College of Natural and Computational Sciences

Department of Statistics

Determinants of Diarrhea Among Under-Five Children in Kenya

By: Kedija Ahmed

A Thesis Submitted to the Department of Statistics in Partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics (Biostatistics)

Advisor: Prof. Eshetu Wencheko

Addis Ababa, Ethiopia

June 2024

Department of Statistics

College of Natural and Computational Sciences

Addis Ababa University

This is to certify that the thesis prepared by Kedija Ahmed entitled: "Determinants of diarrhea among under five children in Kenya" submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Biostatistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

______________	______________	______________
Examiner	Signature	Date
______________	______________	______________
Examiner	Signature	Date
______________	______________	______________
Advisor	Signature	Date

____

Chair of Department or Graduate Program Coordinator

Abstract

Determinants of diarrhea among under-five children in Kenya

Kedija Ahmed

Addis Ababa University, 2024

Diarrhea is defined as having loose or watery stools at least three times a day or more frequently. Diarrhea disease is the second most common cause of death for children under- five in the world and the first leading cause of death in Kenya. Every year, there are over 1.7 billion cases of diarrhea in children worldwide, and 525,000 children under the age of five died from preventable diarrheal illnesses. Descriptive statistics and multilevel binary logistic regression analysis were conducted to assess the prevalence of diarrhea and to identify factors that affect childhood diarrhea, respectively. The overall prevalence of diarrhea in Kenya was 14.80%. Multilevel binary logistic regression analysis showed that age of child, Birth order, region, residence, education level of mother, HH wealth index, number of U-5 children in the HH, mother age, media exposure and rotavirus vaccine have statistically significant associated with occurrence of diarrhea in Kenya. Government and concerned stakeholders should work for further reduction of diarrhea prevalence among under-five children in Kenya.

Acknowledgments

I am immensely grateful to Allah for granting me the strength and dedication to embark on and successfully accomplish this thesis. I extend my sincerest appreciation to my advisor, Eshetu Wencheko (Prof.), for his continual guidance and support throughout the entirety of this final thesis. Furthermore, I would like to express my gratitude to the other instructors and my classmates.

Contents

Abstract iii

Acknowledgments iv

List of Tables vii

ABBREVIATIONS AND ACRONYMS viii

CHAPTER ONE 1

INTRODUCTION 1

1.1 Background of the study 1

1.2 Statement of the Problem 3

1.3 Objectives of the Study 3

1.3.1 The general objective 3

1.3.2 Specific objectives 4

1.4 Significance of the study 4

1.5 Limitation of the study 4

LITERATURE REVIEW 5

2.1 General Concept about diarrhea disease 5

2.2 Empirical Review 5

2.3. Related Literature on determinants of diarrheal diseases among U-5 children 6

CHAPTER THREE 10

DATA AND METHODOLOGY 10

3.1 Description of Study Area and Data Source 10

3.2. Variables in the Study 11

3.2.1. The response variable 11

3.2.2. Explanatory variables 11

3.3 Methodology 12

3.3.1 Descriptive Statistics 12

3.3.2 Inferential Statistics 12
3.4 Logistic Regression Model 13
3.5 Binary Logistic Regression Model 13
3.6 Mixed Effect Binary Logistic Regression Model 14
3.6.1Empty model 14
3.6.2. Random intercept model 15
3.6.3 Random Coefficient Model 15
3.7 Tests of Model Goodness of Fit 17
3.8 Methods of Variable Selection 18
3.9 Model Selection 19
RESULTS 20
4.1 Results of Descriptive Statistics 20
4.2 Results of Inferential Statistics 23
4.2.1 Multilevel analysis results 24
4.2.2 Model adequacy checking results 29
4.2.3 Model goodness of fit test result 30
5.1 Discussion 31
CHAPTER SIX 33
6.1 Conclusions 33
6.2 Recommendation 35
References 36

List of Tables

Table 1: List of independent variables with their category.. 11
Table 2: Weighted frequency and percentage of diarrhea 20
Table 3: Weighted frequency and percentage of diarrhea versus individual level predictors 20
Table 4: Weighted Frequency and percentage of diarrhea versus community level predictors ... 22
Table 5: Results of intercept-only model... 24
Table 6 : Results of intraclass correlation.. 24
Table 7: Result of Random Intercept Model... 24
Table 8 Result of Random Coefficient Model.. 25
Table 9 : Results of model selection criteria.. 29
Table 10: Result of chi-square test of association 39

ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immunodefficency Syndrome
AOR	Adjusted Odds Ratios
AIC	Akaike's Information Criterion
CDC	Center for Disease Control and Prevention
CI	Confidence Interval
DHS	Demographic and Health Survey
EA	Enumeration Area
HIV	Human Immunodeficiency Virus
ICC	Intraclass Correlation Coefficient
KDHS	Kenya Demographic and Health Survey
KHMSF	Kenya Household Master Sample Frame
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimation
MC	Multicollinearity
U-5	Under-Five
UNICEF	United Nation Childrens Fund
VIF	Variance Inflation Factor
WBB	World Bank Blog
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

According to the World Health Organization (WHO, 2017), having loose or watery stools at least three times a day or more frequently than is typical for an individual is defined as diarrhea. It typically indicates a digestive tract infection, which can be caused by a number of different bacterial, viral, and parasitic pathogens. Diarrhea disease is the second most common cause of death for children under five. Every year, there are over 1.7 billion cases of diarrhea in children worldwide, and 525,000 children under the age of five die from preventable diarrheal illnesses.

Acute watery diarrhea, bloody diarrhea, and persistent diarrhea are the three main types of childhood diarrhea. When people have acute watery diarrhea, such as cholera, they often lose a lot of fluid and become dehydrated quickly. *Rotavirus* and *escherichia coli* bacteria, as well as *vibrio cholera* are the pathogens that cause acute watery diarrhea. Visible blood in the feces is indicative of the second type of diarrhea, sometimes known as dysentery or bloody diarrhea. *Shigella* is the source of bloody diarrhea, and it is linked to intestinal damage and nutritional losses in an infected person. An episode of diarrhea that lasts at least 14 days, with or without blood, is considered persistent diarrhea (UNICEF, 2017).

Worldwide, 1 in 3 people lack access to clean drinking water. According to a recent report by UNICEF and WHO, billions of people worldwide still suffer from inadequate access to water, sanitation and hygiene. Globally, 4.2 billion people lack access to securely managed sanitation services, 2.2 billion to safely managed drinking water services, and 3 billion to even the most basic hand washing facilities (WHO, 2019).

Diarrhea substantially raises the chance of later stunting of physical growth and cognitive development. This is supported by a study that examined the non-fatal burden of childhood diarrhea, and discovered that the condition markedly raised the likelihood of later impairments to the area's physical and cognitive development (Khalil et al., 2016).

The decline in the incidence rates of diarrhea in West and Central Africa has slowed down over time (1995–2009), with minimal progress made between 2010 and 2017 (Simen-Kapeu et al., 2021).

Globally, the proportional distribution of diarrhea cases among children under five years of age, Africa (696 million) and South Asia (783 million) account for over half the cases of childhood diarrhea; East Asia and Pacific (35 million) and rest of the world (480 million) account for diarrhea cases (UNICEF, 2017).

According to the Center for Disease Control and Prevention (CDC), children with human immunodeficiency virus (HIV) have a significantly higher mortality rate when they experience diarrhea. In fact, the death rate for these children is 11 times greater compared to children without HIV. However, advancements in healthcare have demonstrated that preventing diarrhea through measures such as rotavirus vaccination, breastfeeding, and promoting safe water, improved hygiene, and sanitation is not only feasible but also economically beneficial.

A study conducted in 34 sub-Saharan African countries, including Kenya, revealed that the prevalence of childhood diarrhea morbidity in sub-Saharan Africa was high. Maternal age, wealth index, maternal education, maternal occupation, age of child, time of initiation of breast feeding and time to get a water source were significantly associated with diarrhea (Demissie et al., 2021).

Based on the Kenya DHS 2022 report, within a span of two weeks prior to the survey, approximately 14% of children below the age of 5 experienced an episode of diarrhea (DHSProgram, 2022).

The findings of a research carried out in Kenya revealed a notable correlation between child age, caregiver's low level of education and improper disposal of children's feces with the occurrence of diarrhea episodes (Mulatya et al., 2020).

The peak of diarrhea cases occurred one to two months following intense rainfall (Tornheim et al., 2010).

1.2 Statement of the Problem

Globally in 2021, diarrhea remains a significant cause of mortality among children under the age of 5, responsible for around 9% of all child deaths. Shockingly, this equates to over 1,200 young lives lost every single day, totaling approximately 444,000 children per year. It is disheartening to realize that despite the existence of a straightforward treatment option, this preventable tragedy continues to persist (UNICEF, 2024).

In Kenya, diarrheal illnesses rank fourth among all outpatient visits in terms of morbidity and mortality, and they are the primary cause of death for children under five (Wanjala, 2022)

Diarrheal disease continues to be a significant public health concern among under-five children in Kenya, contributing to morbidity and mortality rates. Despite existing interventions, the determinants of diarrhea disease in this population remain poorly understood. Studies carried out in Kenya previously exhibit differences in their design, sample size, and the tools used for data collection. No research has been conducted on the determinants of diarrheal disease using the latest data from KDHS 2022 and also most of the previous studies have examined a limited number of determinants, without considering other potential factors that could contribute to the the childhood diarrhea.

This study aimed to identify and examine the key determinants of diarrheal among under-five children in Kenya using the most recent data from the 2022 Kenya Demographic and Health Survey (KDHS) by incorporating potential factors that may contribute to the occurrence of diarrhea disease. By exploring these determinants, this research seeks to provide valuable insights into the factors influencing diarrhea disease and guide targeted interventions to reduce its prevalence and impact on child health in Kenya.

1.3 Objectives of the Study

1.3.1 The general objective

The main objective of this study was to identify determinants of diarrhea among under-five children in Kenya.

1.3.2 Specific objectives

- To assess the prevalence of childhood diarrhea in different region of Kenya.
- To assess the variation in childhood diarrhea within and between enumeration areas in Kenya.

1.4 Significance of the study

The study on the determinants of diarrhea disease among under-five children in Kenya holds immense significance. By examining the factors that influence the occurrence of this disease, the study has the potential to enhance public health outcomes. It offers valuable insights into the specific determinants that lead to diarrhea disease in this vulnerable population, which can guide the development of targeted interventions and policies. Moreover, the research addresses gaps in existing literature and empowers communities with the information needed to take preventive actions against diarrhea disease.

1.5 Limitation of the study

The stud design utilized in this study was cross-sectional, resulting in cross-sectional bias and capturing only prevalence cases, not incident cases of diarrhea.

The data utilized in this study was gathered in a cross-sectional manner; in these types of studies, it can be challenging to consider the seasonal variations in the prevalence of diarrhea.

CHAPTER TWO

LITERATURE REVIEW

2.1 General Concept about diarrhea disease

Acute diarrhea and respiratory infections are the most common pediatric illnesses and reasons for visiting health care in low-income and middle-income nations. The regions with the highest rates of severe diarrhea episodes were Southeast Asian (26%) and African (26%). Sub-Saharan Africa had the highest rate of childhood fatalities in 2011, with 50% of deaths attributed to diarrhea (Walker et al., 2013).

A variety of organisms, like bacteria, viruses, and protozoa, can cause diarrhea. Diarrheal illnesses remain a major concern in low- and middle-income nations. Diarrhea ranks among the top causes of death for children under five. There is a higher chance of diarrheal sickness in those who are impoverished. The fact that diarrheal illnesses cause deaths in children nowadays is evidence of unequal resource distribution worldwide (Ahs et al., 2010).

Dehydration happens when electrolytes and water are lost and not sufficiently restored. Dehydration can cause symptoms such as sunken eyes, sunken fontanels in babies, loss of skin turgor, restlessness, or irritability, and thirst. A patient who is severely dehydrated may exhibit low blood pressure, a fast pulse that may be difficult to detect, a sluggish and impaired level of consciousness, a decreased or absent urine output, and chilly, damp extremities that may appear cyanotic (WHO, 2005).

2.2 Empirical Review

According to the proportionate distribution of diarrheal disease-related mortality worldwide, 84% of child deaths from diarrheal diseases occur in South Asia (38%), and Africa (46%). Additionally, East Asia and the Pacific (9%), as well as the rest of the globe (7%) (UNICEF, 2017). Diarrhea kills 2,195 (nearly 32 school buses full of children each day) more than AIDS, malaria, and measles combined (CDC).

The World Health Organization's 2004 Global Burden of Disease estimates show that 15 nations accounted for around 75% of the annual child fatalities caused by diarrhea. Included were the following countries: China (40,000), Uganda (29,300), Kenya (27,400), Niger (26,400), Burkina

Faso (24,300), Tanzania (23,900), Mali (20,900), Angola (19,700), Afghanistan (82,100), Ethiopia (73,700), Pakistan (53,300), Bangladesh (50,800), and India (386,600) (WHO, 2004).

Roughly 88% of deaths linked to diarrhea are caused by contaminated water, poor sanitation, and inadequate hygiene. Diarrhea can negatively affect a child's growth and mental development (CDC).

Around 2 billion people in the world live without access to safely managed drinking water service. Among them, 771 million people cannot access even basic drinking water services (water from improved source with a round-trip collection time of less than 30 minutes including queuing). People with low access to drinking water services are concentrated in sub Saharan Africa (WBB).

According to the study conducted to assess diarrhea incidence in low- and middle-income countries in 1990 and 2010, the incidence of diarrhea decreased between 1990 and 2010 from 3.4 episodes per child year to 2.9 episodes per child year. As in the past, babies aged 6 to 11 months have the highest incidence rates (4.5 episodes per child year in 2010). In 1990, there were about 1.9 billion cases of pediatric diarrhea across these 139 nations, and by 2010, there were almost 1.7 billion cases (Fischer Walker., 2012).

2.3. Related Literature on determinants of diarrheal diseases among U-5 children

Berde et al., (2018) used multivariate logistic regression models to investigate the causes of childhood diarrhea in children under five years old in Nigeria. The multivariate analysis findings showed that child's age, region and disposal of the child’s last fecal matter are important determinants.

In Kenya, a research was carried out utilizing KDHS 2008-09 to ascertain the contributing factors of diarrhea disease in children under the age of five. The study employed multivariable analysis and identified several variables including child's age, mother's place of residence, mother's level of education, source of drinking water, and size of the household. The research paper highlighted the significance of maternal literacy and access to high-quality drinking water sources in reducing the prevalence of childhood diarrhea (Mbugua et al., 2014).

Mosisa et al., (2021) conducted a community-based unmatched case-control study in Jima Geneti District, Oromia region, Ethiopia, to identify covariates that contribute to diarrheal illnesses in children under five. The results of multivariable logistic regression analysis showed that child's age, the availability of a hand-washing facility, proximity of a latrine, mothers' and caregivers' history of diarrheal illness within the previous two weeks, use of latrines, practice of hand-washing during a critical period, refusal to use domestic solid waste, and status of rotavirus vaccination were significantly associated with diarrhea.

In Kenya, a study was carried out utilizing the KDHS 2014 data, which was a survey conducted on a population base to examine the disease burden and risk factors of diarrhea in children below the age of five. The researchers employed both bivariate and multivariable analysis techniques to determine the factors contributing to diarrhea in this age group. The findings of the study revealed a significant association between diarrhea and factors such as child's age, caregiver's low level of education, and unsafe disposal of children's feces (Mulatya and Ochieng, 2020).

In order to determine the factors that contribute to childhood diarrhea in Tiko town, Cameroon in Sub-Saharan Africa, a cross-sectional community household survey was carried out from August 1 to August 31, 2012 by Tambe et al. (2015). The study demonstrated a substantial correlation between diarrhea and the number of children under the age of ten living in a family, the child's restroom, container for storing drinking water, and the caregiver's awareness about safe water source.

Tesfaye et ale., (2020) studied if moderate to severe diarrhea are associated factors among under-five children in Wonago district, South Ethiopia. Multivariable logistic regression model was employed in the study to account for confounding variables after the bivariate logistic regression technique was used to determine the weak relationship between the independent variables and the dependent variable. The findings demonstrated a substantial correlation between moderate to severe diarrhea and number of family members, presence of animals in homes, availability of latrines, and practice of hand washing. The percentage of cases of moderate to severe diarrhea was 30.9%.

In Medebay Zana District, northwest Tigray, Ethiopia, Asfaha et al., (2018) carried out a community-based unmatched case–control study to determine the factors of childhood diarrhea.

The study demonstrated that maternal educational status, age of child, number of children under five, child feeding habits, maternal history of diarrhea, bathroom facilities, solid waste disposal and household drinking water were associated with childhood diarrhea.

A research conducted in Kasarani, Nairobi, examined the impact of water, sanitation, and hygiene on the prevalence of diarrhea among children under five years old. The study, which utilized a cross-sectional design, identified the age of the child, water treatment methods, and solid waste storage practices as the key risk factors for childhood diarrhea. Additionally, the findings indicated that water quality in Kasarani posed a significant risk for childhood diarrhea. (Kimani et al., 2019).

Shine et al., (2020) conducted a community-based cross-sectional survey in Debre Berhan town, Ethiopia, from April 13–28, 2018, to assess the prevalence of diarrhea and its associated factors among children under the age of five. The study employed a multivariable logistic regression analysis to analyze the data. The results revealed that 16.4% of U-5 children in the selected area experienced diarrhea. Interestingly, children aged between seven and eleven months had a higher likelihood of developing diarrhea compared to those younger than seven months. Additionally, the study found that second-born children had a higher incidence of diarrhea than first-born children. Surprisingly, children who received rotavirus immunization had a 10.3 fold increased risk of diarrhea. These findings highlight the significant role of rotavirus in contributing to the burden of diarrheal cases among children in the town.

A pooled regression analysis was conducted in Ghana utilizing data from the Ghana Demographic and Health Survey, focusing on 15,808 children under the age of five. The aim was to evaluate the collective impact of environmental factors on the prevalence and severity of childhood diarrhea over a span of 21 years. The prevalence of childhood diarrhea exhibited a gradual decrease from 20% in 1993 to 16% in 2003, followed by an increase to 20% in 2008. Subsequently, there was a notable reduction to 12% in 2014. The findings indicated that various factors such as the child's current age, geographical location, religion, mother's educational attainment, ethnicity, access to clean drinking water and sanitation facilities, living conditions, birth order, maternal age, and the child's gender were significantly linked to diarrhea (Afitiri et al., 2020).

Owusu et al., (2023) carried out a research to examine the occurrence and factors influencing diarrhea and acute respiratory infections in children under the age of five in West Africa through a cross-sectional study utilizing DHS data from 13 West African nations. The study utilized multivariable logistic regression analysis to pinpoint potential predictors of diarrhea and ARIs. The findings indicated an overall diarrhea prevalence of 13.7%. Moreover, factors such as place of residence, child's age, mother's age, mother's educational level, wealth index, sanitation, water source, and nutritional status were found to have a significant correlation with diarrhea prevalence.

In Mathare informal settlement, Nairobi, Kenya, a cross-sectional study was conducted to investigate the factors influencing childhood diarrhea. The findings revealed that gender of the caregiver, relationship between primary caregivers, as well as the number of individuals in the household and children in the household exhibited statistical significance in relation to the prevalence of diarrhea (Guillaum et al., 2020).

A research carried out in Kawangware slum, Nairobi in Kenya to examine the factors that contribute to the occurrence of childhood diarrhea among the under-five population indicated hand washing, bottle feeding, latrine usage, disposal of infant feces, and water storage methods were identified as noteworthy factors linked to diarrhea in children below the age of five (Mutama et al., 2019).

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Description of Study Area and Data Source

This study is conducted in Kenya, East Africa. Kenya shares borders with South Sudan to the NorthWest, Ethiopia to the North, Somalia to the East, Uganda to the West, Tanzania to the South, and the Indian Ocean to the Southeast. The geography, climate, and population of Kenya exhibit significant diversity. The country encompasses frigid snow-capped mountaintops, such as batian, nelion, and point lenana on mount Kenya, surrounded by expansive forests, wildlife, and fertile agricultural regions (Wikipedia).

The seventh Demographic and Health Survey (DHS) conducted in Kenya, known as KDHS 2022, provided the population-based cross-sectional survey data for this study.

The 2022 KDHS utilized a stratified sampling method that involved two stages of selection from the Kenya household master sample frame (K-HMSF). In the first stage, 1,692 clusters were chosen from the K-HMSF using equal probability with independent selection within each sampling stratum. A household listing was conducted in all selected clusters, and the resulting list of households served as the sampling frame for the second stage of selection, where 25 households were chosen from each cluster. However, it was later discovered that some clusters had fewer than 25 households during the listing process. As a result, all households from these clusters were included in the sample. This brought the total number of sampled households for the 2022 KDHS to 42,022 (DHSProgram, 2022).

The implementation of the 2022 KDHS was successful in 1,691 clusters, with the exception of one cluster in Mandera that couldn't be visited due to security concerns. Due to the non-proportional allocation to the sampling strata and nonresponse, the survey was not self-weighting. Therefore, the resulting data has been weighted to ensure representation of the various survey domains (DHSProgram, 2022). For this study we used STATA 16 software to perform descriptive statistics and multilevel analysis.

In this study, the inclusion criteria encompassed children who were under the age of five and had a de jure residence. Conversely, the exclusion criteria involved children who were age five and above, as well as those with de facto residences were not included in the study.

We use weight in this study to adjust for differences in the probability of selection and interview between cases in a sample. These weights are necessary due to the unequal probability with which samples are selected in DHS surveys. By applying weights, we ensure that the tabulated results provide a proper representation of the target population. The weighting process helps account for design effects and nonresponse rates, ultimately improving the accuracy and reliability of the estimates derived from the survey data.

3.2. Variables in the Study

3.2.1. The response variable

The response variable is dichotomous or binary having two status of diarrhea in the two weeks prior to the survey. We code this dichotomous variable as “1” denoting having diarrhea and "0" otherwise.

3.2.2. Explanatory variables

The explanatory variables were grouped as individual and community-level. The independent variables for this study were listed in line with previous studies conducted on the factors affecting childhood diarrhea.

Table 1: List of independent variables with their category

Variables	Categories
Individual-level factors	
Age of child (in months)	<6, 6-11,12-23, 24-35, 36-47, 48-59
Sex of child	Male, Female
Birth order	1st, 2nd, 3rd and above
Breastfeeding	Yes, No
Age of Mother (in years)	15–24, 25-34,35-49
Wealth index of the HH	Poorest, Poorer, Meddle, Richer, Richest
Educational level of mother	No education, Primary, Secondary, Higher

Mother work	Yes, No
Media exposure	Yes, No
Number of U-5 children	<=2, >=3
Family size	≤5, > 5
Rotavirus vaccinated	Yes, No
Community-level factors	
Place of residence	Urban, Rural
Region	Cost, North, Eastern, Central, Nairobi, Nyanza, Rift Valley, Western
Type of toilet facility	Improved, Unimproved
Source of drinking water	Improved, Unimproved

3.3 Methodology

3.3.1 Descriptive Statistics

In this study we presented our finding with descriptive statistics (frequencies, percentages and cross tabulations).

3.3.2 Inferential Statistics

Chi-square test of independence

The objective of the chi-square test of statistical independence is to test whether there is an association between two categorical variables. The test assumes that the observation must be independent of each other, the sample must be randomly selected from the population, the expected frequency of each cell must be at least 5, the data are mostly categorical (nominal or ordinal) and must be in the form of frequencies, the observed frequencies in any one cell of the table must not be zero (empty cell) and the frequencies data must have a precise numerical value and must be organized into categories or groups.

The null hypothesis assumes that there is no association between the categorical variables being examined; on the other hand, the alternative hypothesis suggests that there is a significant association between the variables.

The test statistic for chi-square test of independency is given by:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}\ .$$

In the above E_{ij} and O_{ij} respectively, represent the expected and observed number of counts in the $(ij)^{th}$ cell, Where *c* and *r* represent number of columns and rows respectively. We reject the null hypothesis for p-value less than specified alpha level of significance.

3.4 Logistic Regression Model

Logistic regression models are often called logit models. Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables where the dependent variable is dichotomous or polychromes (Agresti., 2012).

3.5 Binary Logistic Regression Model

The binary logistic regression is a type of regression which is used when the response variable is dichotomous and the predictor variables are of any type. Binary logistic regression model is used to investigate the effect of predictors on the probability of the response variable (Agresti., 2012).

In this study, the dependent variable is given as: $Y_i = \begin{cases} 1, Having\ diarrhea \\ 0,\ Otherwise \end{cases}$

In binary logistic regression, the outcome variable Y_i (i=1, 2,…,n) follows a Bernoulli probability distribution that takes on the value 1 with probability π_i and 0 with probability $1- \pi_i$.

$$P(Y_i = 1) = \pi_i, \quad P(Y_i = 0) = 1 - \pi_i$$

Where π denote the Probability of success (having diarrhea in the two weeks prior to the survey).

Hypothesis Test for Regression Coefficient

The Wald Test. The Wald test is used to test the statistical significance of each coefficient, β_j, in the model. H_0: $\beta_1 = \cdots = \beta_k = 0$ versus H_1: $\beta j \neq 0$ at least for one *j*, *j* =1, …, *k*.

The Wald test statistic is given by: $Z = \frac{\hat{\beta}}{SE(\hat{\beta})}$, where SE is the unrestricted standard error of $\hat{\beta}$.

Under H_0 is true, Z has a standard normal distribution; it then follows that Z^2 has an approximate chi-squared distribution with df = 1 (Agresti., 2012).

3.6 Mixed Effect Binary Logistic Regression Model

Mixed-effect binary logistic regression contains both fixed effects and random effects. It is used to model binary outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables when data are clustered or there are both fixed and random effects.

Due to the hierarchical nature of DHS data and children age 0–59 months were nested within a household and that household were nested with in cluster/EAs, the assumption of independence of observations and equal variance across clusters is violated. Thus mixed effect models which included both fixed and random effects were used to assess the clustering effect of childhood diarrhea.

The fixed effects are used to estimate the association between the likelihood of childhood diarrhea and explanatory variables at both individual and community levels. In the final multivariable analysis, the associations between the response and independent variables were presented using adjusted odds ratio (AOR) and 95% confidence interval (CI) with their respective p-value.

3.6.1Empty model

The empty model is a model without any explanatory variables:

$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij}$

Where, γ_{00} $is\ the\ intercept$

U_{0j} denotes the random effect for the jth enumeration area (EA).

R_{ij} denotes residual

Variance decomposition:

$$var(Y_{ij}) = var(U_{0j}) + var(R_{ij}) = {\tau_0}^2 + \sigma$$

Covariance between two individuals ($i \neq i'$) in the same group j:

Cov $(Y_{ij}, Y_{i'j}) = var(U_{0j}) = {\tau_0}^2$ and their correlation is given by:

$$\rho(Y_{ij}, Y_{i \cdot j}) = ICC = \frac{{\tau_0}^2}{({\tau_0}^2 + \sigma^2)}$$

This is the intraclass correlation coefficient (Snijders and Bosker, 2011).

3.6.2. Random intercept model

The random intercept model is a type of multilevel logistic regression model that allows for the intercept to vary randomly across different groups or clusters in the data. This model is particularly useful when there is clustering or nesting of data points within higher-level units. The random intercept accounts for the variation in the baseline level of the outcome variable between these groups (Snijders and Bosker, 2011).

The formula for the random intercept model in multilevel logistic regression can be expressed as follows:

$$logit\ (p_{ij}) = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_1 X_{ij}$$

Where:

- p_{ij} represents the probability of the binary outcome for individual i in group j.
- β_{0j} is the group-specific intercept that varies randomly across groups.
- β_1 is the fixed effect coefficient associated with the predictor variable Xij.

3.6.3 Random Coefficient Model

On the other hand, the random coefficient model extends the random intercept model by allowing not only the intercept but also one or more coefficients to vary randomly across groups. This means that not only do the baseline levels differ between groups, but also how predictors influence the outcome can vary across these groups (Snijders and Bosker, 2011).

The formula for a simple random coefficient model with one predictor variable can be represented as:

$$logit(p_{ij}) = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + (\beta_{1j}+u_{1j})X_{ij}$$

Where:

- p_{ij} represents the probability of success for individual i in group j.

- β_{0j} is still the group-specific intercept.
- (β_{1j}+u_{1j}) denotes a group-specific slope where both fixed effect coefficient and random effect are included.
- X_{ij} is the predictor variable.
- u_{1j} represents the random effect on the slope parameter.

In summary, while the random intercept model accounts for varying baseline levels across groups, the random coefficient model further allows for differences in how predictors affect outcomes between these groups.

Assumptions about mixed effects logistic regression model: (i) Linearity: the relationship between the natural log of these probabilities (when expressed as odds) and the predictor variable is linear; (ii) There are no outlier; (iii) No multicollinearity.

Parameter Estimation Method for Mixed Effect Binary Logistic Regression Model

In a mixed effect binary logistic regression model, the parameters are estimated using a method called Maximum Likelihood Estimation (MLE). This method aims to find the parameter values that maximize the likelihood of observing the data given the model. MLE is a widely used statistical method for estimating the parameters of a statistical model. In the context of mixed effect binary logistic regression, MLE works by finding the values of the fixed effects and random effects that make the observed data most probable. The likelihood function is maximized by adjusting the parameter estimates iteratively until convergence is reached (Twisk, 2013).

Assumption Checking for Multilevel Binary Logistic Regression Model

Test of multicollinearity: The assumption can be verified with the variance inflation factor (VIF), which determines the correlation strength between the independent variables in a regression model. The VIF is defined as VIF ($\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$) $= \frac{\mathbf{1}}{\mathbf{1}-\boldsymbol{R_i}^{\mathbf{2}}}$. Here $\boldsymbol{R_i}^{\mathbf{2}}$ is the coefficient of determination obtained when the X_i variable is regressed on the remaining explanatory variables. A rule of thumb: If VIF ($\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$) exceed 10, then $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ is poorly estimated because of multicollinearity (MC) (or the jth regressor variable, meaning that X_j is responsible for MC.

Test for outliers: This assumption can be verified by calculating Cook's distance (Di) for each observation to identify influential data points that may negatively affect the regression model. In situations when outliers exist, one can implement the following solutions: remove the outliers, consider a value of the mean or median instead of outliers or Keep the outliers in the model but maintain a record of them while reporting the regression results.

3.7 Tests of Model Goodness of Fit

Deviance Goodness-of-Fit Test. The deviance goodness-of-fit test assesses the discrepancy between the current (simpler model), M, and the full model (the saturated model), S. Let L_M denote the maximized log-likelihood value for a model M of interest. Let L_S denote the maximized log-likelihood value for the most complex model, S, possible. This model has a separate parameter for each observation and it provides a perfect fit to the data. The model is said to be saturated. Because the saturated model has additional parameters, its maximized log-likelihood L_S is at least as large as the maximized log-likelihood L_M for a simpler model M. The deviance of a GLM is defined as:

$$\text{Deviance} = 2(L_S - L_M).$$

A large value of deviance and small *p*-values provide strong evidence of model lack of fit.

The Hosmer-Lemeshow Test: The Hosmer-Lemeshow goodness-of-fit test compares the observed and expected frequencies of events and non-events to assess how well the model fits the data. The aim here is to test:

$$H_0: \beta_1 = \cdots = \beta_k = 0 \text{ versus } H_1: \beta_j \neq 0 \text{ at least for one } j,\ j = 1, \ldots, k.$$

We use the test statistic due to Hosmer and Lemeshow (2000):

$$\text{HL} = \sum_{g=1}^{G} \frac{(O_{1g} - E_{1g})^2}{N_g \pi_g (1 - \pi_g)}$$

Under the null hypothesis, the statistic HL is distributed as chi-square distribution with (g −2) degrees of freedom; g (g = 1, …, n); g is the number of rows (groups) in a $g \times 2$ table into which the n observations are classified (see Hosmer and Lemeshow, 2000). H_0 is rejected if the computed value of HM ≥ chi-squared distribution (g −2) at a preset level of significance α.

Pearson's chi-square Test: The overall goodness of fit of a model with k covariates can be tested by performing the hypothesis:

H_0: $\beta_1 = \cdots = \beta_k = 0$ versus H_1: $\beta j \neq 0$ at least for one j, $j = 1, \ldots, k$.

The goodness-of-fit test statistic is Pearson's chi-square

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

Where O_i and E_i, respectively, represent the observed class and expected class frequencies.

The Pearson chi-square is compared with the chi-square distribution having $(r - 1)$ $(c - 1)$ degrees of freedom, where there are r rows and c columns. Rejection of the hypothesis states that all the independent variables are not significant for large value of Pearson chi-square.

Likelihood Ratio Test (LRT): The likelihood ratio test compares the fit of a full model with a reduced model by examining the difference in log-likelihood values. In multilevel logistic regression, this test can be used to compare nested models with different levels of complexity.

3.8 Methods of Variable Selection

Forward Selection

In this methodology it is assumed that there were no explanatory variables in the model except an intercept term. Forward selection adds terms sequentially. At each stage it selects the term giving the greatest improvement in the fit. The minimum p-value for testing the term in the model is a sensible criterion, since reductions in deviance for different terms may have different degrees of freedom value (Hosmer Jr et al., 2013).

Backward Elimination

Backward elimination begins with a complex model and sequentially removes terms. At each stage, it select the term whose removal has the least damaging effect on the model (remove variables that have largest p-value). The process stops when any further deletion leads to a significantly poorer fit (Agresti, 2012).

Stepwise variable selection

This methodology is based on choosing the explanatory variables in the subset model in steps which could be either adding one variablie at a time or deleting one variable at a time. One of the drawbacks of using stepwise methods for variable selection is that they can be unreliable or inconsistent, depending on the data and the criteria used. Stepwise methods are sensitive to the sample size, the order of variables, the correlation among variables, and the significance level (Hosmer Jr et al., 2013).

3.9 Model Selection

Akaike's Information Criterion (AIC)

The Akaike's information criterion judges a model by how close its fitted values tend to be the true mean values, in terms of a certain expected values.

AIC = -2(log -likelihood) + 2K

Where, K is number of model parameter (the number of variables in the model plus the intercept).

Out of a set of candidate models, the model with the lowest AIC is preferable (Akaike, 1974).

CHAPTER FOUR

RESULTS

4.1 Results of Descriptive Statistics

We used descriptive statistics to furnish fundamental details regarding variables in a dataset and to emphasize potential correlations among variables by using summary tables and bar charts. The response variable in this study was childhood diarrhea, categorized as yes or no.

Table 2: Weighted frequency and percentage of diarrhea

Diarrhea	Frequency	Percent
No	13,380	85.20
Yes	2323	14.80
Total	15703	100.00

This study included 15703 children under the age of five. The findings revealed that 13,380 children (85.2%) did not have diarrhea, whereas 2,323 children (14.80%) did have diarrhea.

Table 3: Weighted frequency and percentage of diarrhea versus individual level predictors

Covariates	Total number of children		Diarrhea status			
			No		Yes	
Sex of child	Frequency	Percent	Frequency	Percent	Frequency	Percent
Male	7983	50.84	6760	50.53	1223	52.63
Female	7720	49.16	6620	49.47	1100	47.37
Age of child in Months						
<6 months	1653	10.53	1418	10.60	235	10.12
6-11	1770	11.27	1362	10.18	408	17.56
12-23	3139	19.99	2396	17.91	743	31.98
24-35	2966	18.89	2550	19.06	416	17.91
36-47	3146	20.03	2821	21.08	325	13.99
48-59	3029	19.29	2833	21.17	196	8.44
Wealth index						
Poorest	3653	23.26	3100	23.17	552	23.76
Poorer	2862	18.23	2451	18.32	411	17.70
Middle	2712	17.27	2313	17.28	400	17.22
Richer	3112	19.82	2609	19.50	503	21.65
Richest	3364	21.42	2907	21.73	457	19.67

Breastfeeding						
No Yes	7484 8219	47.66 52.34	6547 6833	48.93 51.07	937 1386	40.34 59.66
Number of U-5						
<=2 >=3	13727 1976	87.42 12.58	11624 1756	86.88 13.12	2103 220	90.53 9.47
Age of Mother						
15-24 25-34 35-49	4137 8078 3488	26.35 51.44 22.21	3343 6928 3109	24.98 51.78 23.24	793 1150 380	34.14 49.50 16.36
Education of mother						
No education Primary Secondary Higher	1696 5965 5261 2781	10.80 37.99 33.50 17.71	1467 5033 4427 2453	10.96 37.62 33.09 18.33	230 932 833 328	9.90 40.12 35.86 14.12
Mother work status						
No Yes	7718 7985	49.15 50.85	6561 6819	49.04 50.96	1157 1166	49.81 50.19
Family size						
<=5 >5	8992 6711	57.26 42.74	7564 5816	56.53 43.47	1428 895	61.47 38.53
Birth order						
First Second Third and above	4396 3720 7587	27.99 23.69 48.32	3637 3153 6590	27.18 23.57 49.25	758 567 998	32.63 24.41 42.96
Media exposure						
No Yes	3725 11978	23.72 76.28	3133 10247	23.42 76.58	592 1731	25.48 74.52
Rotavirus vaccinated						
No Yes	7212 8491	45.93 54.07	6578 6802	49.16 50.84	634 1689	27.29 72.71

In the above table the percentages are tabulated with column wise to consider the total population to compare the result the category with same total population.

From the above output, we observed that out of the total children included in the study, 7983 were male and 7720 were female. The incidence of diarrhea among females (47.37%) was found to be lower than males (52.63%). The study revealed that the majority of children were under the

age of 12-59 months. Among children in the age group12-23 months, the prevalence of diarrhea was the highest (31.98%) compared to other age categories.

Most of the children (8219) were breastfeed, and the prevalence of diarrhea was higher among them when compared to not breastfeed. Children who lived in households with two or fewer children under the age of five had a higher incidence of diarrhea compared to those living in households with three or more children under the age of five.

Out of the total number of children, 3653 (23.26%) were born into the poorest families, while 2712 (17.27%) were born into the middle economic status families. The prevalence of diarrhea was highest among childrens who have poorest family as compared to others.

The largest proportion of children, accounting for 51.44% was born to mothers aged 25–34 years. Children born to the mothers in the age group of 25-34 were highly exposed to diarrhea as compared with the other age groups. Most of the children (5965) were born from mothers whose education level was primary and the occurrence of diarrhea is highest among them in Kenya. Out of the total children, 7985 were born to mothers who worked and the occurrence of diarrhea was higher among them.

Most of the children lived with family size less than or equal to 5, and the occurrence of diarrhea was higher among them. Most of the children were born under third and above birth birth order for their family and the occurrence of diarrhea was highest among them.

Out of the total children, 76.28% were coming from family's that had access to mass media exposure and the occurrence of diarrhea was higher among them.

Most of the children (54.07) were vaccinated with rotaviruses, but the occurrence of diarrhea was higher among them when we compared to not vaccinated children in Kenya.

Table 4: Weighted Frequency and percentage of diarrhea versus community level predictors

Covariates	Total number of children		Diarrhea status			
			No		Yes	
Residence	Frequency	Percent	Frequency	Percent	Frequency	Percent
Urban	5741	36.56	4838	36.16	903	38.87

Rural	9962	63.44	8542	63.84	1420	61.13
Toilet facility						
Improved	10353	65.93	8817	65.90	1536	66.12
Unimproved	5350	34.07	4563	34.10	787	33.89
Source of water						
Improved	11990	76.35	10176	76.05	1814	78.09
Unimproved	3713	23.65	3204	23.95	509	21.91
Province						
Cost	1449	9.23	1222	9.13	227	9.77
North	604	3.85	537	4.01	67	2.88
Eastern	1853	11.80	1628	12.17	224	9.64
Central	1764	11.23	1498	11.20	267	11.49
Rift Valley	4856	30.92	4145	30.98	711	30.61
Western	1501	9.56	1226	9.16	274	11.80
Nyanza	1880	11.97	1612	12.05	268	11.54
Nairobi	1796	11.44	1512	11.30	285	12.27

From the above output, the majority of the children, 9962 (63.44%), were from rural areas, while the rest, 5741 (36.56%), were from urban areas of Kenya. Children living in rural areas were highly exposed to diarrheal diseases as compared with those living in urban areas (38.87% versus 61.13%), respectively. The majority of the children used water from improved drinking water sources and improved toilet facilities. The prevalence of diarrhea was higher among them when compared to those who used unimproved water sources and unimproved toilet facilities.

Most of the children, 4856 (30.92%), were from the Rift Valley province of Kenya, while the last, 1449 (9.23%), were from the cost province of Kenya. The prevalence of diarrhea varies from province to province. The highest diarrhea prevalence (30.60%) was in Rift Valley Province, and the lowest was in cost province (9.77%).

4.2 Results of Inferential Statistics

From the result of backward variable selection except breastfeeding the p-value of all independent variables were less than 0.25, so those independent variables were included in the multilevel binary logistic analysis.

4.2.1 Multilevel analysis results

Table 5: Results of intercept-only model

	Coef.	Std. Err.	z	P>|z|	[95% CI]
Constant	-2	0.04	-54.89	0.00	[-2.02, -1.88]
Random effect Var(EA)	0.52	0.05			[0.43, 0.63]

From the above result the intercept term represents the baseline log odds of the diarrhea when all predictor variables are zero. The intercept term captures the overall average log odds of diarrhea across all clusters without considering any specific predictors. As we have seen from the above the constant is statistically significant. The random effect associated with intercept captures how much individual groups deviate from the overall intercept value and the confidence interval of the random effect does not include zero. This means there was a significant variability in random effect associated with intercept, suggesting that there are cluster level differences in log odds that cannot be explained by fixed effect alone.

Table 6 : Results of intraclass correlation

Level	ICC	Std. Err.	95% CI
EA	0.174	.014	[0.15, 0.20]

As we see from the above output, the confidence interval of intraclass correlation does not include zero meaning that it is significant and intra-class correlation coefficient (ICC) =0.174, indicates the proportion of variance of the occurrence of diarrhea between EA. This means that around 17.4% of the variation is between EA whereas the remaining 82.6% of the variation is within EA.

Table 7: Results of Random Intercept Model

Covariates	Coef.	Std. Err.	P-value	AOR	95% CI(AOR)
Constant	-1.36	0.20	0.00	0.26	[0.17, 0.38]
Child sex (ref: male)					
Female	-0.11	0.05	0.03	0.90	[0.82, 0.99]
Child Age (ref: < 6 months)					
6-11	0.46	0.10	0.00	1.59	[1.29, 1.95]
12-23	0.45	0.10	0.00	1.57	[1.30, 1.91]

24-35	-0.26	0.10	0.01	0.77	[0.63, 0.94]
36-47	-0.03	0.13	0.79	0.97	[0.75, 1.24]
48-59	-0.58	0 .14	0.00	0.56	[0.43, 0.73]
Birth order (ref: first)					
Second	-0.17	0.07	0.02	0.85	[0.73, 0.97]
Third and above	-0.29	0.08	0.00	0.75	[0.63, 0.88]
Region (ref: cost)					
North	-0.28	0.21	0.18	0.75	[0.50, 1.14]
Eastern	-0.23	0.14	0.10	0.79	[0.60, 1.05]
Central	-0.01	0.15	0.97	0.99	[0.74, 1.33]
Rift valley	0.03	0.12	0.79	1.03	[0.81, 1.31]
Western	0.35	0.15	0.02	1.42	[1.06, 1.91]
Nyanza	0.00	0.14	0.99	1.00	[0.76, 1.32]
Nairobi	0.14	0.17	0.42	1.15	[0.82, 1.62]
Residence (ref: urban)					
Rural	-0.27	0.10	0.01	0.77	[0.64, 0.93]
Mother education (ref: First)					
Primary	-0.08	0.11	0.47	0.92	[0.75, 1.14]
Secondary	-0.30	0.12	0.01	0.74	[0.59, 0.94]
Higher	-0.62	0.14	0.00	0.54	[0.42, 0.71]
Wealth index (ref: Poorest)					
Poorer	-0.09	0.09	0.33	0.92	[0.77, 1.09]
Middle	-0.02	0.10	0.82	0.98	[0.80, 1.20]
Richer	-0.11	0.12	0.36	0.89	[0.70, 1.14]
Richest	-0.35	0.15	0.02	0.71	[0.53, 0.94]
Mothers work(ref: no)					
Yes	0.07	0.05	0.22	1.07	[0.96, 1.19]
Toilet(ref: improved)					
Unimproved	-0.10	0.07	0.16	0.91	[0.79, 1.04]
Water source (ref: improved)					
Unimproved	-0.06	0.07	0.40	0.94	[0.82, 1.08]
Number of U-5 (ref: <=2)					
>=3	-0.36	0.09	0.00	0.70	[0.59, 0.83]
Family size (ref: <=5)					
>5	-0.05	0.06	0.44	0.96	[0.85, 1.07]
Mother age (ref: 15-24)					
25-34	-0.07	0.07	0.31	0.93	[0.81, 1.07]
35-49	-0.28	0.10	0.00	0.75	[0.62, 0.92]
Media Exposure (ref: no)					
Yes	-0.18	0.07	0.01	0.83	[0.72, 0.96]
Rotavirus vaccinated (ref: no)					
Yes	0.60	0.12	0.00	1.83	[1.44, 2.32]
Var (EA)	0.56	0.06		0.56	[0.46, 0.68]

Table 8 Results of Random Coefficient Model

Covariates	Coef.	Std.Err.	P-value	AOR	95% for AOR
Constant	-1.59	0.21	0.00	0.20	[0.14, 0.31]
Child sex (ref: male)					
Female	-0.08	0.05	0.14	0.93	[0.84, 1.02]
Child Age (ref: < 6 months)					
6-11	0.46	0.11	0.00	1.59	[1.29, 1.96]
12-23	0.46	0.10	0.00	1.58	[1.30, 1.92]
24-35	-0.27	0.10	0.01	0.76	[0.62, 0.94]
36-47	-0.04	0.14	0.78	0.96	[0.73, 1.26]
48-59	-0.59	0.15	0.00	0.56	[0.42, 0.74]
Birth order(ref: first)					
Second	-0.17	0.07	0.024	0.85	[0.73, 0.98]
Third and above	-0.29	0.09	0.00	0.75	[0.63, 0.89]
Region (ref: cost)					
North	-0.27	0.22	0.22	0.77	[0.50, 1.17]
Eastern	-0.24	0.14	0.10	0.79	[0.59, 1.05]
Central	-0.03	0.15	0.83	0.97	[0.72, 1.30]
Rift Valley	0.02	0.12	0.87	1.02	[0.80, 1.30]
Western	0.36	0.15	0.02	1.44	[1.07, 1.94]
Nyanza	-0.02	0.14	0.90	0.98	[0.74, 1.30]
Nairobi	0.07	0.18	0.70	1.07	[0.75, 1.52]
Residence (ref: urban)					
Rural	-0.27	0.10	0.01	0.76	[0.63, 0.92]
Mother education (ref: no education)					
Primary	-0.09	0.11	0.43	0.92	[0.74, 1.14]
Secondary	-0.31	0.12	0.01	0.74	[0.58, 0.93]
Higher	-0.63	0.14	0.00	0.54	[0.41, 0.71]
Wealth index (ref: poorest)					
Poorer	-0.08	0.09	0.40	0.93	[0.77, 1.11]
Middle	-0.02	0.11	0.87	0.98	[0.80, 1.21]
Richer	-0.11	0.13	0.38	0.89	[0.70, 1.15
Richest	-0.36	0.15	0.02	0.70	[0.52, 0.94]
Mother work (ref: no)					
yes	0.07	0.06	0.24	1.07	[0.96, 1.19]
Toilet facility (ref: improved)					
Unimproved	-0.09	0.07	0.23	0.92	[0.80, 1.06]
Water source (ref: improved)					
Unimproved	-0.07	0.07	0.34	0.93	[0.81, 1.08]
Number of U-5 (ref: < 3)					
>= 3	-0.37	0.09	0.00	0.69	[0.58, 0.83]
Family size (ref: <= 5)					
1>5	-0.04	0.06	0.46	0.96	[0.85, 1.08]

Mother age (ref: 15-24)					
25-34	-0.09	0.07	0.23	0.92	[0.79, 1.06]
35-49	-0.31	0.10	0.00	0.73	[0.60, 0.90]
Media exposure (ref: no)					
Yes	-0.18	0.07	0.01	0.83	[0.72, 0.96]
Rotavirus vaccinated (ref: no)					
Yes	0.83	0.14	0.00	2.30	[1.74, 3.05]
Var(Rotavirus)	1.16	0.19		1.16	[0.85, 1.60]
Var(EA)	1.37	0.18		1.37	[1.06, 1.77]
Cov(Rotavirus,_EA)	-0.97	0.17	0.00	0.97	[-1.30, -0.64]

The result in table 8 is different from the result in table 7 by varying rotavirus across EA. In a random coefficient model, the coefficient of the model is allowed to vary randomly across different units of analysis. This type of model is particularly useful when there is heterogeneity in the data that cannot be explained by fixed effect alone. So we conduct this model by considering rotavirus as random coefficient that can have different effect on diarrhea in different EA. when examining how rotavirus varies across EA, one would look at how the estimated coefficient for rotavirus changes from one area to another. In our finding above, a positive coefficient of rotavirus indicated that an increase in rotavirus is associated with increase in the occurrence of diarrhea.

Interpretation of the results in table 8

We would like to mention that all comparison of AOR and 95% confidence interval for AOR are made by keeping the remaining variables constant. Therefore we will not repeat the phrase "keeping other variable constant".

Our finding in above table revealed that, child age influenced the occurrence of diarrhea. The odds of being infected by diarrhea for childrens age 6 to 11 months were higher by 1.59 times compared to childrens age less than 6 months. That means childrens under the age less than 6 months were less affected by diarrhea compared to childrens age 6 to 11 months. The odds of being infected by diarrhea for childrens age 12 to 23 months were higher by 1.58 times compared to childrens age less than 6 months. The odds of being infected by diarrhea for childrens age 24 to 35 months were lower by 0.76 times compared to childrens age less than 6

months. The odds of being infected by diarrhea for childrens age 48 to 59 months were lower by 0.56 times compared to childrens age less than 6 months.

The odds of being infected by diarrhea for second birth order children were lower by 0.85 times compared to childrens first birth order. The odds of being infected by diarrhea for third and above birth order children were lower by 0.75 times compared to children first birth order.

The result showed that, province in which a child lived influenced the occurrence of childhood diarrhea: The odds of being infected by diarrhea for children lived in western province of Kenya were higher by 1.44 times compared to childrens lived cost province.

Place of residence also affect the occurrence of diarrhea. The odds of being infected by diarrhea for children lived in rural area of Kenya were lower by 0.76 times compared to childrens lived in urban area of Kenya.

The educational background of mothers in Kenya had a significant impact on the prevalence of childhood diarrhea. Children whose mothers had higher levels of education were 0.54 times less likely to be affected by diarrhea compared to children whose mothers had no formal education. Children whose mothers had a secondary level of education were 26% less likely to be infected by diarrhea compared to children whose mothers had no education.

The occurrence of childhood diarrhea was influenced by the household wealth index. Children from the richest households had 0.70 times lower odds of experiencing diarrhea compared to children from the poorest households. Children living with three or more under five children in the household had a 0.69 times lower likelihood of experiencing diarrhea attacks compared to those living with one or two under five children.

The findings indicated that maternal age had an impact on the prevalence of childhood diarrhea: The odds of being infected by diarrhea for children whose mothers were in the age group of 35 to 49 were lower by 0.73 times compared to children whose mothers were in the age group of 15 to 24.

Children who had access to social media had a 0.83 times lower risk of being infected by diarrhea compared to children who did not have access to social media.

The odds of being attacked by diarrhea for children got rotavirus vaccination was higher by 2.30 times compared to childrens who were not vaccinated with rotavirus. In this case, an odds ratio of 2.3 means that children who have been vaccinated against rotavirus are 2.3 times more likely to experience diarrhea compared to those who have not been vaccinated. This may be due to incomplete protection, vaccine failure, delayed immune response and environmental factors such as poor hygiene practice, contaminated food and exposure to infectious agents in the communal setting.

Table 9 : Results of model selection criteria

Model	LL	AIC
Intercept only model	-6417.51	12839.02
Random intercept model	-6048.92	12165.85
Random coefficient model	-6002.79	12077.58

The model with smaller value of AIC is better. The random coefficient model, which had the lowest AIC value of 12077.58, demonstrated a better fit compared to the other two models. Hence, we prefer the random coefficient model over the others.

4.2.2 Model adequacy checking results

Table 10: Results of multicollinearity diagnostics

Predictors	VIF	Tolerance	R-Squared
Child sex	1.00	1.00	0.00
Child age	1.69	0.59	0.41
Birth order	1.97	0.51	0.49
Region	1.15	0.87	0.13
Residence	1.80	0.56	0.44
Mother education	2.08	0.48	0.52
Wealth index	3.74	0.27	0.73
Mothers work	1.21	0.83	0.17
Toilet facility	1.58	0.63	0.37
Water source	1.19	0.84	0.16
Number of U-5	1.21	0.83	0.17
Family size	1.31	0.76	0.24
Mother age	1.75	0.57	0.43
Media exposure	1.63	0.61	0.39
Rotavirus	1.63	0.61	0.39

As see in table 10, all VIF values are less than 10 that shows there was no multicollinearity. In column two above we had seen the values of tolerance which is measure of the degree to which an independent variable is not linearly related to other independent variables in the model. It is calculated as 1 - R^2, where R^2 is the coefficient of determination obtained when the X_i variable is regressed on the remaining explanatory variables. Researchers often recommend a minimum tolerance level of 0.10 (Tabachnick & Fidell, 2001). So the tolerance values in the above were greater than 0.10. A low R-square value, like 0.00 for child sex in the above table suggests that the variable in question is not strongly correlated with any other independent variable.

4.2.3 Model goodness of fit test result

Goodness of fit test for this study was assessed by using likelihood ratio test (LRT). If the likelihood ratio (LR) statistic, which measures the relative likelihood of the random coefficient and random intercept models, is significant, it suggests that the random coefficient model provides a better fit to the data than the null model. Accordingly, the LRT test is given as follow.

Likelihood-ratio test LR chi2(2) = 92.30

(Assumption: ri nested in rc) Prob > chi2 = 0.0000

Where ri is the random intercept model and rc is the random coefficient model.

As seen from this result the p-value is significant that means the random intercept model is nested in random coefficient model.

CHAPTER FIVE

5.1 Discussion

The objective of this research was to assess the prevalence of childhood diarrhea in different region of Kenya and to analyze covariates that contribute to childhood diarrhea in children under the age of five in Kenya, using data from the KDHS 2022. To achieve this, descriptive statistics and multilevel binary logistic regression techniques were utilized.

The result of our study revealed that child age has a significant effect on childhood diarrhea in Kenya.This finding is similar to Berde et al., (2018), Mbugua et al., (2014), Mosisa et al., (2020), Mulatya and Ochieng , (2020), Asfaha et al., (2018), Shine et al., (2020), Owusu et al., (2020) and contradict to Kimani et al., (2019). Children in the age of 6 to 23 months were at high risk of diarrhea. This may be children between 6-23 months are typically more mobile and curious, leading them to explore their environment, including touching surfaces and objects that may be contaminated with pathogens. This behavior increases their exposure to infectious agents that can cause diarrhea.

Our finding revealed that child birth order has a negative effect on childhood diarrhea in the study area that means when birth order increase the occurrence of diarrhea decreases. This may be mother's lack of experience to care their first born child from diarrhea. This result is contradicted to Shine et al., (2019).

Region is also found to be significantly related to childhood diarrhea in Kenya. Children who were lived in western provenance are at high risk of diarrhea.

Another most important determinant variable with negative effect to diarrhea in Kenya is education level of mother. When education level of mother increase the occurrence of diarrhea decrease. A child whose mother is not educated was at high risk of diarrhea. This finding is similar to Mbugua et al., (2014), Asfaha et al., (2018), Afitiri et al., (2020), Tesfaye et al., (2020) and Owusu et al., (2020).

Wealth index of the household also has negative effect on the occurrence of diarrhea. Children in the richest household economic status are found to be at low risk of diarrhea than children in poorest household economic status. This finding is similar to Owusu et al., (2020).

Number of U-5 children in the household also has negative association to the occurrence of diarrhea that means when number of U-5 children in the household increased the occurrence of diarrhea decrease. This finding is similar to Tambe et al., (2015) and contradicts to Asfaha et al., (2018)

Mother age is also found to be significantly related to childhood diarrhea in Kenya. When mother age increased occurrence of childhood diarrhea decreased. This may be mother get experience to protect childrens from diarrhea. This finding is similar to Mbugua et al., (2014) and Afitiri et al., (2020).

Exposure to mass media also has negative association with diarrhea. Children coming from household that have access to mass media were at lower risk of diarrhea compared to children coming from households that have not access to mass media exposure.

Rotavirus has positive association to the occurrence of diarrhea which is unexpected. This finding is similar to Shine et al., (2022) and contradicts to Mosisa et al., (2020).

CHAPTER SIX

6.1 Conclusions

The research findings indicated that both community level and individual level factors play a crucial role in the prevalence of diarrhea among children under the age of five in Kenya. Child age, child birth order, region of residence, maternal education, household wealth index, number of U-5 children in the household, maternal age, media exposure, and rotavirus vaccination were found to have a statistically significant impact on the occurrence of childhood diarrhea in Kenya.

The study found that the prevalence of diarrhea in different province of Kenya varied significantly: 9.77% in the Coast, 2.88% in the North, 9.64% in the Eastern 11.49% in the Central, 30.61% in the Rift valley, 11.80% in the Western, 11.54% in Nayanza, and 12.27% in Nairobi. The discrepancy in diarrhea prevalence between EA and within EA was 17.4% and 82.6% respectively.

Child age emerged as a crucial factor influencing the occurrence of diarrhea in under-five children in Kenya. The research indicates that younger children are more prone to experiencing diarrhea compared to older children. Birth order was found to have an impact on the prevalence of diarrhea among under-five children in Kenya. The research suggests that children's birth order within a family can influence their susceptibility to diarrheal infections. The region of residence was identified as a significant determinant of diarrhea among under-five children in Kenya. Disparities in access to clean water, sanitation facilities, and healthcare services across different regions were linked to varying rates of diarrheal illnesses.

The education level of mothers was shown to play a crucial role in influencing the occurrence of diarrhea among under-five children in Kenya. Higher maternal education was associated with better knowledge and practices related to child health and hygiene. Wealth index was identified as a determinant affecting the prevalence of diarrhea among under-five children in Kenya. Socioeconomic status significantly impacts access to clean water, nutritious food, healthcare services, and overall living conditions, which are key factors contributing to diarrheal risks. Mother's age was identified as a factor influencing the prevalence of diarrhea among under-five

children in Kenya. Younger or older maternal age groups were associated with different levels of knowledge, caregiving practices, and access to healthcare services, all impacting child health outcomes. Media exposure was recognized as a determinant affecting the occurrence of diarrhea among under-five children in Kenya. Access to information through various media channels influences parental awareness, behavior change communication, and adoption of recommended practices related to child health and hygiene.

6.2 Recommendation

Based on the findings of the research, it is recommended that public health interventions in Kenya focus on addressing these determinants to reduce the burden of diarrhea among under-five children. Specific recommendations include:

Public health initiatives should prioritize early childhood interventions to prevent and manage diarrhea cases effectively. Targeted strategies for infants and toddlers could help reduce the burden of diarrheal diseases in the age group.

Health education programs should emphasize proper hygiene practices and sanitation measures, especially for younger siblings in larger families. Addressing birth order dynamics may contribute to lowering the incidence of diarrhea in these households.

Policy makers and public health authorities should focus on improving infrastructure and healthcare services in regions with higher prevalence rates of diarrhea.

Investing in maternal education through awareness campaigns and educational programs can empower mothers with essential knowledge on preventing diarrheal diseases. Enhancing maternal literacy levels can lead to improved health outcomes for young children.

Efforts should be made to address socioeconomic inequalities by implementing targeted interventions that provide vulnerable families with resources for improved hygiene practices, sanitation facilities, and healthcare access.

Communication strategies leveraging media platforms should be utilized to disseminate key messages on diarrheal disease prevention, treatment options, and healthy behaviors for parents/caregivers.

References

Afitiri, A.-R., Armah, F. A., Ekumah, B., nyieku, F. E., yawson, D. O. & Odoi, J. O. 2020. Cumulative effects of environmental factors on household childhood diarrhoea in Ghana. *Water Practice & Technology,* 15**,** 1032-1049.

Agresti., A. 2012. *Categorical data analysis*, John Wiley & Sons.

Ahs, J. W., Tao, W., Löfgren, J. & Forsberg, B. C. 2010. Diarrheal diseases in low-and middle-income countries: incidence, prevention and management. *Open Infect Dis J,* 4**,** 113-24.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control,* 19**,** 716-723.

Asfaha, K. F., Tesfamichael, F. A., Fisseha, G. K., Misgina, K. H., Weldu, M. G., Welehaweria, N. B. & Gebregiorgis, Y. S. 2018. Determinants of childhood diarrhea in Medebay Zana District, Northwest Tigray, Ethiopia: a community based unmatched case–control study. BMC pediatrics, 18, 1-9.

Berde, A., Yalçin, S., Ozcebe, L., Üner, S. & Karadag-caman, Ö. 2018. Determinants of childhood diarrhea among under-five year old children in Nigeria: A population-based study using the 2013 demographic and health survey data. Turkish Journal of Pediatrics, 60.

CDC, Centers for Disease Control and Prevention: [Online] Availabel: https://www.cdc.gov/healthywater/pdf/global/programs/globaldiarrhea508c.pdf

Demissie, G. D., Yeshaw, Y., Aleminew, W. & Akalu, Y. 2021. Diarrhea and associated factors among under five children in sub-Saharan Africa: Evidence from demographic and health surveys of 34 sub-Saharan countries. *Plos one,* 16**,** e0257522.

DHSprogram. 2022. *Kenya: Standard DHS, 2022 Dataset* [Online]. Available: https://dhsprogram.com/data/dataset/Kenya_Standard-DHS_2022.cfm?flag=1 [Accessed].

Fischer Walker, C.L., Perin, J., Aryee, M.J., Boschi-Pinto, C. and Black, R.E., 2012. Diarrhea incidence in low-and middle-income countries in 1990 and 2010: a systematic review. BMC public health, 12, pp.1-7.

Guillaume, D.A., Justus, O.O. and Ephantus, K.W., 2020. Factors influencing diarrheal prevalence among children under five years in Mathare Informal Settlement, Nairobi, Kenya. Journal of public health in Africa, 11(1).

Hosmer JR, D. W., Lemeshow, S. & Sturdivant, R. X. 2013. *Applied logistic regression*, John Wiley & Sons.

KhaliL, I. A.-M., Troeger, C., RAO, P., Forouzanfar, M., Columbara, D., Misra, K. & Mokdad, A. Assessing the Nonfatal Burden of Childhood Diarrhea Including Malnutrition, Physical Growth, and Cognitive Development. Open Forum Infectious Diseases, 2016. Oxford University Press, 1220.

KimanI, H. M., Akunga, D. N., Anyango, S. O. & ndwiga, T. 2019. Diarrhea Risk Factors Associated with Water, Sanitation and Hygiene Among the Under Five in Kasarani, Nairobi County. *Central African Journal of Public Health,* 5**,** 272-279.

Mbugua, S., Musikoyo, E., Ndungi, F., Sang, R., Kamau-mbuthia, E. & Ngotho, D. 2014. Determinants of diarrhea among young children under the age of five in Kenya, evidence from KDHS 2008-09. *African Population Studies***,** 1046-1056.

Mosisa, D., Aboma, M., Girma, T. & Shibru, A. 2021. Determinants of diarrheal diseases among under five children in Jimma Geneti District, Oromia region, Ethiopia, 2020: a case-control study. BMC pediatrics, 21, 1-13.

Mulatya, D. M. & Ochieng, C. 2020. Disease burden and risk factors of diarrhoea in children under five years: Evidence from Kenya's demographic health survey 2014. *International Journal of Infectious Diseases,* 93**,** 359-366.

Mutama, R., Mokaya, D. & Wakibia, J. 2019. Risk factors associated with diarrhea disease among children under-five years of age in Kawangware slum in Nairobi County, Kenya. *Food and Public Health,* 9**,** 1-6.

Owusu, D. N., Duah, H. O., Dwomoh, D. & Alhassan, Y. 2023. Prevalence and determinants of diarrhoea and acute respiratory infections among children aged under five years in West Africa: evidence from demographic and health surveys. International Health, ihad046.

Simen-kapeu, A., Bogler, L., Weber, A.-C., Ntambi, J., Zagre, N. M., Vollmer, S. & Ekpini, R. E. 2021. Prevalence of diarrhoea, acute respiratory infections, and malaria over time (1995-2017): A regional analysis of 23 countries in West and Central Africa. *Journal of global Health,* 11.

Shine, S., Muhamud, S., Adanew, S., Demelash, A. & Abate, M. 2020. Prevalence and associated factors of diarrhea among under-five children in Debre Berhan town, Ethiopia 2018: a cross sectional study. BMC infectious diseases, 20, 1-6.

Tambe, A. B., Nzefa, D. L. & Nicoline, N. A. 2015. Childhood diarrhea determinants in sub-Saharan Africa: a cross sectional study of Tiko-Cameroon. Challenges, 6, 229-243.

Snijders, T. A. & Bosker, R. 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling.

Tesfaye, T. S., Magarsa, A. U. & Zeleke, T. M. 2020. Moderate to severe diarrhea and associated factors among under-five children in Wonago District, South Ethiopia: a crosssectional study. Pediatric Health, Medicine and Therapeutics, 437-443.

Tornheim, J. A., Manya, A. S., Oyando, N., Kabaka, S., O'Reilly, C. E., Breiman, R. F. & Feikin, D. R. 2010. The epidemiology of hospitalization with diarrhea in rural Kenya: the utility of existing health facility data in developing countries. *International Journal of Infectious Diseases,* 14**,** e499-e505.

Twisk, J. W. 2013. *Applied longitudinal data analysis for epidemiology: a practical guide*, cambridge university press.

UNICEF.2024.Diarrhoea[Online].Available: https://data.unicef.org/topic/childhealth/diarrhoeal-disease/ [Accessed].

UNICEF, W. 2017. Why children are still dying and what can be done. 2009. Reterieved from: http://apps. who. int/iris/bitstream/10665/44174/1/978924 1598415_eng. pdf.

Walker, C. L. F., Rudan, I., LIU, L., Nair, H., Theodoratou, E., Bhutta, Z. A., O'brien, K. L., Campbell, H. & Black, R. E. 2013. Global burden of childhood pneumonia and diarrhoea. *The Lancet,* 381**,** 1405-1416.

Wanjala, E. W. A. P. 2022. *Diarrhoea in the 21st Century:* [Online]. Available: http://guidelines.health.go.ke:8000/media/NHA_POLICY_BRIEF_Diarrhoea.pdf [Accessed].

WBB. World Bank Blogs: https://blogs.worldbank.org/opendata/world-water-day-two-billion-people-still-lack-access-safely-managed-water#:~:text=Among%20them%2C%20771%20million%20people,concentrated%20in%20Sub%2DSaharan%20Africa.

WHO. 2019. Available: https://www.who.int/news/item/18-06-2019-1-in-3-people-globally-do-not-have-access-to-safe-drinking-water-unicef-who [Accessed].

WHO.2017.Diarrhoeal disease [Online].World Health Organization.Available: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease [Accessed 2 may 2017].

WHO: 2004 : World Health Orginaization.The global burden of disease : 2004 update.

WHO. 2005. Department of Child and Adolescent Health and Development. The treatment of diarrhoea: a manual for physicians and other senior health workers. 4th rev. ed. Geneva: World Health Organaization 2005.

Wikipedia. Kenya [Online]. Available: https://en.wikipedia.org/wiki/Kenya [Accessed].

Appendix

Table 11: Result of chi-square test of association

Independent variables Versus Diarrhea			
	Pearson chi-square	Df	P-value
Sex of child	4.36	1	0.04
Age of child	455.51	5	0.00
Wealth index	9.82	4	0.04
Breastfeeding	53.93	1	0.00
Number of U-5	19.42	1	0.00
Age of Mother	112.15	2	0.00
Education of mother	26.05	3	0.00
Mother work status	2.23	1	0.14
Family size	18.42	1	0.00
Birth order	44.48	2	0.00
Mass media exposure	7.72	1	0.01
Rotavirus vaccinated	335.32	1	0.00
Residence	7.35	1	0.01
Toilet facility	0.01	1	0.91
Source of water	6.53	1	0.01
Province	29.32	6	0.00

Declaration

I, the undersigned, declare that this thesis proposal is my own work, has not been presented for consideration to this or any other university, and all sources of material used for the proposal have been duly acknowledged.

Name of candidate	**Signature**	**Date**
Kedija Ahmed		
______________________________	________________	_____________________

This thesis has been submitted for consideration with my approval as a university advisor.

Name of advisor	**Signature**	**Date**
Prof. Eshetu Wencheko		
_____________________________	____________________	______________________

.