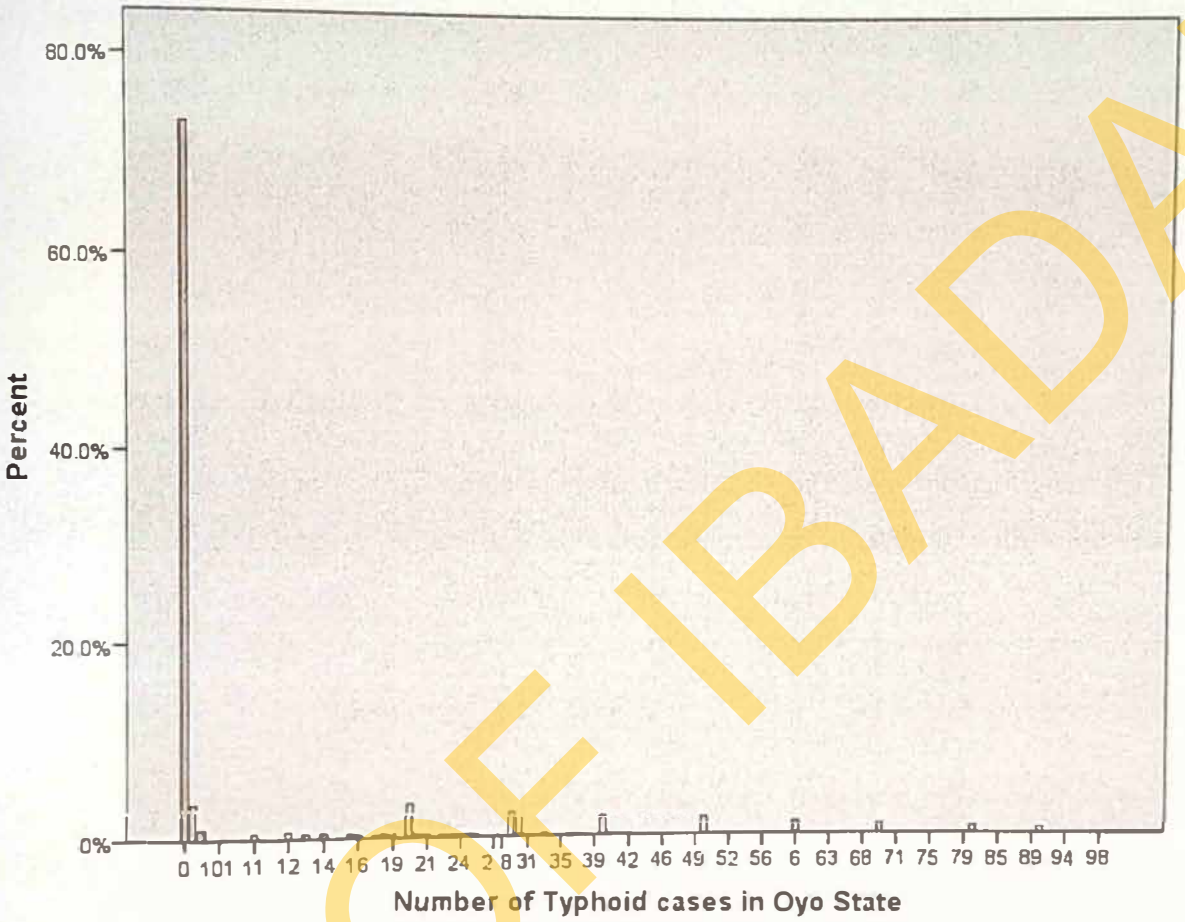

Figure 4. 1:Distribution of Typhoid fever cases in Oyo state in the year 2011 to 2014

4.1.4 Distribution of Typhoid fever cases in Oyo State by Age group.

Figure 4.2 shows the distribution of typhoid fever cases by age group. Those in age group 40 and above has the highest reported incidence of typhoid fever (480) with cases between 2011 and 2014, followed by age group 20 to 40 with 197 reported cases, Also, those in age group 0 to 28 days has the lowest reported incidence of the disease (14).

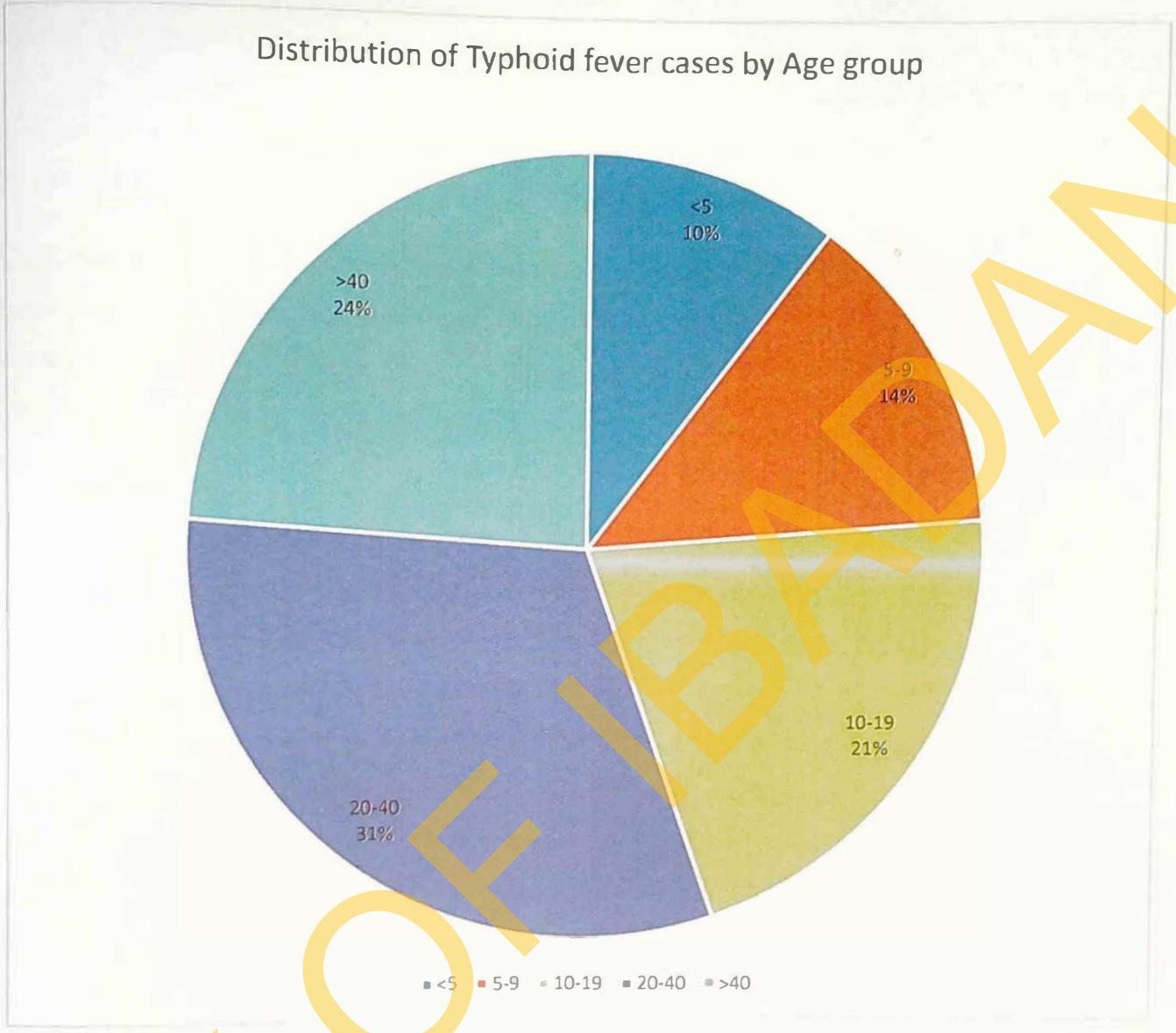

Figure 4.2 Distribution of Typhoid fever cases in Oyo state by Age group

4.2 Effect of Month, Year, LGA and Age group on the occurrence of Typhoid fever cases in Oyo state using the ZINB model

Table 4.3 shows the regression parameters for the zero inflated negative binomial regression 0.308,-0.301), (2013: IRR=0.810, 95% CI=-0.349,-0.735), (2014: IRR=0.742, 95%CI=-0.436,-0.160), compared to year 2011.

There was a significant overall reduction in the reported cases of typhoid fever in February. The risk of typhoid fever decreased in the state by 18.2% in February (IRR=0.818, 95%CI:-0.398, -0.006). There was an excess risk of typhoid in the state by 13.7% in September (IRR=1.137, 95%CI:-0.099, 0.357), 9.3% in October (IRR=1.093, 95%CI:-0.189, 0.367), 12.6% in November (IRR=1.126, 95%CI:-0.109, 0.347) and 1.4% in December (IRR=1.014, 95%CI:-0.221, 0.248) compared to April.

The risk of typhoid fever was highest in Lagelu LGA with an excess risk of 7.2%. There was 4 times higher risk of typhoid compared to Afijio (IRR=4.072, 95% CI: 1.046, 1.762). Also, the odd of typhoid was 3 times higher in Oriire with an excess risk of 18.1% (IRR=3.181, 95% CI: 0.673, 1.642). In addition, there was 2 times excess risk of typhoid in Ibarapa Central by 38.3%, Onaara by 37.3%, and Iwajowa by 2.0% ((IRR=2.383, 95%CI:0.539, 1.198); (IRR=2.373, 95%CI:0.515, 1.212); (IRR=2.02, 95%CI: 0.358, 1.048) respectively. Furthermore, there was an increased risk of typhoid in Ibadan South west by 46.7% and Ibarapa East by 44.8% (IRR=1.467, 95% CI=0.034, 0.732); (IRR=1.448, 95% CI=0.043, 0.698). However, people from Iseyin and Oyo west had the lowest risk of typhoid fever compared to those from Afijio (IRR=-0.715, 95% CI=--0.666,-0.003); (IRR=-0.425, 95% CI=--1.427,-0.284). The risk of typhoid fever was lowest by 82.8% (IRR=0.828, 95% CI=-0.307, -0.070) among children 0-28 days and highest by 19.3% among those aged 20-40 years (IRR=1.193, 95% CI=0.150, 0.204)

From the zero inflated part, risk of typhoid fever decreased in Oyo state by 13.1% (IRR=0.869, 95% CI=-0.279, -0.002) and 15.2% (IRR=0.848, 95% CI=-0.303, -0.027) in year 2013 and 2014 respectively compared to 2011.

Lower risk of typhoid fever were reported in January, February, and March by 31.6% (IRR=0.684, 95% CI=-0.602, -0.157), 33.8% (IRR=0.662, 95% CI=-0.635, -0.19) and 29.5% (IRR=0.705, 95% CI=-0.573, -0.127). While higher risk of typhoid fever were recorded in the

state in September by 66.7% (IRR=1.667, 95% CI=0.265, 0.757), November (IRR=1.667, 95% CI=0.265, 0.757), and December by 86.9% (IRR=1.869, 95% CI=0.374, 0.877). The highest risk of typhoid was reported in October, there was 3 times more incidence of typhoid fever in this month (IRR=3.210, 95% CI=0.882, 1.450) compared to April

Typhoid fever risk decreased in the following LGAs; Atiba: 33.5% (IRR=0.665, 95% CI=-0.787, - 0.03), Ibadan South East: 38.4% (IRR=0.616, 95% CI=-0.863, - 0.106), Ibadan North-East: 39.7% (IRR=0.603, 95% CI=-0.884, - 0.128), Ibadan North-west: 41.2% (IRR=0.588, 95% CI=-0.909, -0.153), Iwajowa: 43.1% (IRR=0.569, 95% CI=-0.938,-0.191), Ona-Ara: 44.0% (IRR=0.66, 95% CI=-0.794,-0.038), Ibadan South-west: 44.7% (IRR=0.553, 95% CI=-0.967, -0.219), Irepo:45.4% (IRR=0.546, 95% CI=-0.982,-0.23), Ibadan North: 46.6% (IRR=0.534, 95% CI=-1.007, -0.249), Ibarapa central: 49.3% (IRR=0.507, 95% CI=-1.049, -0.309), Iseyin: 53.5% (IRR=0.465, 95% CI=-1.14,-0.392), Itesiwaju: 56.7% (IRR=0.443, 95% CI=-1.185, -0.443), Ido: 58.1% (IRR=0.419, 95% CI=-1.14, -0.498), Ibarapa East: 61.0% (IRR=0.390, 95% CI=-1.309, -0.576),

Finally, risk of typhoid fever increased in Atisbo and Akinyele by 54.0% and 54.6% (IRR=1.54, 95% CI=0.003, 0.861); (IRR=1.546, 95% CI=0.004, 0.868) respectively. Also, there was 2 times increase in the risk of typhoid fever in Saki west, Orelope, Oriire and Saki east by 42.6%, 57.9%, 75.0%, and 92.5% (IRR=2.426, 95% CI=0.408, 1.365); (IRR=2.579, 95% CI=0.465, 1.43); (IRR=2.75, 95% CI=0.525, 1.499); (IRR=2.925, 95% CI=0.565, 1.581) respectively. Furthermore, the risk of typhoid was 3 times higher in Oyo East and Oyo west by 56.2% and 89.0%, (IRR=3.562, 95% CI=0.743, 1.798); (IRR=3.89, 95% CI=0.799, 1.918). Lastly, the odds of typhoid was 4 times higher in Surulere by 30.7% (IRR=4.307, 95% CI=0.892, 2.028).All compared to Afijio LGA in a certain zero group.

Finally, there was an exaggerated increased risk of typhoid fever among children 0-28 days with 121 times higher risk (IRR=121.389, 95% CI=4.280, 5.318). Also, children 1-11 months had 9 times higher risk of typhoid fever (IRR=9.478, 95% CI=2.010, 2.488) while there was 70.1% decreased risk among those aged 10-19 years (IRR=0.701, 95% CI=-0.539, -0.170) compared to people older than 40 years.

Table 4.3 Parameter Estimates of the Zero-inflated Negative binomial for Typhoid fever cases in Oyo state.

	ZINB				
Parameters	**IRR**	**Standard Error**	**95% CI for IRR**		**P-value**
			Lower bound	**Upper bound**	
Intercept	27.653	0.161	3.003	3.636	<0.001
2011(Year1)*					
2012(Year2)	0.844	0.071	-0.308	-0.031	0.017
2013(Year3)	0.810	0.070	-0.349	-0.735	0.002
2014(Year4)	0.742	0.070	-0.436	-0.160	<0.001
Jan	0.852	0.100	-0.357	0.037	0.111
Feb	0.818	0.100	-0.398	-0.006	0.043
Mar	0.915	0.101	-0.287	0.108	0.37
April*					
May	1.137	0.102	-0.072	0.328	0.209
Jun	1.058	0.102	-0.145	0.257	0.583
Jul	1.183	0.103	-0.032	0.369	0.101
Aug	1.099	0.102	-0.106	0.296	0.356
Sep	1.137	0.116	-0.099	0.3576	0.269
Oct	1.093	0.114	-0.189	0.367	0.531
Nov	1.126	0.116	-0.109	0.347	0.307
Dec	1.014	0.120	-0.221	0.248	0.910
Afijio*					

Akinyele	0.814	0.211	-0.618	0.207	0.329
Atiba	1.814	0.175	0.252	0.939	<0.001
Atisbo	1.290	0.212	-0.162	0.671	0.230
Egbeda	0.869	0.187	-0.507	0.225	0.451
Ibadan north	0.599	0.176	-0.857	-0.167	0.003
Ibadan north east	0.988	0.178	-0.360	0.336	0.944
Ibadan north west	0.850	0.174	-0.503	0.178	0.349
Ibadan south east	1.052	0.174	-0.291	0.393	0.770
Ibadan south west	1.467	0.178	0.034	0.732	0.031
Ibarapa central	2.383	0.168	0.539	1.198	<0.001
Ibarapa east	1.448	0.167	0.043	0.698	0.027
Ibarapa north	0.910	0.178	-0.443	0.255	0.597
Ido	0.805	0.169	-0.549	0.115	0.201
Irepo	0.954	0.174	-0.389	0.295	0.787
Iseyin	0.715	0.169	-0.666	-0.003	0.048
Itesiwaju	0.947	0.170	-0.388	0.279	0.749
Iwajowa	2.020	0.176	0.358	1.048	<0.001
Kajola	0.906	0.187	-0.466	0.269	0.600
Lagelu	4.072	0.183	1.046	1.762	<0.001
Ogbomoso North	0.762	0.190	-0.644	0.099	0.151
Ogbomoso South	0.762	0.188	-0.640	0.097	0.149
OgoOluwa	0.982	0.184	-0.379	0.343	0.921
Olorunsogo	0.746	0.181	-0.649	0.062	0.105
Oluyole	1.211	0.188	-0.178	0.560	0.310
Onaara	2.373	0.178	0.515	1.212	<0.001
Orelope	1.222	0.243	-0.277	0.678	0.410
Orire	3.181	0.247	0.673	1.642	<0.001
Oyo east	1.574	0.268	-0.072	0.980	0.091
Oyo west	0.425	0.292	-1.427	-0.284	0.003
Saki east	0.567	0.261	-1.079	-0.057	0.294
Saki west	0.776	0.244	-0.731	0.224	0.298

Akinyele	0.814	0.211	-0.618	0.207	0.329
Atiba	1.814	0.175	0.252	0.939	<0.001
Atisbo	1.290	0.212	-0.162	0.671	0.230
Egbeda	0.869	0.187	-0.507	0.225	0.451
Ibadan north	0.599	0.176	-0.857	-0.167	0.003
Ibadan north east	0.988	0.178	-0.360	0.336	0.944
Ibadan north west	0.850	0.174	-0.503	0.178	0.349
Ibadan south east	1.052	0.174	-0.291	0.393	0.770
Ibadan south west	1.467	0.178	0.034	0.732	0.031
Ibarapa central	2.383	0.168	0.539	1.198	<0.001
Ibarapa east	1.448	0.167	0.043	0.698	0.027
Ibarapa north	0.910	0.178	-0.443	0.255	0.597
Ido	0.805	0.169	-0.549	0.115	0.201
Irepo	0.954	0.174	-0.389	0.295	0.787
Iseyin	0.715	0.169	-0.666	-0.003	0.048
Itesiwaju	0.947	0.170	-0.388	0.279	0.749
Iwajowa	2.020	0.176	0.358	1.048	<0.001
Kajola	0.906	0.187	-0.466	0.269	0.600
Lagelu	4.072	0.183	1.046	1.762	<0.001
Ogbomoso North	0.762	0.190	-0.644	0.099	0.151
Ogbomoso South	0.762	0.188	-0.640	0.097	0.149
OgoOluwa	0.982	0.184	-0.379	0.343	0.921
Olorunsogo	0.746	0.181	-0.649	0.062	0.105
Oluyole	1.211	0.188	-0.178	0.560	0.310
Onaara	2.373	0.178	0.515	1.212	<0.001
Orelope	1.222	0.243	-0.277	0.678	0.410
Orire	3.181	0.247	0.673	1.642	<0.001
Oyo east	1.574	0.268	-0.072	0.980	0.091
Oyo west	0.425	0.292	-1.427	-0.284	0.003
Saki east	0.567	0.261	-1.079	-0.057	0.294
Saki west	0.776	0.244	-0.731	0.224	0.298

Surulere	0.712	0.295	-0.918	0.238	0.249
Agegrp 0-28	0.828	0.060	-0.307	-0.070	0.002
Agegrp 1-11	0.275	0.015	-1.3707	-1.214	<0.001
Agegrp 12-59	0.421	0.022	-0.907	-0.822	<0.001
Agegrp 5-9	0.591	0.018	-.0561	-0.492	<0.001
Agegrp 10-19	0.825	0.015	-0.222	-0.161	<0.001
Agegrp 20-40	1.193	0.014	0.150	0.204	<0.001
Agegrp >40*					
Inf.Intercept	4.413	0.171	1.150	1.819	<0.001
Inf.2011(Year1)*					
Inf.2012(Year2)	0.827	0.070	-0.328	-0.052	0.007
Inf.2013(Year3)	0.869	0.071	-0.279	-0.002	0.047
Inf.2014(Year4)	0.848	0.071	-0.303	-0.027	0.019
Inf.Jan	0.684	0.114	-0.602	-0.157	<0.001
Inf.Feb	0.662	0.113	-0.635	-0.190	<0.001
Inf.Mar	0.705	0.114	-0.573	-0.127	0.002
Inf.April*					
Inf.May	0.962	0.116	-0.267	0.189	0.736
Inf.Jun	0.970	0.117	-0.259	0.198	0.793
Inf.Jul	0.989	0.117	-0.240	0.217	0.923
Inf.Aug	0.972	0.116	-0.256	0.200	0.809
Inf.Sep	1.667	0.126	0.265	0.757	<0.001
Inf.Oct	3.210	0.145	0.882	1.450	<0.001
Inf.Nov	1.667	0.126	0.265	0.757	<0.001
Inf.Dec	1.869	0.128	0.374	0.877	<0.001
Inf.Afijio*					
Inf.Akinyele	1.546	0.221	0.004	0.868	0.048
Inf.Atiba	0.665	0.193	-0.787	-0.030	0.035
Inf.Atisbo	1.540	0.219	0.003	0.861	0.048
Inf.Egbeda	0.916	0.202	-0.483	0.309	0.666
Inf.Ibadan north	0.534	0.193	-1.007	-0.249	0.001

Inf.Ibadan North east	0.603	0.193	-0.884	-0.128	0.009
Inf.Ibadan North west	0.588	0.193	-0.909	-0.153	0.006
Inf.Ibadan South East	0.616	0.193	-0.863	-0.106	0.012
Inf.Ibadan South west	0.553	0.191	-0.967	-0.219	0.002
Inf.Ibarapa central	0.507	0.189	-1.049	-0.309	<0.001
Inf.Ibarapa east	0.390	0.187	-1.309	-0.576	<0.001
Inf.Ibarapa north	0.694	0.196	-0.749	0.019	0.062
Inf·Ido	0.419	0.189	-1.140	-0.498	<0.001
Inf.Irepo	0.546	0.192	-0.982	-0.230	0.002
Inf.Iseyin	0.465	0.191	-1.140	-0.392	<0.001
Inf.Itesiwaju	0.443	0.189	-1.185	-0.443	<0.001
Inf.Iwajowa	0.569	0.191	-0.938	-0.191	0.003
Inf.Kajola	0.884	0.201	-0.517	0.271	0.541
Inf.Lagelu	0.776	0.195	-0.637	0.129	0.193
Inf.Ogbomoso North	0.909	0.202	-0.493	0.301	0.636
Inf.Ogbomoso South	0.892	0.202	-0.510	0.281	0.569
Inf.OgoOluwa	0.855	0.199	-0.548	0.236	0.435
Inf.Olorunsogo	0.754	0.198	-0.671	0.105	0.153
Inf.Oluyole	0.953	0.202	-0.445	0.348	0.810
Inf.Onaara	0.660	0.193	-0.794	-0.038	0.031
Inf·Orelope	2.579	0.246	0.465	1.430	<0.001
Inf.Orire	2.750	0.248	0.525	1.499	<0.001
Inf·Oyo east	3.562	0.269	0.743	1.798	<0.001
Inf.Oyo west	3.890	0.285	0.799	1.918	<0.001
Inf.Saki east	2.925	0.259	0.565	1.581	<0.001
Inf.Saki west	2.426	0.244	0.408	1.365	<0.001
Inf.Surulere	4.307	0.290	0.892	2.028	<0.001
Inf.Agegrp 0-28	121.389	0.265	4.280	5.318	<0.001
Inf.Agegrp 1-11	9.478	0.122	2.010	2.488	<0.001
Inf.Agegrp 12-59	2.088	0.099	0.542	0.930	<0.001
Inf.Agegrp 5-9	1.070	0.095	-0.119	0.254	0.476

Inf.Agegrp 10-19	0.701	0.094	-0.539	-0.170	<0.001
Inf.Agegrp 20-40	0.599	0.093	-0.695	-0.329	<0.001
Inf.Agegrp >40*					

Inf.Agegrp 10-19	0.701	0.094	-0.539	-0.170	<0.001
Inf.Agegrp 20-40	0.599	0.093	-0.695	-0.329	<0.001
Inf.Agegrp >40*					

4.3 Test for Comparison of the regression models: ZIP, ZINB, ZIGP and ZAP (Hurdle) using the AIC and the -2logL.

Table 4.4 shows the comparison for typhoid fever in Oyo state.

Ho: Model 1, 2, 3, and 4 are the same

H1: Models 1, 2, 3, and 4 are different

Table 4.4: Test for the comparison of the models for Typhoid fever in Oyo state.

Model	AIC	-2LogL	DF
ZIP	51290.47	51,100	94
ZINB	30733.61	30540	95
ZIGP	51290.47	51,100	94
ZAP or Hurdle	51285.73	51,100	94

CHAPTER FIVE

DISCUSSION

5.1 PATTERN OF TYPHOID FEVER CASES IN OYO STATE POPULATION FROM 2011 TO 2014

The findings from this study indicate that the annual occurrence of typhoid fever cases was high between 2011 and 2012 but showed a sustained decrease from 2012 to 2014. This is in agreement with a study conducted in Nigeria which also reported a decline in prevalence of typhoid fever cases (Yusuff, et al 2014) but contradicts a study conducted in North-central Nigeria that reported yearly increase in prevalence in the state between 2011 and 2013 (Yahaya, et al 2014). According to the most recent estimates published in 2014, approximately 21 million cases and 222,000 typhoid related deaths occur annually worldwide. Historically, in the pre-antibiotic era, the case fatality rate of typhoid fever was 10- 20%, but now less than 1% with prompt treatment (CDC, 2013). A closer view on the distribution of the above suggests a change in disease occurrence pattern in recent years.

Typhoid fever cases seemed to increase towards the latter part of the wet season (September and October) to the start of the dry season (November and December) which is similar to that reported in a study in Tanzania by Malisa and Nyaki (2010), and also by Yahaya, et al. (2014). The high risk during these periods might be an index of higher water pollution. Common problems identified by residents are related to the hardship encountered sourcing for water (Alagiah, 1981). Although, most water sources will be recharged by rains, those without adequate protection especially well and stream will experience influx of run-off, which carries high load of impurities including pathogens (Fedkiw, 1991). Anthropogenic activities in the cities that release vaporous toxins into the atmosphere, just as unprotected well can be polluted by run-off (Enger & Smith, 2003). Global climate change is most apparent around issues related to water (UNICEF, 2014). The monthly report of the typhoid fever cases indicates a clear shift of the disease occurrence pattern in the study areas which correspond very well with the recent change in the timing of the rainy and dry season.

Incidences of the disease in some LGAs are far higher compared to the incidence reported in other LGAs. This might be as a result of the number of rural settlements found in these LGAs. Although, the order of typhoid fever prevalence is lower in some parts of the state, there is still need to strengthen typhoid control protocols that are currently been implemented in other parts of the states. This is in agreement with WHO (2016) report that, there is a higher risk of typhoid fever in areas with low standards of hygiene and water supply facilities. According to University of Warwick researchers, they determined that epidemics of typhoid are more likely to be caused by environmental changes, such as spread to a new geographical area rather than by genetic mutations (PNAS, 2014). High prevalence of typhoid fever in some of the study areas may indicate the prevalence of its infective agent more than the infective agents of other water-borne diseases. However, the low proportion of residents that suffered the disease may reflect the impact of intervention programmes abolition against the back drop of the previous epidemics in some of the LGAs (Lawoyin, et al., 1999).

Children are the most vulnerable to typhoid fever as shown from the result of this study. This is in agreement with the report in China and Pakistan in 2012, 15.3 cases per 100,000 person years among 5 to 6 years and 451.7 cases per 100,000 among 2 to 15 years were reported respectively. High predisposition of children to infections due to their low immunity may be linked to the higher incidence of water-borne diseases among children, especially 1-9 years (Hlupheka & Hailemariam, 2001). Also, children within this age group have little influence on the choice of water source unlike adults who determine where water is sourced for drinking and domestic uses.

Lastly, multi-national eradication and control programs against several water-borne diseases including typhoid fever in western Africa might have influence on this decrease in occurrence (Kerr, 2003). In 2012, WASH initiatives implemented through the United Nations Children's Fund (UNICEF) non-emergency programmes that helped more than 10 million people gain access to improved sanitation. Access to improved water sources increased from 49% in 2009 to 59% in 2011. The core activity of the WASH, supported by the Water supply and sanitation collaborative council (WSSCC) and UNICEF contributed towards achievement of Millennium development Goals (MDG) through combinations of actions to influence policy at national and global level. This helped in effecting behavioral change at the grassroots level. Also, in 2014, UNICEF's water, sanitation and hygiene (WASH) team works in over 100 countries worldwide

to improve water and sanitation services, as well as basic hygiene practices. UNICEF's efforts provided nearly 14 million people with clean water and over 11 million with basic toilets during this period. This might as well be of great influence on the decline of typhoid fever cases re-emerging in Oyo state.

The zero-inflated Poisson (ZIP) and zero-inflated generalized Poisson (ZIGP) models gave almost the same parameter estimates as well as the AIC and the -2logL statistic. The estimates from these models are very high in terms of the comparison statistic which suggests that the two regression models was not very good in modeling data with excess zeros. Also, the zero-altered Poisson (ZAP) gave a fair estimate of parameters and a much lower AIC and -2logL statistic compared to ZIP and ZIGP. This made it a slightly better model for the dataset. However, the zero-inflated negative binomial regression model (ZINB) gave the best parameter estimates and the lowest value for the AIC and -2logL. The ZINB model consistently fits the data best compared to other regression models considered in this study.

5.2 LIMITATION OF THE STUDY

Reliability of the cases of typhoid fever documented by the health workers in charge of the Oyo state integrated disease surveillance and response (IDSR) data might not be certain. In addition, the format in which the year 2010 data was collected does not correspond with the other years (2011-2014) thereby making it difficult to merge. Further, there was no record of socio-demographic variables such as sex and education status in the dataset. This made it difficult to get more insight about the factors that contributes to the occurrence of typhoid fever cases in the state.

5.3 CONCLUSION

The zero-inflated negative binomial distribution better accommodates over-dispersion in the outcome data compared to other zero-inflated distributions. Although all the models fits the data well, the ZINB fits best in predicting the effects of season, year of reporting and LGA on the occurrence of typhoid fever in Oyo state with the lowest AIC and the -2logL.

The month of reporting which is classified into season (wet and dry season) has effect on the re-occurrence of typhoid fever in the state. Some of the local government areas also had effects on the re-emergence of this disease.

5.4 RECOMMENDATION

The zero-inflated negative binomial model (ZINB) should be considered as the statistical method of choice for modeling count data with excess zeros in the presence of all the other zero-inflated models (ZIP, ZIGP, and ZAP) in this study.

The poor supply of potable water to some LGAs has been identified as a primary factor in the prevalence of typhoid fever among their residents, it can also be seen that a lot of improvement has to be made with respect to personal hygiene and environmental sanitation by these residents themselves. It is therefore recommended, that potable water provision and water sanitation projects should be embraced as an authentic intervention option to solving health problems arising from water contamination. Therefore, the prevention of water related diseases should take precedence over cure. To establish the extent of typhoid fever burden, detailed environmental epidemiology of water related diseases in Oyo state and other parts of Nigeria should be embarked upon.

REFERENCES

Agbakwuru EA, Adesunkanmi AR, Fadiora SO, Olayinka OS, Aderonmu AO, Ogundoyin OO, et al 2003. A review of typhoid perforation in a rural African hospital. *West African Journal of Medicine*; 22(1):22-25.

Alagiah, D.R. 1982 The International Drinking Water Supply and Sanitation Decade (IDWSSA). Harare, Zimbabwe.

Baddam, Ramani, Narender Kumar, Kwai-Lin Thong, et al 2012. "Genetic fine structure of a Salmonella enterica serovar Typhi strain associated with the 2005 outbreak of typhoid fever in Kelantan, Malaysia.". *Journal of Bacteriology* 194 (13): 3565–3566.

Bohning, D., Dietz, E., Schlattman, P., Mendonca, L., Kirchner, U. 1999. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A*. 162: 195-209.

Boucher, J.P., Denuit, M., Guillen, M. 2007. Risk classification for claim count: a comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*. 11(4): 110-131

Cameron, A.C., Trivedi, P.K. 1998. Regression Analysis of Count Data. New York: *Cambridge University Press*.

Centre for Disease Control and Prevention (CDC), 2013 www.cdc.gov/ncezid/dfwed/pdfs/typhi-annual-summary-2013-508c.pdf

Centers for Disease Control and Prevention (CDC), 2015. National Typhoid and Paratyphoid Fever Surveillance Annual Summary, 2013. Atlanta, Georgia: US Department of Health and Human Services, CDC,

Cheung YB (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. Statistics in Medicine 21: 1461-1469.

Chomel B, Bruno, Albino B, François-Xavier M. 2007. Wildlife, exotic pets and prevention of emerging parasitic zoonoses. *Emerging Infectious disease*; 13: 6-11.

Consul, P. C. and G. C. Jain (1970). On the generalization of Poisson distribution. *Annals of Mathematics and. Statistics*. 41, 1387.

Consul, P.C. 1989. Generalized Poisson distribution: Properties and Application. *New York: Marcel Dekker*

Crump JA, Luby SP, Mintz ED (2004). The global burden of typhoid fever. *Bull World Health Organ* 82: 346–353.

Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. *Biometrics* 65:1254–1261

Dalrymple, M., Hudson, I. and Barnett, A. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics & Data Analysis*, 41, 491–504.

Enger, E.D. & Smith, B.F. (2003) Environmental Science: A Study of Interrelationships (8th Ed.) *McGraw Hill Higher Education.*

Federal Ministry of Health, 2005. National treatment guidelines. A publication of the Federal Ministry of Health, Abuja, Nigeria. 47.

Federal Ministry of Health, 2009. National technical guidelines for integrated diseases surveillance and response. Abuja: Federal Ministry of Health; p. 2-69.

Fedkiw, J. (1991) Nitrate Occurrence in Waters. USDA, Washington D.C. Available at www.cee.vt.edu/ewr/environmental .

Greene WH (1994). Accounting for excess zeros and sample selection in Poisson and Negative binomial regression models, working paper, Department of Economics, Stern School of business, New York University, New York.

Greene, WH, 2002. *Econometric Analysis*. Prentice Hall, Inc.; Upper Saddle River, NJ: 5th Edition

Gurmu, S. and Trivedi, P. K. 1996. Excess zeros in count models for recreational trips. *Journal of Business and Economic Statistics* 14, 469-477.

Gurmu, S. 1997. Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of Applied Econometrics* 12, 225-242.

Gurmu, S. 1998. Generalized hurdle count data regression models. *Economics Letters*, 58, 263-268.

Heibron, D. 1994. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 36, 531-547.

Hinde, J. P. and Demetrio, C. G. B. 1998. Overdispersion: models and estimation. *Computational Statistics and Data Analysis* 27, 151-170.

Hlupheka, P.C. & Hailemariam, M. (2001) Prevalence of water-borne diseases within the health facilities in Nakuru District, Kenya. A Report Submitted to Applied Epidemiology, University of Nairobi, Kenya.

Hoberg EP, Polley L, Jenkins EJ, Kutz SJ, Veitch AM, Elkin BT. Integrated approaches and empirical models for investigation of parasitic diseases in northern wildlife. Emerg Infect Dis 2008; 14: 10-7.

Hornick RB. 1985. Selective primary health care: Stragegies for control of diseases in developing world. XX.Typhoid fever. *Rev Infect Dis*; 7: 536-46.

Hu MC, Pavlicova M, and Edward VN, 2011. Zero-inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial. *The American Journal of drug and alcohol abuse* VOL. 37, 37(5): 367–375.

Kerr, C., 2003. Success for river blindness control program. *Lancet Infect.Dis*. 3, 65.

Kosek MC, Bern and Guerrant RL, 2003. The global burden of diarrheal Disease, as estimated from studies published between 1992 and 2000. Bulletin of the World Health Organization, 81:197-204

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to random defects in manufacturing. *Technometrics*. 34: 1-14.

Lawoyin, T.O . Ogunbodede, N.A , Olumide, E.A . & Onadeko, M.O 1999. Outbreak of cholera in Ibadan, Nigeria. *European Journal of Epidemiology* 15, 367-370.

Lee AH, Wang K and Yau KKW 2001. Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal* 43: 963-975.

Lee AH, Wang K, Scott JA, Yau KKW, and McLachlan, GJ 2006. Multi-level zero inflated Poisson regression modeling of correlated count data with excess zeros. *Statistical Method in Medical Research* 15:47-61

Long, J. S., & Freese, J. 2006. Regression models for categorical dependent variables using stata (2nd, Ed.). College Station, TX: Stata Press.

Louis MW. Zoonotic parasitic diseases: emerging issues and problems. *Int J Parasitol* 2008; 38: 1209-10.

Lynch MF, Blanton EM, Bulens S, et al, 2009. Typhoid fever in the United States, 1999–2006. *JAMA*;302:859-65.

Malisa.A, and Nyaki. H, 2010. Prevalence and constraints of typhoid fever and its control in an endemic area of Singida region in Tanzania: Lessons for effective control of the disease. *Journal of Public Health and Epidemiology* Vol. 2(5), pp. 93-99

Marques, A.C., 1987. Human migration and the spread of malaria in Brazil. *Parasitol.* 166–170.

Moghimbeigi. A, Eshraghian ME., Mohammad. K, Mcardle. B 2008. Multilevel zero inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics* 35, 1193-1202.

Morens DM, Fauci AS 2012 Emerging infectious diseases in 2012: 20 years after the Institute of Medicine report. *MBio* 3: e00494–12. doi:10.1128/

Morens DM, Fauci AS (2013) Emerging Infectious Diseases: Threats to Human Health and Global Stability. *PLoS Pathog* 9(7): e1003467. doi:10.1371/ journal.ppat.1003467

Mouatassim Y, Ezzahid E, Belasri Y 2012. Operational Value-at-Risk in Case of Zero-inflated Frequency. Int J Econ Finance 4(6):70–77

Mullahy J 1986. Specification and testing of some modified count data models. *Journal of Econometrics* 33: 341-365.

Muyembe-Tamfum JJ, Veyi J, Kaswa M, Lunguya ●, Verhaegen J, Boelaert M 2009. An outbreak of peritonitis caused by multidrug-resistant *Salmonella* Typhi in Kinshasa, Democratic Republic of Congo. *Travel Med Infect Dis* 7 (40): 3. doi:10.1016/j.tmaid.2008.12.006.

Neelon, B.H., O'Malley, A.J., Normand, S.T. 2010. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*. 10(4): 421–439.

Ngongeh LA and Chiejina SN, 2014. Emerging Animal Parasitic Diseases: A Global Overview and Appropriate strategies for their Monitoring and Surveillance in Nigeria. *The Open Microbiology Journal*, 8, 87-94

Oguntoke, O. 2009, "Assessment of Blood and Urine lead levels of some pregnant women residing in Lagos, Nigeria". Environmental Monitoring Assessment 170, 67-474

Omole, D.O., 2013. Sustainable groundwater exploitation in Nigeria. *J. Water Resour. Ocean Sci.*, 2(2): 9-14.

Omole, D.O. and J.M. Ndambuki, 2015. Nigeria's Legal Instruments for Land and Water Use: Implications for National Development. In: Evans, O. (Ed.), In-Country Determinants and Implications of Foreign Land Acquisitions. Business Science Reference, 1. Hershey, PA, USA. pp: 354-373.

Omole D.O, Emenike C.P., Tenebe I.T, Akinde A.O. and Badejo A.A 2015. An Assessment of Water Related Diseases in a Nigerian Community. *Research Journal of Applied Sciences, Engineering and Technology* 10(7): 776-781, 2015

Omole, D.O., J.M. Ndambuki and K.O. Balogun, 2015. Consumption of sachet water in Nigeria: Quality, public health and economic perspectives. *Afr. J. Sci. Technol. Innov. Dev.*, 7(1): 45-51.

Onyango, D., M., Angienda, P., O. 2010. Epidemiology of Waterborne Diarrhoeal Diseases among Children Aged 6-36 Months Old in Busia - Western Kenya. *Int. J. Biol. Life Sci.*, 6(2): 92-99.

Pohlmeier, W., and V. Ulrich. 1995. An econometric model of the two-part decision making process in the demand for health care. *Journal of Human Resources* 30: 339–361.

Ridout, M.S., Hinde, J.P., Demetrio, C.G.B. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*. 57: 219-223.

Oguntoke, O. 2009, "Assessment of Blood and Urine lead levels of some pregnant women residing in Lagos, Nigeria". Environmental Monitoring Assessment 170, 67-474

Omole, D.O., 2013. Sustainable groundwater exploitation in Nigeria. *J. Water Resour. Ocean Sci.*, 2(2): 9-14.

Omole, D.O. and J.M. Ndambuki, 2015. Nigeria's Legal Instruments for Land and Water Use: Implications for National Development. In: Evans, O. (Ed.), In-Country Determinants and Implications of Foreign Land Acquisitions. Business Science Reference, 1. Hershey, PA, USA, pp: 354-373.

Omole D.O, Emenike C.P., Tenebe I.T, Akinde A.O. and Badejo A.A 2015. An Assessment of Water Related Diseases in a Nigerian Community. *Research Journal of Applied Sciences, Engineering and Technology* 10(7): 776-781, 2015

Omole, D.O., J.M. Ndambuki and K.O. Balogun, 2015. Consumption of sachet water in Nigeria: Quality, public health and economic perspectives. *Afr. J. Sci. Technol. Innov. Dev.*, 7(1): 45-51.

Onyango, D., M., Angienda, P., O. 2010. Epidemiology of Waterborne Diarrhoeal Diseases among Children Aged 6-36 Months Old in Busia - Western Kenya. *Int. J. Biol. Life Sci.*, 6(2): 92-99.

Pohlmeier, W., and V. Ulrich. 1995. An econometric model of the two-part decision making process in the demand for health care. *Journal of Human Resources* 30: 339–361.

Ridout, M.S., Hinde, J.P., Demetrio, C.G.B. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*. 57: 219-223.

Saffari, S. E., Robiah, A. and Greene, W. 2011. Handling of over-dispersion of count data via truncation using Poisson regression model. *Journal of Computer Science and Computational Mathematics*, 1(1), 1–4.

Sharma, V.P., 1996. Re-emergence of malaria in India. Indian J. Med. Res.103, 26–45.

Snow, N. D. (1894).The pathology and mode of communication of Cholera. *London Medical Gazette* 9, 745–753.

Swerdlow, D.S. (1992) Water-borne transmission of epidemic cholera in Trujilo, Peru: lessons for a continent at risk. *Lancet* 340, 28-33.

Tyagi, B.K., 2004. A review of the emergence of Plasmodium falciparum dominated malaria in irrigated areas of the Thar Desert, India. *Acta Tropica* 89, 227–239.

Typhoid Fever. World Health Organization

Typhoid Fever. *cdc.gov*. May 14, 2013..

Typhoid vaccines: WHO position paper 2008. (PDF). *Wkly Epidemiol Rec*. 83 (6): 49–59.

UNICEF/WHO, 2012. Progress on Drinking Water and Sanitation: 2012 Update. UNICEF and World Health Organization. ISBN: 978-92-806-4632-0.

UNICEF, 2012.United Nations Children's Fund, Raising Even More Clean Hands: Advancing Learning, Health and Participation through WASH in Schools, (New York), p. 2.

UNICEF, 2014 (United Nations Children's Fund).Sanitation updates. News, opinions and resources for sanitation for all http://sanitationupdates.wordpress.com

UNICEF, 2014.Water, Sanitation and Hygiene. https://www.unicef.org/wash/index_3951.html; www.un.org/sustainabledevelopment/water-and-sanitation.

USAID 2005 *USAID Health: Environmental Health, Overview*. Available at www.usaid.gov/ oue_work/goal_health/eh.

Vuong Q 1989. Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica.
; 57:307–334.

Winkelmann, R. 2003. Health care reform and the number of doctor visits – an econometric analysis. *Journal of Applied Econometrics*, 19, 455–472.

WHO 2014. Taeniasis/cysticercosis. Fact sheet No: 376. http://www.who.int/neglected_diseases/diseases/cysticercosis/en.

World Health Organization, 2007. Global plan to combat neglected tropical diseases, 2008-2015. Geneva, Switzerland: World Health Organization; Available from URL: http://whqlibdoc.who.int/hq/2007/WHO_CDS_NTD_2007.3_eng.pdf

World Health Organization (WHO), 2008. Prepared for World Water Day 2001. Reviewed by staff and experts from the cluster on Communicable Diseases (CDS) and the Water, Sanitation and Health unit (WSH).

World Health Organization (2008). The Global Burden of Disease: 2004 Update (WHO, Geneva).

World Health Organization, 2014; http://www.who.int/water_sanitation_health/diseases/burden/en/

USAID 2005 *USAID Health: Environmental Health, Overview*. Available at www.usaid.gov/ oue_work/goal_health/eh.

Vuong Q 1989. Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica.
; 57:307–334.

Winkelmann, R. 2003. Health care reform and the number of doctor visits – an econometric analysis. *Journal of Applied Econometrics*, 19, 455-472.

WHO 2014. Taeniasis/cysticercosis. Fact sheet No: 376. http://www.who.int/neglected_diseases/diseases/cysticercosis/en.

World Health Organization, 2007. Global plan to combat neglected tropical diseases, 2008-2015. Geneva, Switzerland: World Health Organization; Available from URL: http://whqlibdoc.who.int/hq/2007/WHO_CDS_NTD_2007.3_eng.pdf

World Health Organization (WHO), 2008. Prepared for World Water Day 2001. Reviewed by staff and experts from the cluster on Communicable Diseases (CDS) and the Water, Sanitation and Health unit (WSH).

World Health Organization (2008). The Global Burden of Disease: 2004 Update (WHO, Geneva).

World Health Organization, 2014; http://www.who.int/water_sanitation_health/diseases/burden/en/

Wright J, Gundry S, Conroy R. 2004. Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. *Trop Med Int Health* 9(1):106–117.

Yahaya UB, Olayemi, I.K, Spencer O, Yakubu M, 2014. Assessment of a Vulnerable Rural Community to Typhoid Fever using Geospatial-Temporal Analysis: Case Study of Ejule, Kogi State of Nigeria. *Journal of Environment and Earth Science* www.iiste.org ISSN 2224-3216 (Paper) ISSN 2225-0948 (Online) Vol.4, No.22.

Yap, Kien-Pong; et al. 2012. Insights from the genome sequence of a Salmonella enterica serovar Typhi strain associated with a sporadic case of typhoid fever in Malaysia. *J. Bacteriol.* 194 (18): 5124–5125. doi:10.1128/jb.01062-12.

Yau K, Wang K, and Lee A (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal* 45: 437-452.

Yusuff SA, John W, and Olohuntoba AC, 2014. Review on Prevalence of Waterborne Diseases in Nigeria. *J. of Advancement in Medical and Life Sciences*. V1I2.