DEBRE BERHAN UNIVERSITY

ASSOCIATED FACTORS INFLUENCING VACCINATION OF CHILDREN AGED 12 TO 35 MONTHS IN ETHIOPIA: AN APPLICATION OF MULTILEVEL ORDINAL LOGISTIC REGRESSION MODEL

BY

GARDEW BOGALE

A THESIS SUBMITTED AS A PARTIAL FULFILLMENT TO THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN STATISTICS (BIOSTATISTICS)

DEPARTMENT OF STATISTICS

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

COLLEGE OF POST GRADUATE STUDIES

FEBRUARY, 2024

DEBRE BERHAN, ETHIOPIA

DEBRE BERHAN UNIVERSITY

ASSOCIATED FACTORS INFLUENCING VACCINATION OF CHILDREN AGED 12 TO 35 MONTHS IN ETHIOPIA: AN APPLICATION OF MULTILEVEL ORDINAL LOGISTIC REGRESSION MODEL

A MASTER'S THESIS

BY

GARDEW BOGALE

ADVISOR;

AYELE ADIMASU (ASSISTANT PROF.)

FEBRUARY, 2024

DEBRE BERHAN, ETHIOPIA

DECLARATION

I, the undersigned, declare that the thesis entitled on **"Associated Factors Influencing Vaccination of Children Aged 12 to 35 Months in Ethiopia: An Application of Multilevel Ordinal Logistic Regression Model"** is my original work, and has not been presented for achieving any degree or diploma award in this university or any other institutions. All source of materials used for the thesis have been duly acknowledged. The thesis has been submitted in partial fulfillment for the requirements of Masters of Science Degree in Statistics (Biostatistics), Debre Berhan University.

Gardew Bogale Abay ____________________ ______/____/ _______

Name of Advisor Signature Date

Place of submission; Department of Statistics, College of Natural And Computational Science, Debre Berhan University

Date of submission; February, 2024

This thesis has been submitted for examination with my approval as a University advisor.

Ayele Adimasu (assistant prof.) ____________________ ______/____/ _______

Name of Advisor Signature Date

DEBRE BERHAN UNIVERSITY

COLLEGE OF POSTGRADUATE STUDIES

DEPARTMENT OF STATISTICS

APPROVAL SHEET

This is to certify that the thesis entitled on **“Associated Factors Influencing Vaccination of Children Aged 12 to 35 Months in Ethiopia: An Application of Multilevel Ordinal Logistic Regression Model”** submitted in partial fulfillment of the requirements for the degree of Masters of Science in Statistics (Biostatistics) prepared by **Gardew Bogale Abay** under my supervision and no part of the thesis has been submitted for any degree or diploma award in this university or any other institutions. Hence, I recommend that it can be accepted as fulfilling thesis requirements.

Approved by the board of examiners

____________________________	______________________	______/____/ ________
Name of Chairman	Signature	Date
____________________________	______________________	______/____/ ________
Name of Advisor	Signature	Date
____________________________	______________________	______/____/ ________
Name of Internal Examiner	Signature	Date
____________________________	______________________	______/____/ ________
Name of External Examiner	Signature	Date

ACKNOWLEDGEMENT

First and foremost, I would like to thank the almighty God, who helped and blessed me in all my activities, and to the Holy Virgin Mary. Secondly, I would like to express my deepest gratitude to my advisor; Ayele Adimasu (assistant prof.) giving his support, valuable comments and advises by dissipates his golden time throughout the end. Really thank you very much for your involvement in my work and I sincerely thank all other lectures at Debre Berhan University, Statistics Department who helps me with valuable advises for my work.

I would like to thank also my classmates for your valuable contribution by sharing your knowledge from the beginning of data management till the end of my thesis. I would like to thank Dilla University for sponsoring and giving an opportunity to attend postgraduate program.

At last but not the least, I would like to thank all my best friends, my family members spatially for my father and mother for their unreserved encouragement and supports ever time to be my accomplishment successful.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iii
LIST OF TABLES vii
LIST OF FIGURES viii
LIST OF ACRONYM AND ABBREVIATIONS' ix
ABSTRACT x
CHAPTER ONE 1
1. INTRODUCTION 1
1.1. Background of the Study 1
1.2. Statement of the Problem 3
1.3. Objective of the Study 5
1.3.1. General Objective 5
1.3.2. Specific Objective 5
1.4 Significance of the Study 5
1.5 Scope of the Study 6
1.6 Organization of the Paper 6
CHAPTER TWO 7
2. LITERATURE REVIEW 7
2.1. Theoretical literature review 7
2.1.1. Concepts and backgrounds of vaccination status and coverage 7
2.1.2. The EPI program and vaccination coverage in Ethiopia 11
2.1.3. Factors Associated with vaccination incompletion of children 13
2.1.3.1. Geography and governance 13
2.1.3.2. Demographic and socio – economic characteristics of women 13
2.1.3.3. Access 14
2.4. Empirical Literature Review on Vaccination 14
CHAPTER THREE 19
3. METHODOLOGY 19
3.1. Study Design 19
3.2. Source of Data 19

3.3. Data management, processing and analysis methods 20
3.4. Inclusion and Exclusion Criteria 20
3.5. Study Variables 20
3.5.1 Dependent Variable: 20
3.5.2 Independent variables: 21
3.6. Method of Statistical Analysis 22
3.6.1. Logistics Regression Analysis 23
3.6.1.1. Ordinal Logistic Regression Analysis 24
3.6.1.2. The generalized ordinal logistic regression (GOLOGIT) model 28
3.6.1.3. Variable Selection and Model Building Techniques 30
3.6.1.4. Parameter Estimation in Ordinal Logistic Regression Model 30
3.6.1.5. Goodness-of-Fit of the Ordinal logit model 31
3.6.2 Multilevel Analysis 32
3.6.2.1. Multilevel ordinal Logistic Regression Analysis 35
3.6.2.1.1. The empty two level ordinal logistic regression model 37
3.6.2.1.2. Random intercept and fixed effect ordinal logistic regression model 37
3.6.2.1.3. Random coefficient two level ordinal Logistic Regression Model 38
3.6.2.1.4. Intra-class correlation 40
3.6.2.1.5. Parameter Estimation (Maximum likelihood via Quadrature) 40
3.6.2.1.6. Goodness of Fit Test for multilevel ordinal logistic regression model 41
3.6.2.1.7. Test of Overall Model Fit 42
3.8. Model Diagnostics checking 43
3.9. Ethical Consideration 46
CHAPTER FOUR 47
4. RESULTS AND STATISTICAL DATA ANALYSIS 47
4.1 Results for Descriptive Statistics 47
4.2. Ordinal logistics regression analysis 55
4.2.1. Variable selection and Model building method 55
4.2.2. Model comparison for all ordinal logistics regression models 57
4.3. Results of Multilevel ordinal logistics regression model 57
4.3.1. Multilevel ordinal logistics regression model comparison 58

4.3.2. Random coefficient (slope) model 59
4.3.3. Goodness-of-fit test of multilevel random coefficient model 63
4.4. Model Diagnosis checking 64
CHAPTER FIVE 66
5. DISCUSSION, CONCLUSION AND RECOMMENDATION 66
5.1. Discussions 66
5.2. Conclusions 68
5.3. Recommendations 70
5.4. Strength and Limitation of the Study 70
REFERENCE 71
APPENDIX 78

LIST OF TABLES

Table 3. 1: Classification of independent and dependent variables 21
Table 4.1: Proportion of childhood vaccination status, 2019EMDHS……………………47
Table 4.2: Demographic determinant factors of vaccination status 49
Table 4.3: Socio-economic determinant factors of vaccination 50
Table 4.4: Reproductive health and service related determinant factors 52
Table 4.5: Characteristics of the cluster variables 53
Table 4.6: Results of stepwise variable selection for ordinal logistic model 55
Table 4.7: Rests of parallel line regression (PO) assumption 56
Table 4.8: AIC, BIC and -2LL values of all ordinal models 57
Table 4.9: Results of multilevel ordinal logistics model selection criteria 59
Table 4.10: Results of random coefficient model 61
Table A.1: Maximum likelihood parameter estimation of PPOM……………………………78
Table A.2: Results of the empty model with random intercept 80
Table A.3: Test of multi-collinearity [Correlation Matrix] 80

LIST OF FIGURES

Figure 4.1: Childhood vaccination coverage in Ethiopia based on 2019 EMDHS 48
Figure 4.2: Normal probability plot of the standardized residual... 64
Figure 4.3: Plots of standardized residual against predicted probability...................................... 65
Figure 4. 4: Plots of Leverage Value against Predicted Probability... 65

LIST OF ACRONYM AND ABBREVIATIONS'

AIC	Akaike Information Criterion
ANC	Ante Natal Care
BCG	Bacillus Calmette-Guerin
BIC	Bayesian Information Criterion
CI	Confidence Interval
DPT	Diphtheria, Pertussis, Tetanus
DHS	Demographic and Healthy Survey
EA	Enumerative Area
EDHS	Ethiopian Demographic and Healthy Survey
EMDHS	Ethiopian Mini Demographic and Healthy Survey
EPI	Expanded Programme on Immunization
FMOH	Federal Ministry of Health
GAVI	Global Alliance for Vaccine and Immunization
HIB	Haemophilus Influenza Type B
HSDP	Health Sector Development Plan
ICC	Intra Class Correlation
MDG	Millennium Development Goal
ML	Maximum Likelihood
MOH	Ministry of Health
NGO	None Governmental Organization
OPV	Oral Polio vaccines
OR	Odds Ratio
PCV	Proportional Change Variance
PPOM	Partial Proportional Odds Model
Std.Error	Standard Error
SNNP	South Nations Nationalities and People
SPSS	Statistical Package for Social Science
STATA	South Texas Art Therapy Association
TT	Tetanus Toxoid
UNICEF	United Nations Children Fund
USAID	United States Agency for International Development
VPD	Vaccine Preventable Disease

ABSTRACT

Associated Factors Influencing Vaccination of Children Aged 12 to 35 Months in Ethiopia: An Application of Multilevel Ordinal Logistic Regression Model

Gardew Bogale

Debre Berhan University February, 2024

Background*; One of the safest and most economical ways to lower childhood morbidity and death is through vaccination. Even though individual vaccine coverage has increased in Ethiopia, it is still uncommon to see a child completely vaccinated with all recommended vaccines. Therefore, the main objective of this study was to identify associated factors influencing vaccination of children aged 12 to 35 months in Ethiopia.*

Method*; Vaccination status was examined in a sample of 5,753 children aged 12–35 months from the 2019 EMDHS data. The study used percentages to show the prevalence of vaccination coverage among children of aged 12 – 35 month in Ethiopia. Multilevel ordinal logistic regression modes were constructed, and the best-fitting model was selected to identify variables that significantly associated with childhood vaccination and assess regional variability of childhood vaccination in Ethiopia.*

Results*; The prevalence of childhood vaccination coverage among children was 83.73% (20.91% completely vaccinated and 62.82% partially vaccinated), while about 16.27% of children were non-vaccinated. From the multilevel ordinal logistic regression models, it was found that the random coefficient ordinal logistic model is the best among all the ordinal logistic models and the result revealed that 64.3% of the community-level variation on childhood vaccination has been explained by the combined factors at both the individual and community levels this imply that there exist a clear difference in childhood vaccination incompletion across regions in Ethiopia. The fitted random coefficient model showed that childhood vaccination was significantly associated with age, region, residence, religion, number of antenatal care visit, place of antenatal care visit, wealth index, and place of delivery at 5% level of significance. Moreover, the effect of these significant variables on childhood vaccination is the same for each region except place of delivery which was differed between regions.*

Conclusion*; based on the findings; treatments are better to focus on targeting essential childhood vaccinations by promoting rural maternal access to health care service that are included the low economic family, because the finding showed low vaccination coverage in the rural resident and poor economic level.*

Key words; Vaccination, Multilevel Ordinal Logistic Model, Partial Proportional Odds

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the Study

Globally, vaccination prevents an estimated 2-3 million deaths from measles, pertussis (whooping cough), tetanus, and diphtheria annually (WHO, 2009). Almost 5.2 million children died in 2019, and 14,000 children die every day worldwide. Approximately 80 percent of the 5.2 million child fatalities have happened in Central and Southern Asia and sub-Saharan Africa. The continent of Sub-Saharan Africa continues to have the highest rate of child mortality worldwide (Khazaei Z, et al, 2018).

The Global Vaccine Action Plan (GVAP) endorsed by the 194 Member States of the World Health Assembly in May 2012, is a framework to prevent millions of deaths by 2020 through more equitable access to existing vaccines for people in all communities. It established a common vision and a forum in which immunization stakeholders could collectively discuss matters of concern, as well as a mechanism to connect the activities of global partners. And it acted as a key focal point, maintaining the high global profile of immunization. GVAP was developed to help realize the vision of the Decade of Vaccines, that all individuals and communities enjoy lives free from vaccine preventable diseases (GVAP, 2020).

Vaccination has a significant potential to enhance people's health. It is thought to be crucial for raising the rate of child survival. Public vaccination programs have helped to lower vaccine preventable illnesses and mortality worldwide since their start. It is revered as one of the greatest medical achievements of the 20th century, having saved more lives than any other medical procedure. Furthermore, vaccination is said to be second only to clean infection in terms of lowering the prevalence of infectious diseases. It is claimed to be the most effective and economical way to manage certain illnesses. This is evident in the drastic decline, and in some cases elimination, of certain infectious diseases since the introduction of vaccines in the 20th century (JAMA, 2015).

Even though the third dosage of the diphtheria, tetanus, and pertussis vaccination (DTP) was more widely distributed globally in 2011 up from 5% in 1974 roughly 25% of children worldwide had not finished the three-dose DTP series. Similar to projections from 2010 and 2011, 83% of newborns globally were anticipated to have gotten at least three doses of the DTP immunization in 2012. Of the 1974 WHO member states, 131 (68%) had DTP3 coverage of 90% or higher in every district, while 59 (30%) had DTP3 coverage of 80% or higher in every district. The six most prevalent vaccine-preventable diseases that mostly impact children are covered by the EPI. Along with other illnesses, the EPI program now covers tuberculosis, measles, poliomyelitis, whooping cough, diphtheria, and tetanus. Hepatitis and yellow fever are two of these. (WHO, 2009).

Eight vaccines are currently administered by the EPI: the measles vaccine, the tetanus toxoid (TT) vaccine, the hepatitis B (HepB) vaccine, the hemophilia's influenza type B, the oral polio vaccine (OPV), the diphtheria pertussis-tetanus (DPT) vaccine, and the yellow fever vaccine. There are now eight vaccine-preventable diseases in Ethiopia according to the EPI program. And immunizations are administered regularly. Routine immunizations are administered from birth and must be finished by the time a child turns one year old (EDHS, 2011). Of the 22.6 million infants who were not given three DTP doses in their first year of life, 12.4 million (or 55%) were residents of three different nations. These are Nigeria (17%), Indonesia (7%) and India (30%), and 16.3 million people (72%), were residing in ten nations. Roughly 10 million children (44%) began the three-dose DTP series but did not finish it, and an estimated 12.6 million children (56%) did not receive the initial DTP dosage (UNICEF & WHO, 2009). One significant finding is that over 70% of these kids reside in ten nations (five of which are in Africa): Afghanistan, Chad, Ethiopia, Democratic Republic of the Congo, India, Indonesia, Nigeria, Pakistan, Philippines, and South Africa. This makes these nations the center for these highly contagious diseases and presents difficulties for efforts to eradicate them (WHO, 2012).

Children who were either not vaccinated at all or only received a portion of the required vaccinations run a higher risk of developing VPD and passing on illnesses to younger children, adults who are contraindicated for vaccines, and individuals who have vaccine failure (Borba RC, et al, 2015).

Every day, illnesses that are easily preventable claim the lives of about 704 children (Nour TY, et al, 2020). One of the safest and most economical ways to lower childhood morbidity and death is through vaccination. An estimated 2.5 million fatalities among children under five are avoided annually because to vaccinations. 19.4 million babies did not obtain the recommended vaccinations in 2019. (Mekonnen ZA, et al, 2020). Ethiopia ranks fifth globally in terms of the proportion of unvaccinated children, more than 1.2 million children did not receive the first dose of the measles vaccination in 2018 and over 9,000 children did not receive the third dose of the pentavalent vaccine (Mekonnen ZA, et al, 2020). It will need long-term improvements in service delivery to shield Ethiopian children from needless pain and death. (Fenta S, et al, 2021).

Even though individual vaccine coverage has increased in Ethiopia, it is still uncommon to see a child completely vaccinated with all recommended vaccines (EMOH, 2018). Ethiopia has some of the highest global rates of disability and infant mortality. The complete childhood vaccination coverage in many parts of Ethiopia is far from optimal. This study used multilevel analyses to deal with the hierarchical nature of the EMDHS data and reduce regional discrepancies in childhood vaccination incompletion among children of aged 12 – 35 month in Ethiopia. Therefore, the main objective of this study was to assess the coverage and factors associated with childhood vaccination incompletion in Ethiopia.

1.2. Statement of the Problem

One of the most important strategies for accomplishing the Sustainable Development Goals (SDGs) is vaccination, particularly the one pertaining to lowering the death rate among under five children (UNICEF & WHO, 2009). In low and middle-income nations, childhood vaccinations have been demonstrated to be successful in preventing diseases that can be prevented by vaccines. Each year, vaccinations save the lives of about 2.5 million children (Maman, K et al, 2005). Delayed vaccinations can increase a child's chance of contracting a disease.

Ethiopia updated its health policy in 1997; and Health Sector Development Plan (HSDP) was introduced as the framework for implementing the health strategy. The development of the preventive, promotional, and curative components of health care, ensuring that health care is accessible to all population segments, and encouraging the involvement of the private sector and non-governmental organizations in the health sector are the main goals of the health policy. The EPI in Ethiopia focuses on eight illnesses: measles, poliomyelitis, diphtheria, pertussis, tetanus, hepatitis B, and yellow fever (WHO, 2015).

The percentage of children who have received the necessary number of vaccine doses, regardless of age during vaccination, is the accepted indicator of vaccination coverage (Luman, E. et al, 2008). The national understanding of the social factors of low vaccination coverage and non-uptake of immunization services is lacking, notwithstanding a small number of research conducted in various parts of Ethiopia. Numerous investigations have been carried out in several developing nations, to determine the factors influencing children vaccination rates (Abadura et al., 2015). Some of those studies was used binary logistic regression model (i.e., fully vaccinated or not fully vaccinated), in every instance. When it comes to binary logistic regression, infants who have received one or more vaccination are deemed to be non-vaccinated if they do not meet the necessary conditions. This gives researchers enough data to examine the vaccination pattern of these infants. However, taking into account natural ordering, a child's vaccination status is typically categorized as completely vaccinated, partially (incompletely vaccinated), and not vaccinated. The researcher used the multilevel ordinal logistic regression model to determine the factors that influence childhood vaccination in order to provide an answer. Additionally, the assumptions of regression's independence are compromised when a standard logistic regression analysis approach is used to examine data in a hierarchical arrangement, such as children nested within communities [regions]. In order to overcome these constraints and evaluate the important influence of both individual and community-level factors, this study employed a multi-level ordinal logistic regression analysis.

In order to determine the factors that contribute to vaccination incompletion in Ethiopia, several earlier investigations were carried out in various contexts (Aragaw K, 2014), (Sisay, H, 2018), (Metkie, K. A. et al, 2023), (Abadura et al., 2015) and (Gizachew W, 2020). Nevertheless, most of the research done in this field did not take into consideration the clustering structure of the EDHS data, and the majority of the studies are thought to be illogical since they suggest that children who are partially vaccinated and those who are not vaccinated have identical risks of contracting disease. Therefore, no researches are done in Ethiopia using the 2019 Ethiopia Mini Demographic and Health Survey data by applying multilevel ordinal logistics regression. And hence, the purpose of this study was to use multilevel ordinal logistic regression to discover the factors that influence vaccination coverage among Ethiopian children aged 12 – 35 months and attempt to explore factors influencing vaccination incompletion and assessing regional variation of vaccination coverage taking in to account various demographic, socio-economic, health and service related variables. Relevant data regarding vaccination of 12 – 35 month-old children were retrieved from the 2019 EMDHS.

In line with the objectives of this study, the following research questions have been addressed:-

1. What are the key socio-economic and demographic predictors of childhood vaccination incompletion in Ethiopia?
2. Is there significant variation in children's vaccination incompletion across regions?
3. What factors have made significant contribution to the variations of childhood vaccination incompletion among regions of Ethiopia?

1.3. Objective of the Study

1.3.1. General Objective

The general objective of this study is to identify associated factors influencing vaccination of children aged 12 to 35 months in Ethiopia by applying a multilevel ordinal logistic regression model.

1.3.2. Specific Objective

The specific objectives of this study which should be accomplished to achieve the general objective stated above are;-

- To identify key determinant factors that may influence childhood vaccination
- To examine the extent of variability of incomplete vaccination within and between regions for under-five children in Ethiopia.
- To identify the factor(s) those explain the variation of vaccination incompletion between regions for under-five children in Ethiopia.

1.4 Significance of the Study

The World Health Organization defines fully vaccinated children as having received three doses of the DPT and polio vaccines, a measles vaccination by the age of 12 months, and a vaccination against tuberculosis (BCG). Considering this EPI, the study may help stakeholders to achieve the target goals regarding children's vaccination. Additionally, it fills a research gap as little is known about Ethiopia's research on explaining variability rather than identifying risk factors of vaccination incompleteness in under-five year’s old children. The percentage of children vaccinated currently is one indicator of how much a nation has contributed to the Sustainable Development Goals (SDG). This study will be significant because it will improve our knowledge of the factors that influence children's vaccination incompletion.

Furthermore, by developing relevant, evidence-based suggestions for addressing problems associated with vaccination incompletion rate, the study will support the efficient use of resources. Moreover, the factors influencing the complete vaccination of under-five children in our nation are not frequently analyzed using statistical techniques like multilevel ordinal logistics regression analysis. So that policy makers in government and non-governmental organizations can use this study as a source of information and to raise knowledge of the proper statistical techniques to apply when dealing with correlated data. Ultimately, the research will serve as foundation for future investigations.

1.5 Scope of the Study

This study is about “assessment of vaccination incompletion and associated factors influencing vaccination of children of aged 12 to 59 months in Ethiopia using 2019 EMDHS Data by applying a multilevel ordinal logistics regression analysis”. Then the study scope is limited only in Ethiopia to investigate the relationships between the key socio-economic, demographic and health related predictors on children’s vaccination incompletion.

1.6 Organization of the Paper

The study is organized into five chapters. Chapter one covers background of the study, introduction, statement of the problem, general and specific objectives, scope and significance of the study. The second chapter presents a brief review of both theoretical and empirical literatures, and the third chapter deals with methodology including study design, type of variables, source of data and methodology of statistical analysis, the fourth chapter deals with results of statistical data analysis, finally the last chapter deals with discussion, conclusion, recommendations and limitation of the study.

CHAPTER TWO

2. LITERATURE REVIEW

In this chapter, the researcher presented a review of the literature on factors associated with childhood vaccination incompletion of children aged 12 to 35 months. Relevant studies were reviewed giving for special focus on findings and discussion in the study. And also, theoretical and Empirical literature reviews on childhood vaccination incompletion for under-five children were given below.

2.1. Theoretical literature review

2.1.1. Concepts and backgrounds of vaccination status and coverage

Vaccination is a tried-and-true method of preventing two to three million deaths per year and managing and curing serious vaccine-preventable diseases. Vaccination is defined as a type of health technology that is essential to the practice of pediatric medicine that aims to prevent, reduce, or shield a person from an epidemic (Dessie & Negeri, 2018). Expanded program of immunizations (EPIs) were introduced by the World Health Organization (WHO) to all of its members in 1974 with the goal of eradicating vaccine-preventable diseases (VPDs), including tuberculosis, tetanus, measles, pertussis, polio virus, and diphtheria. The process of establishing immunity by vaccination in children is known as childhood vaccination. In high-endemic settings, it is acknowledged as one of the most economical public health strategies to prevent infectious disease-related morbidity and mortality (WHO, 2012).

According to a recent WHO report, 12 million unvaccinated children lived in just ten countries: Angola, Brazil, the Democratic Republic of the Congo, Ethiopia, India, Indonesia, Nigeria, Pakistan, the Philippines, and Vietnam. Additionally, 19.4 million infants had not received the third dose of the DPT vaccine. Under-five deaths worldwide varied from 96 to 41 per 1000 live births between 1990 and 2018. Sub-Saharan Africa (SSA) saw a range of 189 to 83/1000 live births in the same years. Ethiopia has a higher under-five mortality rate in 2018 than most other SSA nations, at 61 per 1000 live births (WHO, 2015).

One of the world's most economical public health initiatives to lower morbidity and mortality from infectious illnesses in children is the Expanded Program on Immunization (EPI). It serves as a fundamental component of health systems, delaying nearly two million fatalities annually on a global scale (Waisbord & Larson, 2015). According to estimates from the World Health Organization, vaccines could avert an extra two million deaths annually among children under five if all currently available vaccines against childhood diseases are widely adopted and immunization programs can raise vaccine coverage to an average of 90% globally (United Nations, 2006). In order to reduce under-five mortality by two thirds in 2015, increasing vaccination coverage is crucial. Vaccination coverage is also one of the measures used to track progress toward MDG4 achievement (UN, 2013).

While immunization is one of the most successful public health interventions, coverage plateaued in the decade prior to COVID-19. The COVID-19 pandemic, associated disruptions and vaccination efforts strained health systems in 2020 and 2021, resulting in dramatic setbacks. However, from a global perspective recovery is on the horizon: in 2022 diphtheria, pertussis, and tetanus (DTP) immunization coverage almost recovered to 2019 levels. During 2022, about 84% of infants worldwide (110 million) received 3 doses of diphtheria-tetanus-pertussis (DTP3) vaccine, protecting them against infectious diseases that can cause serious illness and disability or be fatal. These global figures however hide significant disparity become countries of different income strata, with low-income countries lagging behind (WHO, 2022).

Haemophilus influenzae type B (HiB) causes meningitis and pneumonia. Hib vaccine had been introduced in 193 Member States by the end of 2022. Global coverage with 3 doses of Hib vaccine is estimated at 76%. There is great variation between regions. The WHO European Region and South-East Asia Region are estimated to have 93% coverage and 91% coverage, respectively, while it is only 32% in the WHO Western Pacific Region. Hepatitis B is a viral infection that attacks the liver. Hepatitis B vaccine for infants had been introduced nationwide in 190 Member States by the end of 2022. Global coverage with 3 doses of hepatitis B vaccine is estimated at 84%. In addition, 113 Member States introduced nationwide 1 dose of hepatitis B vaccine to newborns within the first 24 hours of life. Global coverage is 45% and is as high as 80% in the WHO Western Pacific Region, while it is only estimated to be at 18% in the WHO African Region (WHO, 2022).

Human papillomavirus (HPV) is the most common viral infection of the reproductive tract and can cause cervical cancer in women, other types of cancer, and genital warts in both men and women. One hundred thirty Member States had the HPV vaccine in their national immunization services by the end of 2022, including 14 new introductions. Since many large countries have not yet introduced the vaccine and vaccine coverage continues to be suboptimal in 2022 in many countries, global coverage with the first dose of HPV among girls is now estimated at 21%. This is a proportionally large increase from 16% in 2021 and was driven in particular by the effect of new introductions and programs that resumed after interruptions. Bacterial meningitis is an infection that is often deadly and leaves 1 in 5 individuals with long-term devastating sequelae after the acute infection. Before the introduction of MenAfriVac in 2010(a revolutionary vaccine) Neisseria meningitidis serogroup A (NmA) accounted for 80–85% of meningitis epidemics in the African meningitis belt. By the end of 2022, more than 350 million people in 24 out of the 26 countries in the meningitis belt had been vaccinated with MenAfriVac through campaigns and 14 countries had included MenAfriVac in their routine immunization schedule. No case of NmA meningitis has been confirmed since 2017 in the meningitis belt.

Measles is a highly contagious disease caused by a virus, which usually results in a high fever and rash, and can lead to blindness, encephalitis or death. By the end of 2022, 83% of children had received 1 dose of measles-containing vaccine by their second birthday, and 188 Member States had included a second dose as part of routine immunization and 74% of children received 2 doses of measles vaccine according to national immunization schedules. Mumps is a highly contagious virus that causes painful swelling at the side of the face under the ears (the parotid glands), fever, headache and muscle aches. It can lead to viral meningitis. Mumps vaccine had been introduced nationwide in 123 Member States by the end of 2022. Pneumococcal diseases include pneumonia, meningitis and febrile bacteraemia, as well as otitis media, sinusitis and bronchitis. Pneumococcal vaccine had been introduced in 155 Member States by the end of 2022 and global third dose coverage was estimated at 60%. There is great variation between regions. The WHO European Region is estimated to have 83% coverage, while it is only 23% in the WHO Western Pacific Region (WHO, 2022).

Polio is a highly infectious viral disease that can cause irreversible paralysis. In 2022, 84% of infants around the world received 3 doses of polio vaccine. In 2022, the coverage of infants receiving their first dose of inactivated polio vaccine (IPV) in countries that are still using oral polio vaccine (OPV) is estimated at 84% as well. Targeted for global eradication, polio has been stopped in all countries except for Afghanistan and Pakistan. Until poliovirus transmission is interrupted in these countries, all countries remain at risk of importation of polio, especially vulnerable countries with weak public health and immunization services and travel or trade links to endemic countries (WHO, 2022).

Rotaviruses are the most common cause of severe diarrhoeal disease in young children throughout the world. Rotavirus vaccine was introduced in 121 countries by the end of 2022. Global coverage was estimated at 51%. Rubella is a viral disease which is usually mild in children, but infection during early pregnancy may cause fetal death or congenital rubella syndrome, which can lead to defects of the brain, heart, eyes and ears. Rubella vaccine was introduced nationwide in 174 Member States by the end of 2022, and global coverage was estimated at 68%. Tetanus is caused by a bacterium which grows in the absence of oxygen, for example in dirty wounds or the umbilical cord if it is not kept clean. The spores of C. tetani are present in the environment irrespective of geographical location. It produces a toxin which can cause serious complications or death. Maternal and neonatal tetanus persist as public health problems in 12 countries, mainly in Africa and Asia. Yellow fever is an acute viral haemorrhagic disease transmitted by infected mosquitoes. As of 2022, yellow fever vaccine had been introduced in routine infant immunization programs in 37 of the 40 countries and territories at risk for yellow fever in Africa and the Americas. In these 40 countries and territories, coverage is estimated at 45% (WHO, 2022).

The COVID-19 pandemic drastically interrupted the health system and routine immunization in the African Region in 2020 and 2021, which resulted in a notable drop in vaccine coverage. It is commonly known that stable communities and operational healthcare facilities are necessary for regular immunization programs. (UHC/UCN/WHO, 2022).

Statistics from 2011 alone show that there were over seven million under-five deaths worldwide, of which approximately 41 percent happened in sub-Saharan Africa. The vast majority of these deaths might have been prevented if the region had comprehensive immunization coverage (Rutherford et al, 2010).

According to WHO regional office in Africa (2018) report revealed that there has been considerable success with the development and introduction of new vaccines in the Region. However, uptake of these vaccines has not matched the level of success in new vaccine introduction. This has made the goal of reaching high and equitable immunization coverage a mirage in the Region. Multiple barriers have been blamed for this, chief among which are inadequate commitment of national governments and weak community engagement to immunization service delivery in the Region. Steps are taken to address these issues, including sensitization of government of the African Region to prioritize Universal Access to Immunization as a Cornerstone for Health and Development in Africa. This is because it is argued that development efforts are link to the human beings for whom progress is targeted and/or agents that bring about development.

2.1.2. The EPI program and vaccination coverage in Ethiopia

One indicator used to track progress toward millennium development goal four (MDG4) achievements is the coverage of immunizations. In Ethiopia, the childhood immunization rate is not very high. Universal immunization of children against six common vaccine-preventable diseases, namely tuberculosis, diphtheria, whooping cough (pertussis), tetanus, polio, and measles, is crucial in reducing infant and child mortality. Other childhood vaccines given in Ethiopia protect against hepatitis B, and Haemophilus influenzae type b (Hib). The government of Ethiopia introduced the pneumococcal conjugate vaccine (PCV 13) and monovalent human rotavirus vaccine (RV1) into the nation’s infant immunization programme in November 2011 and October 2012, respectively. The pneumococcal vaccine protects against Streptococcus pneumonia bacteria, which cause severe pneumonia, meningitis, and other illnesses. Rotavirus causes gastroenteritis, an inflammation of the stomach and intestines. If left untreated, it can lead to severe dehydration and death. Unlike the previous EDHS surveys, 2019 EMDHS has also captured information related to the second dose of measles vaccine (MCV 2), an effort recently launched in early 2019 (EMDHS, 2019).

According to the guidelines developed by WHO, children are considered to have received all basic vaccinations when they have received a vaccination against tuberculosis (also known as BCG), three doses each of the DPT-HepB-Hib (also called pentavalent) vaccine, vaccines against polio, and a vaccination against measles. The BCG vaccine is usually given at birth or at first clinical contact, while the DPT-HepB-Hib and polio vaccines are given at approximately age 6, 10, and 14 weeks. Measles vaccinations should be given at or soon after age 9 months. The Ethiopia immunization programme considers a child to have all basic vaccinations if the child has received three doses of the pneumococcal conjugate vaccine (PCV, also given at age 6, 10, and 14 weeks), and two doses of the rotavirus vaccine (at age 6 and 10 weeks) (WHO, 2022).

According to 2019 EMDHS report showed that data on vaccination coverage among children age 12-23 months who received specific vaccines at any time before the survey (according to a vaccination card or the mother's recall), showed that only 4 out of 10 children (43%) have received all basic vaccinations. Close to 2 in 10 children (19%) in this age group have not received any vaccinations at all. 73% of children received BCG, 76% received the first dose of pentavalent, 78% received the first dose of polio, 74% received the first dose of PCV, and 73% received the first dose of rotavirus vaccine. Fifty-nine percent of children received a measles vaccination (MCV1). Coverage rates decline for subsequent doses of these vaccines, with 61% of children receiving the recommended three doses of the pentavalent, 60% all three doses of polio, 60% all three doses of PCV, and 67% the two doses of the rotavirus vaccine. Nine percent of children age 24-35 months received the second dose of the measles vaccine (MCV2) (EMDHS, 2019).

The Ethiopia EPI was started in 1980 with an intention of reaching 100% coverage by 1990 (WHO: Expanded program on immunization). The EPI program in Ethiopia is administered by the Ministry of Health with technical support from the WHO and other organizations. International partners provide extra support in expanding coverage of EPI. For instance, the Reaching Every District (RED) approach is collaboration with the WHO, USAID, UNICEF, Global Alliance for Vaccines and Immunization (GAVI), and Centers for Disease Control (CDC) (Kidane et al, 2008). The RED approach consists of strengthening social mobilization activities, and developing culturally appropriate behavioral change communication strategies. Vaccine-preventable illness outbreaks have been primarily caused by incomplete immunization on multiple occasions. Ethiopian children who receive a full 12-month childhood vaccination are protected against measles, poliomyelitis, tetanus, diphtheria, and TB (Kidane et al, 2008).

2.1.3. Factors Associated with vaccination incompletion of children

According to a review of the literature, vaccination defaulting happens in a variety of scenarios, and factors such as parent's/caregivers, socio-economic and demographics of the population as a whole, cultural norms, and service accessibility affect the uptake of childhood vaccinations.

2.1.3.1. Geography and governance

A child's immunization process is influenced by their lifestyle, geography, and demographic variety. Ethiopia is governed by a federal structure. Linguistic and cultural disparities separate Ethiopia's two established municipalities and nine regional states. The Southern Nationals, Nationalities and Peoples Regional State (SNNP), Afar, Benishangul-Gumuz, Gambella, Harari, Oromia, Somali, and Tigray are the nine regional states. The two administrative cities are Dire Dawa and the capital, Addis Ababa. This system was established in the 1994 constitution and was instituted as a means for providing the right of self-governance to identity groups, thus furthering democratic governance (United Nations Statistics Division). Religion, food, livelihood and other cultural practices vary from region to region. There are monsoon seasons in Ethiopia. This makes unsafe roads even riskier, which affects parents' ability to get to immunization sites as well as the movement of vaccines from the capital to outlying areas. The Afar and Somali regions are home to seasonal pastoralists. Reaching nomads requires a different approach than addressing sedentary populations.

2.1.3.2. Demographic and socio – economic characteristics of women

Parents' various economic traits have an impact on the underlying propensity to neglect childhood vaccinations. In the 2011 Human Development Report, Ethiopia was placed 174th out of 187 countries based on a variety of social and economic aspects that impact possibilities for success in life (Klugman, J, 2011). The Human Development Report provides information on Ethiopia's position in the international arena as well as the country's economic health and standard of life, all of which may have an impact on vaccination policies for children. From a financial perspective, a household's cost-benefit analysis may discourage parents from demanding that their children receive vaccinations. According to (Byström, E et al, 2020) parents and caregivers may find that the current cost of vaccinations outweighs the potential expense of future sickness. Immunization defaulting has been found to be significantly predicted by family income. Moreover, vaccination uptake is influenced by education, which is correlated with money and other forms of social capital. Maternal education and child immunization are substantially correlated.

The place of delivery, maternal health services such as prenatal care, and mother TT vaccinations are all related to the immunization status of the offspring. Widespread disparities in immunization coverage continue to harm parents in the lowest socioeconomic quintile, especially those who live in rural regions and have no formal education, which has an impact on a child's immunization status. The place of delivery and accessibility to medical facilities are additional characteristics that are linked to a child's immunization status (Adebiyi, F, 2013). According to two studies, since mothers will be exposed to health information, children who are born in a medical institution have an increased probability of receiving all recommended vaccinations (Mutua, M et al, 2011).

2.1.3.3. Access

Structural access to health facilities influences parents' self-efficacy regarding their children's immunizations. Health Extension Workers (HEW) equipped kebeles (districts) had greater coverage rates, according to a national study of Ethiopian defaulters. Moreover, Kebeles with vaccination locations within an hour's walk showed higher coverage than Kebeles with a greater walking distance. Mothers who are eligible for postnatal continuity of care via a medical facility may choose to vaccinate their kid. Mothers who received treatment following delivery were more likely to finish their immunization program (Tadesse, et al, 2009).

2.4. Empirical Literature Review on Vaccination

Detailed empirical literature review was presented below focuses on the socio-economic, demographic and health service related factors of childhood vaccination incompletion.

Coverage of all basic vaccines and/or any vaccination coverage has been strongly associated with better wealth status, better education of care givers, and living in urban areas. 57% of children living in urban areas have received all basic vaccinations compared with only 37% of children in rural areas. Children in the highest wealth quintile (65%) are more than twice as likely to have received all basic vaccinations as children in the lowest quintile (25%). Sixty-five percent of children whose mothers have more than secondary education been received all basic vaccinations compared with 34% of children whose mothers have no education. Coverage of all basic vaccinations is highest in Addis Ababa (83%) and lowest in Afar (20%) (EMDHS, 2019).

According to (Tadesse et al, 2009) the result showed that Four hundred eighteen (41.7%) of the children were fully vaccinated and four hundred twelve (41.2%) of the children were partially vaccinated. The BCG: measles defaulter rate was 76.2%. Monthly family income, postponing child immunization and perceived health institution support were the best predictors of defaulting from completion of child immunization. Another study conducted in south-east Nigeria, a cross-sectional and a cluster sampling design were implemented; the result revealed that 180 (48.1%) females, and 194 (51.9%) males' children were immunized; Less than half 155 (41.9%) of the children had 1 missed dose, considered as partial immunization cases indicating low coverage. Of the reasons given for incomplete immunization, mothers agreed that immunization centers are far from home (Umoke et al, 2021).

According to a study by (Fenta S, et al, 2021) using 2016 EDHS data a multilevel Proportional Odds Model was used the result revealed that among 1, 929 children, only 48.6% (95% CI: 46.3 to 50.8%) were fully vaccinated while 37.8% (95% CI: 35.7 to 40.1%) were partially vaccinated. The multilevel ordinal logistic regression model revealed that housewife mother (AOR =1.522, 95%CI: 1.139, 2.034), institutional delivery (AOR =2.345, 95%CI: 1.766, 3.114),four or above antenatal care visits (AOR = 2.657; 95% CI: 1.906, 3.704), children of mothers with secondary or higher education (AOR = 2.008; 95% CI: 1.209, 3.334),Children whose fathers primary education (AOR = 1.596; 95% CI: 1.215, 2.096), from the rich households (AOR = 1.679; 95% CI: 1.233, 2.287) were significantly associated with childhood vaccination.

A study in Ghana using binary logistics model the study found that being divorced (p = 0.048) and working part-time (p = 0.049) has a significant and positive association with immunization incompletion. Women who were divorced [AOR (95% CI) 3.01 (1.59–58.2)] were 3 times less likely to complete immunization than those who were cohabiting, married and widowed taken into account the effect due to all the additional confounder variables included in the analysis. Women who were working part-time were 2.28 times less likely to complete immunization schedule than those working full-time; [AOR (95% CI) 2.28 (1.031–9.11)] (Anokye et al, 2018).

Another study conducted in Ghana using binary logistics regression the result showed that in total, 89.5% (537/600) of the children were fully immunized, 9.5% partially immunized and 1.0% received no vaccine. In the multivariate analysis, the following determinants were significantly associated with the likelihood of being not fully vaccinated (Odds Ratio larger than 1) : age of the mother/caregiver 40–49 years (AOR = 0.15, 95%CI = 0.05–0.87) compared to less than 20 years; marital status (compared to never married/single: being married AOR = 0.29, 95%CI = 0.13–0.68), ethnicity (compared to the main ethnic group Akan: Frafra (AOR = 4.71, 95%CI = 146– 15.18) and Kusaasi (AOR = 0.09, 95%CI = 0.02–0.51), religion (compared to Islam: Christianity AOR = 0.17, 95%CI = 0.06–0. 50), sex of child (compared to male: female AOR = 0.39, 95%CI = 0.19–0.80) and possession of immunization card (compared to those having the card: those without the card AOR = 84.43, 95%CI = 17.04–418.33). We found no association between immunization status and child's relationship to respondent, parity, education, occupation and child's age (Adokiya et al, 2017).

A study conducted in Ethiopia using binary logistics regression the result revealed knowledge about the benefits of vaccination (AOR=6.1 (95% CI 1.3, 28.9), the age at which the child begins vaccination (AOR=2.4 (95% CI 1.09, 8.4) time taken to reach nearby health facility and means of transportation to nearby health facility (AOR=0.22 95% CI 0.06, 0.9) have statistically significant association with incomplete vaccination (Yismaw et al, 2019). Another study done in Mizan-Aman Town, Bench-Maji Zone, Southwest Ethiopia depicted that mothers/caretakers educational level, fathers' educational level, place of delivery, maternal health care utilization, and mothers/caretakers knowledge about vaccine and vaccine-preventable disease showed significant association with full child immunization. The finding from this study revealed that child immunization coverage in the studied area was low (Meleko et al , 2017).

A study by (Alemu, K, 2014) using 2011EDHS applying multilevel binary logistics regression This study identified that region, wealth index, sex of household head, birth order of a child, frequency of a mother listening to radio, and mothers' getting of prenatal care and assistant from a health professional were statistically significant at 5% level for a child 12 to 59 month of age being partially immunized. The multilevel logistic regression analysis revealed that there was heterogeneity across regions. Between regional variations for vaccination incompletion of children 12 to 59 months of age was 1.0596 with standard error 0.4603. Intra-class correlation between two individual mothers in the same region was 0.322067. A low ICC indicates relatively small between regional variations, but in this study it was high. Region Tigray, Afar, Oromia, Somali (at 10% significance level), Gambela, Addis Ababa and Dire Dawa were statistically significant regions for the variation of children vaccination defaulting in the country. The chance of a child 12 to 59 month of age being partially immunized was 67% on average in the country. The average probability of a child 12 to 59 months of age being under vaccinated was below the average (0.6661) in Tigray (0.3747), Addis Ababa (0.1773) and Dire Dawa (0.4596) while it was higher in Afar (0.8997),

A spatial modeling conducted by (Melaku et al, 2020) based on 2005, 2011 and 2016 EDHS data shows the proportion of incomplete immunization was 74.6% in 2005, 71.4% in 2011, and 55.1% in 2016. The spatial distribution of incomplete immunization was clustered in all the study periods (2005, 2011, and 2016) with global Moran's I of 0.3629, 1.0700, and 0.8796 respectively. Getis-Ord analysis pointed out high-risk regions for incomplete immunization: Geographically weighted regression identified different significant variables; being not educated and poor wealth index was the two common for incomplete immunization in different parts of the country in all the three surveys.

A study carried out in Ethiopia using multilevel analysis the study employed a community-based cross sectional study design, of the 774 children included for analysis, 498 (64.3%) were fully vaccinated while 247 (31.9%) were fully vaccinated on-time. Caregivers who had secondary education and above (AOR = 2.391; 95% CI: 1.317–4.343), from richest households (AOR = 2.381; 95% CI: 1.502–3.773), children whose mother attended four or more ante natal care visits (AOR = 2.844; 95% CI: 1.310–6.174) and whose mother had two or more post natal care visits (AOR = 2.054; 95%CI: 1.377–3.063) were positively associated with on-time full vaccination. In contrary, caregivers aged above 35 years (AOR = 0.469; 95 % CI: 0.253-0.869], being vaccinated at health post (AOR = 0.144; 95%CI: 0.048– 0.428) and travelling more than 30 min to the vaccination site (AOR = 0.158; 95%CI: 0.033–0.739) were negatively associated with on-time full vaccination. The random effects indicated that 26% of the variability in on-time full vaccination was attributable to differences between communities (Mekonnen ZA, et al, 2020).

A study conducted using 2011EDHS by binary logistics regression the prevalence of fully immunized children was 24.3%. Specific vaccination coverage for three doses of DPT, three doses of polio, measles and BCG were 36.5%, 44.3%, 55.7% and 66.3%, respectively. The multivariable analysis showed that sources of information from vaccination card [AOR 95% CI; 7.7 (5.95-10.06)], received postnatal check-up within two months after birth [AOR 95% CI; 1.8 (1.28-2.56)], women's awareness of community conversation program [AOR 95% CI; 1.9 (1.44-2.49)] and women in the rich wealth index [AOR 95 % CI; 1.4 (1.06-1.94)] were the predictors of full immunization coverage. Women from Afar [AOR 95% CI; 0.07 (0.01-0.68)], Amhara [AOR 95% CI; 0.33 (0.13-0.81)], Oromia [AOR 95% CI; 0.15 (0.06-0.37)], Somali [AOR 95% CI; 0.15 (0.04-0.55)] and SNNPR [AOR 95% CI; 0.35 (0.14-0.87)] were less likely to fully vaccinate their children (Lakew et al , 2015).

A study conducted in Sinana district, Southeast Ethiopia shows more than three fourth (76.8%) of the children aged 12 to 23 months were fully vaccinated by card plus history. Factors significantly associated with full immunization were antenatal care follow up (AOR = 3.7; 95% CI: 2.3, 5.9), being a farmer (AOR = 1.9; 95% CI: 1.1, 3.1), being father with secondary and above educational level (AOR = 3.1; 95% CI: 1.3, 7.4), having household family income greater than 1000 ETB or 52 USD (AOR = 3.2; 95% CI: 1.4, 7.4), those whose average walking time from home to health facilities is less than an hour (AOR = 3.1; 95% CI: 1.5, 6.3), those who had ever discussed about immunization with health extension workers (AOR = 2.4, 95% CI: 1.3, 4.2) and mothers' with sufficient knowledge on immunization (AOR = 2.5; 95% CI: 1.5, 4.2) (Legesse & Dechasa, 2015).

Furthermore, a study conducted using 2019 Ethiopian mini demographic and health survey data. Immunization status was examined in a sample of 1843 children aged 12–24 months and the result shows that the immunization prevalence among children was 72.2% (34.2% fully immunized and 38.0% partially immunized), while about 27.8% of children were non-immunized. The fitted partial proportional odds model revealed that child immunization status was significantly associated with region afar (OR = 7.90; CI: 4.78–11.92), family planning use (OR = 0.69; CI: 0.54–0.88), residence (OR = 2.22; CI: 1.60–3.09), antenatal visit (OR = 0.73; CI: 0.53–0.99), and delivery place (OR = 0.65; CI: 0.50–0.84) (Metkie, K. A. et al, 2023).

CHAPTER THREE

3. METHODOLOGY

3.1. Study Design

Study designs were community based cross-sectional study, and The Ethiopian Public Health Institute (EPHI), in collaboration with the Central Statistical Agency (CSA) and the Federal Ministry of Health (FMoH), conducted the 2019 Ethiopian Mini Demographic and Health Survey, which is a nationally representative cross-sectional survey. Data collection lasted from March to June 2019. The demographic health survey (DHS) collects a wide range of objectives and self-reported data with a strong focus on maternal and child health, nutrition, and mortality. 2019 EMDHS data were used to examine the factors that influence children's vaccination incompletion in Ethiopia that was extracted from DHS web platform after explicit permission sought to use the data. The data sets have been accessed through the web page of the international DHS program after subscription and being an authorized user. EMDHS data were collected from the nine regions and two administrative cities of Ethiopia.

3.2. Source of Data

The data for this study was from the Ethiopian Mini Demographic and Health Survey (EMDHS) conducted from March to June 2019. The sample frame for the 2019 EMDHS is a composite of all census enumeration areas (EAs) created for the 2019 Ethiopian Population and Housing Census (EPHC) by the Central Statistical Agency (CSA). The census frame lists all 149,093 EAs designed for the EPHC 2019.

An enumeration area is a geographical region that encompasses 131 households on average. In two stages, the 2019 EMDHS sample was stratified and selected. There were 21 sampling strata in each region, divided into urban and rural areas. A total of 305 enumeration areas (EAs) were chosen in the first stage, with probability proportional to EA size. The second stage involved systematic sampling in selecting 30 households per EA. vaccination data was gathered from vaccine card records; if a vaccine card was unavailable, mothers were asked to recall their child's vaccination history. If vaccination was not documented on the infant vaccination Card or the health card, the mother was asked to remember if it had been given (EMDHS, 2019). The researcher was used 5,753 children's age of 12 – 35 month old samples from among 23,007 total birth that obtained from 2019 EMDHS data set.

3.3. Data management, processing and analysis methods

The Ethiopian mini demographic and health survey data were taken from the demographic and health survey program website (http://www.dhsprogram.com) after gaining consent from the EDHS program. In general, the data extraction, management, data cleaning, merging and joining of data were done by SPSS 20, descriptive, summary statistics and the inferential part of the analysis like ordinal logistic regression, multilevel ordinal modeling were used STATA 14.

3.4. Inclusion and Exclusion Criteria

The study was included all children aged 12 to 35 months with known vaccination status, while the study was excluded children whose vaccination status was unknown for some reason.

3.5. Study Variables

3.5.1 Dependent Variable:

Based on the WHO recommendation vaccination, the three categories of child vaccination status were fully (completely) vaccinated, partially (incompletely) vaccinated, and not vaccinated. a child who has received one dose of Bacille Calmette-Guerin (BCG), at least three doses of pentavalent, three doses of oral polio vaccine (OPV), and one dose of measles vaccine is considered completely vaccinated, if they are between the ages of 12 and 35 months. A child between the ages of 12 and 35 months who has missed at least one of the eight vaccines is considered partially (incompletely) vaccinated. A child aged 12 to 35 months who has not had any vaccinations is also considered to be non-vaccinated (unvaccinated).

Since the response variable is the vaccination status of a child. Hence we have divided the dependent variable in to three categories. Suppose the response variable denoted by Y_{ij} , for i^{th} child in the j^{th} region.

$$Y_{ij} = \begin{cases} 1 & \text{if a child did not receive any vaccine (Not vaccinated)} \\ 2 & \textit{if a child only received some of the recommended vaccines (partially vaccinated)} \\ 3 & \text{if a child has recieved all of the recommended vaccines } (\textit{completely vaccineted}) \end{cases}$$

3.5.2 Independent variables:

Independent variable were grouped into level-1(women level), and level-2 (community level) variables. Community level variables includes place of residence and region. The researcher was tried to aggregate individual level into community level variable by using aggregation function in SPSS software. For example, community education level was generated by aggregate respondent's education level. However, when aggregate individual level variable to group level variable, we lose within-group information, 80-90% variance could be wasted, and the relationship between aggregated variables was inflated and distorted. The bias results in ecological fallacy. Due to this reason, the researcher were not used the community level variables that obtained by aggregation. Generally, the selection of the independent variables was based on published literature, past expertise and availability on 2019 EMDHS data set.

Table 3. 1: Classification of independent and dependent variables

Variable name	Description and codes
Individual level [child's and parent]	Demographic, Socio-Economic, health and Service Related Characteristics of child and parent that included in Level-1 Hierarchical Order
Age	Age of respondent [15-19='0' (ref.), 20-24='1', 25-29='2', 30-34='3', 35-39='4', 40-44='5', and 45-49='6']
Religion	Respondent's religion [Orthodox='0' (ref.), Catholic='1', Protestant='2' Muslim='3', Traditional/other='4']
Education-Level	Highest educational level [No-education='0'(ref.), Primary='1' Secondary='2', Higher='3']
Place of antenatal care	Place of antenatal care delivered [at home ='0' (ref.), at health facility ='1']
Wealth index	Wealth index of the respondents [Poor = '1'(ref.), middle = '2', riche = '3']
Number of ANC visit	Number of antenatal care visits during pregnancy[No visit='0' (ref.), one to three ='1', four to five='2', six and above = '3']
Sex of household head	

	Sex of husband/household head[Male ='1'(ref.), Female ='2']
Current marital status	Current marital status of a respondent (single ='0' (ref.), married = '1', windowed = '2', divorced = '3')
Assistance professional	Assistance Health professionals of the woman[No ='0' (ref.), Yes ='1']
Number_ Living _ Children	Number of living children [One to Three children='1' (ref.), four to six children ='2' and seven or more children ='3']
Sex of child	Sex of child [Male ='1'(ref.), Female ='2']
Birth order	Birth order of a child among all birth in the family [first birth = '1'(ref.), two to four = '2', five and more ='3']
Place of delivery	Place of delivery of child[at Home ='1' (ref.), Government Health facility = '1', private health facility ='2', NGO Health facility = '3' and other ='4']
Possession of Radio	Household has a radio[No='0' (ref.), Yes ='1']
Possession of TV	Household has television [No='0' (ref.), Yes ='1']
Community Level Variables	Variables that Determine the Communities Level and Included under the Cluster (Community Level, Level-2)
Region	Region [Tigray= '1'(ref), Afar ='2', Amhara= '3', Oromia='4', Somali='5', Benshangul ='6', SNNPR='7', Gambela='8', Harari='9', Addis Ababa='10' and Dire Dawa ='11']
Residence	Type of place of residence [Urban='1'(ref), and Rural='2']
Dependent Variable	Response Variable that Affected by Level-1 and Level-2 Variables
VSTA_Child	Vaccination status of a child [Not vaccinated ='1' (ref.) partially vaccinated='2' and completely vaccinated='3']

Note: ref. indicates the reference of variables that involved in the model after the researcher's selected these variables by using stepwise variable selection methods using STATA 14.2

3.6. Method of Statistical Analysis

There are different statistical techniques that have been used in analysis of collected data, however method of data analysis is depending on the nature of the variables incorporated in the study and the data type of the basic variables. Depending on the above point, the study was used descriptive statistics and inferential statistics. Descriptive measures were used to summarize the characteristics of the study participants using frequencies and percentages for community-level and individual-level variables. Descriptive statistics is a collection, organization, summarization and presentation of data in a meaningful form by using tables, charts, and graphs to describe the frequency distribution, percentage, mean, variance, mode, standard deviation, skewness, kurtosis etc. it uses the data to provide description of the population either through numerical calculations or graphs or tables. Whereas, inferential statistics is the procedure by which we reach conclusion about population based on the information contained in the sample drawn from that population in the question.

3.6.1. Logistics Regression Analysis

Logistic regression is the statistical technique used to predict the relationship between predictors (independent variables) and a predicted variable (the dependent variable) where the dependent variable is dichotomous or polytomous. There must be two or more independent variables, or predictors, for a logistic regression. The independent variables can be continuous (interval/ratio) or categorical (ordinal/nominal). Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name binary logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regressions, the practical use of the procedure is similar. Depending on the number of categories and type of data set for the response variable, logistic regression can be binary logistic regression, multinomial logistic regression, and ordinal logistic regression.

Logistics regression analysis is one of the most popular and widely used analyses that are similar to linear regression analysis except the outcome is dichotomous. Simple logistic regression refers to the regression application with one dichotomous response variable and one independent variable. The general model for binary logistic regression is;

$$\mathbf{logit}(\boldsymbol{\pi}_i) = \mathbf{log}\left(\frac{\boldsymbol{\pi}_i}{\mathbf{1}-\boldsymbol{\pi}_i}\right) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{x}_{1i} + \boldsymbol{\beta}_2\mathbf{x}_{2i} + \boldsymbol{\beta}_3\mathbf{x}_{3i} +, \ldots, +\boldsymbol{\beta}_k\mathbf{x}_{ki} \quad (3.\ 1)$$

Where π_i *is the probablity of success*

The assumption of binary logistic regression model: logistic regression does not assume a linear relationship between the dependent and independent variables, does not require the independent variable be interval and unbound, the dependent variable need not be normally distributed and homoscedasticity for each level of independent and should be categorical(binary)variable, and absence of multi-collinearity. An important interpretation of binary logistic regression model uses the odds ratio. The odds of response are given as $\frac{\pi}{1-\pi}$, hence the ratio of probability of success to probability of failure is odds of success.

3.6.1.1. Ordinal Logistic Regression Analysis

Numerous models have been examined in the literature for the ordinal dependent variable; the fundamental goal of ordered logit models is to calculate the accumulative probability that the dependent variable would exceed the category (Habyarimana, F, et al, 2014). This model, known as the proportional odds model (POM), is based on the fundamental premise that the independent variable's effect is constant across all dependent variable categories. This premise is also known as the parallel lines assumption or proportional odds assumption. To loosen up on these proportionate odds we employed a generalized model known as the partial proportional odds model, which relaxes the proportional odds assumption for some or all of the predictors.

Logistic regression is a popular modeling approach for predicting the value of a categorical dependent variable with one or more independent variables. Depending on the nature of the categories of the response variables, logistic regression models were divided into binary, multinomial, and ordinal models. The ordinal logistic regression model is a type of logistic model used to examine ordinal dependent variables with more than two categories. The most commonly used ordinal logistic regression models are the continuation ratio, adjacent category, partial proportional, and proportional odds models (O'Connell, 2006). The proportional odds model (POM) estimates the probability of being at or below a certain level of the response variable. It considers the likelihood of both this event and all previous events. Other unique ordinal models are used to find significant explanatory variables when the proportionality assumption, which states that the relationship between the independent variables and the dependent variable does not vary with the categories of the dependent variable, is not met (Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X., 2013).

When a dependent variable contains at least three categories, such as vaccination status of a child (not vaccinated, partially vaccinated, and completely vaccinated), ordinal logit models are used to account for the ordinal structure of the dependent variable (Agresti, A, 2010). States that cumulative link models are an effective model class for these types of data because they appropriately handle observations as categorical, take use of their ordered character, and provide a flexible regression framework that enables detailed analysis. For a response variable Y with C categories and a set of predictors X having the effect parameters β the probability of response variable being less than or equal to category j can be modeled by the logistic distribution as;

$$\boldsymbol{Y_j = \mathrm{P}(Y \leq y_j|X), p(Y \leq y_j|X) = \frac{\exp(\alpha_{j\_}(\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots,\beta_j x_{ij})}{1+\exp(\alpha_{j\_}(\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots,\beta_j x_{ij})}} \qquad (3.\ 2)$$

$$with\ j = 1, 2, \dots, c-1$$

The above proportional odds model gives the cumulative probability of category j and for the response variable having categories C we find the c-1 cumulative probabilities as for the last category the cumulative probability is always equal to one. The above model can also be written as;

$$\boldsymbol{\mathrm{P}(Y \leq y_j|X) = \frac{1}{1+\exp(-\alpha_j+(\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_j x_{ij})}} \qquad (3.\ 3)$$

The probability that response variable lies in the categories greater j is

$$\mathrm{P}(Y > y_j|X) = 1 - \mathrm{P}(Y \leq y_j|X) = 1 - \frac{\exp(\alpha_{j\_}(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij})}{1 + \exp(\alpha_{j\_}(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij})}$$

$$\boldsymbol{\mathrm{P}(Y > y_j|X) = 1 - \frac{1}{1+exp(-\alpha_j+(\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_j x_{ij})}} \qquad (3.\ 4)$$

The odds of response variable being less than or equal to category j to category greater than j can be found as;

$$\boldsymbol{\frac{\mathrm{P}(Y \leq y_j|X)}{\mathrm{P}(Y > y_j|X)} = exp(\alpha_{j\_}(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij})} \qquad (3.\ 5)$$

And the logit model is the natural log of odds ratio and is the linear function of k independent variables

$$\boldsymbol{\log\left[\frac{\mathrm{P}(Y \leq y_j|X)}{\mathrm{P}(Y > y_j|X)}\right] = \alpha_j - (\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}),\ \text{with}\ j = 1, 2, \dots, c-1} \qquad (3.\ 6)$$

The proportional odds assumption is the presumption made by the proportional odds model that the explanatory variables have the same effect on the response variable across all response variable categories. The only thing that changes for various responses variable categories is the intercepts, which stay constant under the proportionate odds assumption. They demonstrate how a one unit increase in predictor increases in log odds of falling into the group bigger than j, given that the sign of β_j is negative (subtracted).

The model of cumulative logit is presented below;

$$\mathbf{Y_{ij}} = \frac{\exp(\alpha_j - (\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}))}{1 + \exp(\alpha_j - (\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}))} \qquad (3.\ 7)$$

Where, α_jAre the intercepts and are different for each comparison ordinal categorical variable and relation between to insure that β_j the slope coefficients, are the same for all the categories of dependent variable, for continuous variables the slope coefficients change in log odds for one unit change in predictor and for nominal predictors the slope coefficient represent the effect of each category of nominal variable as compared to reference category.

The Odds Ratio (OR)

The odds ratio is defined as the ratio of the probability of the occurrence of an event to non-occurrence of an event(Wang, 2011). In binary logistic regression analysis, odds ratio is the exponent of the estimated coefficient $exp(\hat{\beta})$ for each continuous covariate let say j, $exp(\widehat{\beta_j})$) is the predicted change in the odds of success for a unit increase in predictor j variable (McHugh, 2009). In case of categorical predictor variable, $exp(\widehat{\beta_j})$is the predicted change in the odds of success for a given category of the predictor variable with respect to the reference category.

Odds Ratio measures the association of vaccination incompletion with individual and community level variables. The odds ratio of a coefficient indicates how the risk of outcome falling in the comparison group compared to the risk of the outcome in the reference group changes with the variables in the question. An odds ratio >1 indicates that the risk of outcome falling in the comparison group relative to the risk of outcome falling in the reference group increases as the variable increases. In other words, the comparison outcome is more likely. Whereas, an odds ratio <1 indicates that the risk of outcome falling in the comparison group relative to the risk of outcome falling in the reference group decreases as the variable increases. In other words, the comparison outcome is less likely.

The ratio of two odds is the conventional definition of OR, calculates how much each independent variable in the model influences the dependent variable's log chances. yet in this instance, odds can instead be described in terms of cumulative probabilities (Williams, D, 2001).

Proportional Odds Model (POM)

The proportional odds model is also called Cumulative Odds Model (POM), which belongs to the class of Generalized Linear Models and is frequently used for the analysis of ordinal categorical data. When there are more than two ordinal categories in the response variable, the binary logistic regression model is generalized. According to (Agresti, A, 2012), it is utilized to calculate the probability of falling within a specific range of the response variable. Each cumulative logit in POM has a unique threshold value. The categories of dependent variables have no bearing on the equality's coefficients. Accordingly, in any cumulative logit model, the coefficients of the independent variable are equal to one another (McCullagh,P & Nelder, J , 1989). Logit can immediately include the ordering when response categories are arranged in a certain way.

This study is focused on estimating the vaccination status of children an ordered categorical response variable Y_{ij} having three ordered categories (Not Vaccinated, Partially Vaccinated, and completely Vaccinated). Ordered logit model estimate the cumulative probability γ_j or cumulative log odds $\log(\frac{\gamma_j}{1-\gamma_j})$ up to j^{th} category where, j = 1, 2, 3, One category (Last) of the response variable is taken as reference category and cumulative probability for reference category is always equal to one. The cumulative logit probability model takes the form as:

$$\mathbf{log}\left(\frac{\gamma_1}{1-\gamma_1}\right) = \boldsymbol{\alpha_1} + \boldsymbol{X'_i\beta_i} \quad \text{(3. 8), and}$$

$$\mathbf{log}\left(\frac{\gamma_2}{1-\gamma_2}\right) = \boldsymbol{\alpha_2} + \boldsymbol{X'_i\beta_i} \quad \text{(3. 9)}$$

Proportional Odds Assumption

Assuming parallel lines at every level of the categorical dependent instead of normality and constant variance makes the ordinal regression model a more advantageous modeling tool. Testing the proportionate odds assumption is crucial before interpreting the model statistics. The regression functions are parallel on the logit scale. The null hypothesis states that the regression functions are parallel on the logit scale. We want to know whether the general model results in a sizeable improvement in fit from the null hypothesis model. The entry labeled chi-square (G) is the difference between the two -2log-likelihood values. And is calculated as;

$G = 2 \times [L(\hat{\beta}) - L(\beta^o)]$. The degree of freedom for this comparison is $[(j+1)-2] \times P]$ and the approximate p-value is $p[X^2((j+1)-2] \times P)] \geq G$ the related slope coefficients did not differ significantly among response categories, according to the significance p value, indicating that the model's assumption of parallel lines had not been violated in the model involving the link function. $L(\hat{\beta})$. Is the log-likelihood function for the model with the estimated P parameters, and $L(\beta^o)$ is the log-likelihood, without any predictor, using just the j thresholds.

Testing the assumption of parallel lines/proportional odds assumption

An essential premise of ordinal odds is present in ordinal logistic regression models. This presumption states that parameters shouldn't vary for various groups. The test of parallel lines is used to evaluate whether it is acceptable to presume that the location parameters' (slope coefficients) values are the same for all answer categories.

The assumption behind fitting an ordinal logistic regression with the proportional odds model is that all $logits$ have the same relationship with independent variables. This indicates that there is a test of parallel lines or planes for each category of the response outcome in this set of findings. According to (Xu et al., 2022), there are two log-likelihood functions for the test of parallel lines or planes: -2log-likelihood for the model that believes the lines or planes are parallel and -2log-likelihood for the model that assumes the lines or the planes are separated.

There are a few choices available if the proportional odds model is not met. These include: - Collapse two or more levels, especially if there are few observations in some of the levels. Multinomial logistic regression, partial proportional odds model, and bivariate ordinal logistic analyses are used to determine whether a specific independent variable is acting differently at various levels of the dependent (Agresti, A, 2010).

3.6.1.2. The generalized ordinal logistic regression (GOLOGIT) model

As proposed by (Clogg & Shihadeh, 1994) and (Williams, 2006) generalized ordered logit model is an ordinal logistic regression which considers order of category of the response variable with k set of explanatory variables. This model results $J-1$ logits without constrained the effect of each explanatory variable is equal across the logits. Generalized ordered logit model estimates the regression parameters for each explanatory variable on $J-1$ logit of the probability being beyond the J^{th}category in every logit to have different estimated values. The generalized ordered logit model (GOLM) is used when the proportional odds assumption is fully or partially relaxed for the explanatory variables. At the same time, the partial proportional odds model (PPOM) is used when the proportional odds assumption is satisfied for some but not all explanatory variables. The continuation ratio logistic model (CRM) contrasts the likelihood of responding to a particular category with the probability of responding to a higher response, when developing a logit for adjacent categories, the categorization of the response variable is taken into account, and logits are calculated for each pair of categories (Agresti, A, 2019).

In the case where the proportional odds assumption is violated, the proportionality constraint may be completely or partially relaxed for the set of explanatory variables. In this situation we apply the generalized ordered logit (GOLOGIT) model and can be expressed as;

$$\boldsymbol{logit}\left(\boldsymbol{p}\left(\boldsymbol{Y}>\boldsymbol{y_j}\middle|\boldsymbol{X}\right)\right)=\mathbf{log}\left(\frac{\boldsymbol{p}(\boldsymbol{Y}>\boldsymbol{y_j}|\boldsymbol{X})}{\boldsymbol{p}(\boldsymbol{Y}\leq\boldsymbol{y_j}|\boldsymbol{X})}\right)=\boldsymbol{\alpha_j}+(\boldsymbol{\beta_1}\,\boldsymbol{x_{i1}}+\boldsymbol{\beta_2}\,\boldsymbol{x_{i2}}+\cdots+\boldsymbol{\beta_j}\boldsymbol{x_{ij}},\boldsymbol{j}=\mathbf{1},\ldots,\boldsymbol{J} \quad (3.\ 10)$$

Where, α_j *is the intercept or cut points*, β_j *is the logit coefficients*

Positive logit coefficient indicates that an individual is more likely to be in a higher category as opposed to a lower category of the outcome variable. Major strength of gologit is that it can fit three special cases of the generalized model: the proportional odds/parallel-lines model, the partial proportional odds model, and the logistic regression model. Hence, gologit can fit models that are less restrictive than the parallel-lines models fitted by ologit (whose assumptions are often violated) but more parsimonious and interpretable than those fitted by a non-ordinal method, such as multinomial logistic regression (i.e., mlogit). Some well-known models are special cases of the gologit model. When $J=2$, the gologit model is equivalent to the logistic regression model. When$J>2$, the gologit model becomes equivalent to a series of binary logistic regressions where categories of the dependent variable are combined; in this study we have$J=3$, then for $j=1$ category 1 is contrasted with categories 2 and 3; for $j=2$ the contrast is between categories 1 and 2 versus 3.

3.6.1.3. Variable Selection and Model Building Techniques

The number of variables to be included in the model should be the minimum possible that is parsimonious and deliver optimum information. In this study the variable selection process begins with a univariate analysis of each variable. Tests to determine whether a systematic relation or association between each predictor variable with the response variable exists are made before the final model was selected. A univariate logistic regression and a likelihood ratio (LR) chi-square test would be employed to examine the importance of each predictor variables to the outcome variable (Hosmer,D and Lemeshow, 2003).

Another approach to variable selection is to use stepwise selection procedure. Stepwise selection of variables has been widely used in linear regression. In this method, variables are selected for either inclusion or exclusion from the logistic regression model in a sequential fashion based on statistical criterion that checks for the importance of variables. The importance of variables is defined in terms of a measure of the statistical significance of the coefficient for the variable. In stepwise selection procedure, backward selection and/or forward selection procedure are used (Hosmer,D and Lemeshow, 2003). In Forward selection procedure, we add terms sequentially until further additions do not improve the fit. The backward selection on the other hand begins with a complex model and sequentially removes terms. Stepwise selection procedure is the combination of forward selection and backward selection to identify the best model (Hosmer,D and Lemeshow, 2003).

3.6.1.4. Parameter Estimation in Ordinal Logistic Regression Model

The Maximum Likelihood (ML) method is the most often used technique for estimating a logistic regression model's parameters. The likelihood equations in logistic regression are explicit, non-linear functions of the unknown parameters. As a result, we solve the equations using the well-known and highly efficient Newton-Raphson iterative approach, also referred to as the iteratively reweighted Lewis squares algorithm (Ngunyi, A.et al , 2014)

Therefore, the maximum likelihood estimation technique is applied in this study to estimate the model's parameters. The maximum likelihood estimate method, in general, yields values of the unknown parameters that most closely match the observed and projected probability values. A fisher scoring system was presented by (Widyaningsih, P et al, 2017) for MLE fitting of all cumulative link models. As a result, it's a popular and highly useful technique for obtaining maximum likelihood estimates for ordinal logistic regression parameters.

3.6.1.5. Goodness-of-Fit of the Ordinal logit model

As in linear regression, goodness of fit in logistic regression attempts to get at how well a model fits the data. It is usually applied after a final model‖ has been selected (Schermelleh-Engel,et al , 2003). Much of the goodness of fit literature is based on the following hypothesis:

H_0: The model fit the data well Vs. **H_1**: The model does not fit the data well.

The measure of goodness of a fit is done by testing whether a model fits is to compare observed and expected values. From the observed and expected frequencies, we can compute the usual Pearson and Deviance goodness-of-fit measures. For a sample of n independent observations, the deviance and Pearson chi-square for a model with p degrees of freedom, both and D has chi-square distribution with (n-p) degrees of freedom.

A) Test of Overall Model Fit

Likelihood-Ratio Test

The overall model fit in ordinal logistic regression can be based on the change in minus 2loglikelihood when the variables are added to a model that contains only the intercept.

The likelihood-ratio test statistic is given by (Agresti, A., 2010).

$$\boldsymbol{G^2 = -2\ln\left(\frac{L_0}{L_1}\right) = -2[InL_0 - InL_1]} \qquad (3.\ 11)$$

Where; L_0 is the likelihood function of the reduced model and L_1 is the likelihood function of the full model evaluated at the MLEs. This natural log transformation of the likelihood functions yields an asymptotically chi-squared statistics. G^2 is distributed with degrees of freedom equal to the difference between the numbers of parameters estimated in the two models (Menard, 2002). It is important to test the null hypothesis that all population logistic regressions coefficients are not significance difference except the constant one.

Pseudo R-squared statistics: In logistic regression model, McFadden's pseudo R-squared statistic used to compute based on the log likelihood for the model with predictors compared to the log likelihood for the model without predictors (Smith, T & McKenna, C. M, 2013).

B) Individual predictors test of statistical significance in ordinal logit model

The Wald statistic is an alternative test which is commonly used to test the significance of individual logistic regression coefficients for each independent variable. The hypothesis to be tested is, $H_0: \beta_j = 0$ Vs $H_1: \beta_j \neq 0$, j=1, 2…,p at α level of significance. This test is based on descriptive measures of the log-likelihood function at the maximum likelihood estimate $\hat{\beta}_j$ the parameters of the model. The Wald test statistic Z^2 is the squared ratio of the logistic coefficient to its standard error.

$$\mathbf{Z^2} = \left[\frac{\hat{\boldsymbol{\beta}}_{\mathbf{j}}}{\mathbf{SE}(\hat{\boldsymbol{\beta}}_{\mathbf{j}})}\right]^{\mathbf{2}} \approx \mathbf{X^2}(\mathbf{1}) \text{ , } j = 1, 2 \ldots p \qquad (3.\ 12)$$

The Wald test is one of a number of ways of testing whether or not the parameters associated with a group of explanatory variables are zero. If the Wald test is significant for a particular explanatory variable then we would conclude that the parameters associated with these variables are not zero so that the variables should be included in the model otherwise the explanatory variables can be omitted from the model((Agresti, A., 2012).

3.6.2 Multilevel Analysis

Multilevel analysis is a methodology for the analysis of data with complex patterns variability, with a focus on nested sources of such variability. Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. The two basic important uses of multilevel models are: Multilevel model take into account the hierarchical structure usually present in data. They provide a flexible framework for analyzing a variety of different types of response variables and for incorporating covariates at different levels of hierarchical structure.

The main statistical model of multilevel analysis is the hierarchical linear model, an extension of classical linear regressions to a model that includes nested random coefficients. The term multilevel relates to the level of analysis in public health research, which usually, but not always, consists of individuals (at lower level) who are nested with in spatial units (at higher levels). Multilevel analysis is the analysis of the relationships between variables characterizing individuals and variables characterizing groups. In multilevel analysis, the data structure in the population is hierarchical, and the sampler data a sample from this hierarchical population.

The full multilevel regression model assumes that there is a hierarchical data set with one single dependent variable that is measured at the lowest level and explanatory variables at all existing levels. Conceptually the model can be viewed as a hierarchical system of regression equation.

Assumption of multilevel models

In multilevel analysis, you have to make strong assumptions (Hox, J.et al, 2017).

- That your random effects are normal (or, if you have random slopes as long as random intercepts, that the joint distribution is multivariate normal),
- That your model contains all relevant variables, so that you are safe assuming that errors and repressor’s are uncorrelated at all levels,
- You have enough observations at each level to really utilize the asymptotic theory results concerning the likelihood ratio test statistics and inverse of the information matrix as the estimator of the variances of the parameter estimates.

Multilevel methods consist of statistical procedure and applicable for different reason when; the observations that are being analyzed are correlated or clustered along spatial, non-spatial and sequential dimension, accounting for individual and group level variation in estimating group level estimation, estimating regression coefficients for specific group, there is a basic interest in describing the variability and heterogeneity in the population, over and above the focus on average relationships, and the causal processes are thought to operate simultaneously at more than one level.

Heterogeneous proportion

The basic data structure of two-level logistic regression is a collection of N groups (units at level-2: regions) and within group $j. j = 1, 2, \dots N$ a random sample of n_j level-1 (women’s level) units. The outcome variable is ordinal and denoted byY_{ij} $i = 1,2, \dots n_j$ $j = 1, 2, \dots, N$. For level-1 unit i in group j the total sample size is $= \sum_{j=1}^{N} n_j$. If one does not take explanatory variables into account, the probability of success is assumed constant in each group. Let the success probability in group j be denoted byp_j, and hence it can be estimate for the group-dependent probability p_j as follow; (Snijders and Bosker , 1999). The overall proportion p is given as.

$$\hat{\mathbf{p}}_{\mathbf{j}} = \bar{\mathbf{Y}}_{..} = \frac{\mathbf{1}}{\mathbf{M}} \sum_{\mathbf{j=1}}^{\mathbf{N}} \mathbf{1} \sum_{\mathbf{i=1}}^{\mathbf{n_j}} \mathbf{y_{ij}} \qquad (3.\ 13)$$

Testing Heterogeneous proportion

Testing the heterogeneity of proportions between regions was an obvious first step towards the appropriate application of multilevel analysis. To test whether there are indeed systematic differences between groups, the well-known chi-square test for contingency tables can be used.

The test statistic of the chi-squared test for a contingency table is often given in the familiar form. The chi-square test was most frequently employed to look for proportional heterogeneity between regions. The test statistics is provided by:

$$\mathbf{X^2} = \sum_{\mathbf{i=1}}^{\mathbf{ni}} \frac{(\mathbf{O_i} - \mathbf{E_i})^2}{\mathbf{E_i}^2} \qquad (3.\ 14)$$

Where, O_i is the observed and E_i is the expected count in a cell of contingency table.

Estimation of between and within groups' variance

Consider a population having two-levels, the basic data structure of two-level logistic regression analysis is a collection of N groups (units at level-two i.e. regions) and within group j (j= 1, 2, …,N) a random sample of level-one units (women). The outcome variable is dichotomous and denoted byY_{ij} ,(i= 1,2, …n_j , , j= 1,2, … ,N) for level-one unit i in group j. The total sample size is $M = \sum_{j=1}^{N} n_j$

Then, the (theoretical) true variance between the group (regions) dependent probabilities, i.e., the population value of $varp_j$ can be estimated by;

$$\mathbf{varp_j} = \tilde{\boldsymbol{\tau}}^{\mathbf{2}} = \mathbf{S^2}_{\mathbf{between}} - \frac{\mathbf{S^2}_{\mathbf{within}}}{\tilde{\mathbf{n}}} \qquad (3.\ 15)$$

Where, $\tilde{n}$ is given by;

$$\mathbf{\ddot{n}} = \frac{\mathbf{1}}{\mathbf{N-1}}\left[\mathbf{M} - \frac{\sum_{j=1}^{N} \mathbf{n_j}^2}{\mathbf{M}}\right] = \mathbf{\bar{n}} - \frac{\mathbf{S^2}_{(nj)}}{\mathbf{N\bar{n}}} \qquad (3.\ 16)$$

For dichotomous dependent variable, the observed between-groups variance is closely related to the chi-squared test statistic (Snijders and Bosker , 1999). They are given by the formula:

$$\mathbf{S^2}_{\mathbf{between}} = \frac{\mathbf{\hat{p}(1-\hat{p})}}{\mathbf{\ddot{n}(N-1)}}\boldsymbol{\chi}\mathbf{2} \qquad (3.\ 17)$$

Where, $\chi 2$ is chi-square and the within-group variance in the dichotomous case is a function of the group:

$$\mathbf{S^2}_{\mathbf{within}} = \frac{\mathbf{1}}{\mathbf{M-N}}\sum_{j=1}^{N} \mathbf{n_j p_j}\,(\mathbf{1} - \mathbf{p_j}) \qquad (3.\ 18)$$

3.6.2.1. Multilevel ordinal Logistic Regression Analysis

Multilevel statistical techniques can be used to predict a dichotomous or polytomous dependent variable from a set of independent variables. It can be employed in the simplest case without explanatory variables (usually called the empty model) and also with explanatory variables by allowing only the intercept term or both the intercept and slopes (regression coefficients) to vary randomly, and the coefficients are assumed to follow a multivariate normal. The basic data of two-level logistic regression is a collection of N groups and within group j (j = 1, 2, 3…, N) a random sample of n_j level-1 units (individual level).

Consider the outcome variable Y_{ij} is dichotomous response. There is i^{th} individual level nested in the j^{th} cluster, where i= 1, 2, 3…, n_j indicates individual woman under level-1 with n_j sample size. j = 1, 2, 3… N indicates group of individuals make a community level (Level-2). Then, the total sample size is $M = \sum_{J=1}^{N} n_j$. If one does not take explanatory variables into account, the probability of success is assumed constant each group (Tom, A. et al, 1999).

Suppose k explanatory variables $X_1, X_2, X_3, \dots X_k$ the values of X_h $(h = 1,\ 2, 3 \dots k)$ are indicated in the usual way by $X_{hij}, \big(h = 1, 2, \dots, k\text{: } i = 1, 2, 3, \dots n_j : j = 1, 2, 3, \dots N\big)$. since some or all of these variables could be level -1 variables, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success probability depends on the individuals as well as on the group and is assigned by p_{ij} or π_{ij}. The response variable is expressed as the sum of success probability and a residual term ε_{ij} that is.

$\mathbf{Y_{ij} = P_{ij} + \varepsilon_{ij}}$ (3. 19)

The residuals $\varepsilon_{ij}$are assumed to have mean zero and variance $var(\varepsilon_{ij}) = \delta^2_{\varepsilon} = P_{ij}(1 - P_{ij})$. the logistic regression models with random coefficients express the log-odds. The logit of p_{ij}as a sum of a linear function of the explanatory variables with randomly varying coefficients is.

$$logit\ (p_{ij}) = log(\frac{p_{ij}}{1 - p_{ij}}) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij} +, \dots, +\beta_{kj}x_{kij}$$

where $\beta_{0j} = Y_{00} + U_{0j}$, $and\ \beta_{hj} = \gamma_{h0} + U_{hj}$ $h = 1, 2, 3, \dots k$ Then rewrite the model

$$logit(\ p_{ij}) = Y_{00} + \sum_{h=1}^{k} \gamma_{ho}\ x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj}\ x_{hij}$$

The p_{ij} can be solved as

$$\mathbf{p_{ij} = \frac{exp(\ Y_{00} + \sum_{h=1}^{k} \gamma_{ho}\ x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj}\ x_{hij})}{1 + exp(Y_{00} + \sum_{h=1}^{k} \gamma_{ho}\ x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj}\ x_{hij})}} \quad (3.\ 20)$$

Two level ordinal logistic regression model

The multilevel model for a univariate case with two levels is given by;

$\mathbf{y_{ij} = \beta_0 + \beta_{1j}X_{ij} + u_{0j} + e_{ij}}$ (3. 21)

$j = 1, 2, \dots, N\ and\ i = 1,2 \ldots n_i\ , where, n_i = is\ number\ of\ individual\ in\ each\ group\ group$

Where, $N =\ is\ number\ of\ groups,$

u_{0j} the error at the second level, e_{ij} the error at the first level

y_{ij} is the response variable at i^th^ first level and j^th^ second level

X_{ij} is the explanatory variable at i^th^ first level and j^th^ second level

Model building strategies for 2-level dichotomous response logistic regression, suppose there are p explanatory variables at lowest level (level-1) and q explanatory variables at highest level (level-2 variable), then we were constructed different multilevel regression model.

3.6.2.1.1. The empty two level ordinal logistic regression model

Empty level-2 model for ordinal outcome variable refers to a population of groups: level two units (regions) and specifies the probability distribution for group dependent probabilities without taking explanatory variables into take account. Let the $p_{ij}{}^{(C)}$ be ordered categorical response of i th individual in the j th region with C ordered categories coded as C = 0, 1, 2. . . c.

$$\mathbf{logit}(\mathbf{p_{ij}}{}^{(\mathbf{C})}) = \mathbf{logit}\left(\frac{\mathbf{p_{ij}}{}^{(\mathbf{C})}}{\mathbf{1-p_{ij}}{}^{(\mathbf{C})}}\right) = \boldsymbol{\alpha}^{(\mathbf{C})} + \mathbf{u_{0j}} \qquad (3.\ 22)$$

Which measures the odds of $p_{ij}{}^{(C)}$ being in the category less than or equal to C as compared to greater than category C. Where, $\alpha^{(C)}$ cutoff points (intercept) for ordered categories and u_{0j} is random effect of level-2 and assumed normal distribution$N(0, \delta_{u0}^2)$.

3.6.2.1.2. Random intercept and fixed effect ordinal logistic regression model

In the random intercept model intercept is the only random effect, implies that groups differ with respect to the average value of the response variable (status of vaccination). It represents the heterogeneity between regions in the overall response. Random intercept model expresses the log-odds as a sum of a linear function of explanatory variables and random part of the model.

Random intercept model for ordinal response data is given by;

$$\boldsymbol{logit}(\boldsymbol{p_{ij}}{}^{(\boldsymbol{C})}) = \boldsymbol{logit}\left(\frac{\boldsymbol{p_{ij}}{}^{(\boldsymbol{C})}}{\mathbf{1}-\boldsymbol{p_{ij}}{}^{(\boldsymbol{C})}}\right) = \boldsymbol{\alpha}^{(\boldsymbol{C})} + \boldsymbol{X_{ij}}\boldsymbol{\beta} + \boldsymbol{u_{0j}} \qquad (3.\ 23)$$

These models possess proportional odds property (McCullagh, 1980). For all C the fixed or random effects operate on cumulative odds by a constant multiplicative factor.

As in the single-level model, the thresholds (intercepts) allow response probabilities π_{ijC} to vary across response categories C. Parameter $\alpha^{(C)}{}_{ij}$ is interpreted as the log-odds that an individual with $x = 0$ and $u = 0$ has a response of C or lower. Parameter β is the effect of a one unit changes in x on the log-odds that$p_{ij} \leq C$, after adjusting for or holding constant the group effect$u$.

In random intercept model, addition of group-level residual u_{0j} allows intercepts to vary from region to region, according to a normal distribution. This in turn allows the cumulative response probabilities, $\Pr(p_{ij}{}^{(C)})$ and response probabilities, π_{ijC}, to vary across regions. The between-region variation is due to unobserved group-level influences on y (after accounting for the effects of x) represented by u_{0j}. We estimate $var(u_{0j}) = \delta_{0j}^2$, is the residual between-region variance in the log-odds that $p_{ij} \leq C$.

3.6.2.1.3. Random coefficient two level ordinal Logistic Regression Model

In logistic regression analysis, linear models are constructed for the log odds. The multilevel analogue, the random coefficient logistic regression is based on linear models for the log-odds that include random effects for the groups or other higher level units. The multilevel modeling strategy accommodates the hierarchical nature of the DHS data and corrects the estimated standard errors to allow for clustering of observations within units(Goldstein,2011). A significance random effect may represent factors influencing the outcome variable that cannot be quantified in a large scale social survey. A random effects model thus provides a mechanism for estimating the degree of correlation in the outcome that exists at higher level, while also controlling a range of all indicators may potentially influence the outcome. If there is p lower levels explanatory variable and q higher level variable with random regression coefficient.

$$\text{L}ogit\ p_{ij} = \beta_{0j} + \beta_{pj}X_{pij} + \gamma_{0q}Z_{qj} \text{ By substituting } \beta_{0j} = \gamma_{00} + U_{0j} \text{ and } \beta_{pj} = \gamma_{p0} + U_{pj}$$

$$Logit\ p_{ij} = \gamma_{00} + U_{0j} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + U_{pj}X_{pij}$$

$$\boldsymbol{Logit\ p_{ij}} = [\boldsymbol{\gamma_{00}} + \boldsymbol{\gamma_{p0}}\,\mathbf{X_{pij}} + \boldsymbol{\gamma_{0q}}\,\mathbf{Z_{qj}}] + [\mathbf{U_{pj}}\,\mathbf{X_{pij}} + \mathbf{U_{0j}}] \qquad (3.\ 24)$$

$[\gamma_{00} + \gamma_{p0}\mathrm{X_{pij}} + \gamma_{0q}\mathrm{Z_{qj}}]$ – is fixed part of model and $[U_{pj}X_{pij} + U_{0j}]$ -is random component of this model. The intercept β_{0j} and the slope β_{pj} are group – dependant. these group dependent coefficients can be split into an average coefficient and the group dependent deviation. i.e $\beta_{0j} = \gamma_{00} + U_{0j}$ *and* $\beta_{pj} = \gamma_{p0} + U_{pj}$

There are two random group effects, the random intercept U_{0j} and the random slope U_{pj}. it is assumed that the level two residuals U_{0j} and U_{pj} have means zero given the value the values of the explanatory variable X.

Where, $\gamma_{00-}$is the average regression intercept across the group

γ_{p0-} is the average regression coefficient given fixed explanatory variable X_{pij}.

U_{oj} −Random intercept in the model

$U_{pj}\, X_{pij}$ – is the random intersection between the group and predictors (X).

The above model implies that two random effects characterized the groups (cluster) U_{oj} and U_{pj} are correlated. Moreover, it is assumed that for different groups, the pairs of random (U_{oj}, U_{pj}) effects are independent and identically distributed. Hence, the variance and covariance of the level-two random effects are (U_{oj}, U_{pj}) denoted by:

$$\text{var}\,(U_{0j}) = \delta_{00} = \delta_0^2$$

$$\text{var}(\ U_{pj}) = \delta_{pp} = \delta_p^2$$

$$\text{cov}(\ U_{0j},\ U_{pj}) = \delta_{0p}$$

$U_{0j}, U_{1j}, \dots, U_{pj}$- are assumed to be independent between groups but may be correlated within groups. Therefore the components of the vector $U_{0j}, U_{1j}, \dots, U_{pj}$ are independently distributed as a multivariate normal distribution with zero mean vector, variances and co-variances Ω given by:

$$\Omega = \begin{bmatrix} \delta_0^2 & \delta_{10} & \delta_{20} & . & . & \delta_{p0} \\ \delta_{01} & \delta_1^2 & \delta_{21} & . & . & \delta_{p1`} \\ \delta_{02} & \delta_{12} & \delta_2^2 & . & . & \delta_{p2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \delta_{0p} & \delta_{1p} & \delta_{2p} & . & . & \delta_p^2 \end{bmatrix}$$

In the random intercept model the intercept is the only random effect implies that the groups differ with respect to the average value of the dependent variable. In random intercept model assumed that effects of explanatory variables are the same for each region. Random coefficients model assumed that allowing the difference between explanatory variables within a regions to vary across regions. So, random coefficient model represents heterogeneity in relationship between the response and explanatory variables.

Two-level random slope (coefficient) model is given by;

$$\boldsymbol{logit}\left(\boldsymbol{p_{ij}}^{(C)}\right) = \boldsymbol{logit}\left(\frac{\boldsymbol{p_{ij}}^{(C)}}{\mathbf{1}-\boldsymbol{p_{ij}}^{(C)}}\right) = \boldsymbol{\alpha}^{(C)} + \boldsymbol{X_{ij}}\boldsymbol{\beta} + \boldsymbol{Z_{ij}}\mathbf{u}_{\boldsymbol{j}} + \boldsymbol{u_j} \quad (3.25)$$

Where, $X_{ij} = (1, X_{1ij}, X_{2ij}, \dots, X_{nij})$, $Z_{ij} = (1, Z_{1ij}, Z_{2ij}, \dots, Z_{nij})$, $\beta^T = (\beta_0, \beta_1, \dots, \beta_k)$ and $u_j{}^T = (u_{0j}, u_{1j}, \dots, u_{kj})$. However, X_{ij} are co-variates in the data vector, β^T are a vector of fixed effect coefficients, $u_j{}^T$ are random variables at the level-2 assumed dependent multi-variates normal with expectation zero and $Z_{ij}$are variables at the level-2 which are a subset of the X_{ij}.

3.6.2.1.4. Intra-class correlation

The unconditional model decomposes the variance into two independent components. i.e. level-1 variance (σ^2_ε) and level -2 variance (δ^2_{u0}). The intra-class correlation indicates the proportion of the variance explained by the community structure in the population and simply define that the intra-class correlation is the proportion of community-level variance compared to the total variance. Where, σ^2_ε is the variance of individual level units and σ^2_{u0} is the variance of the community-level residual errors. In multilevel logit model, level-1 residual variance $\sigma^2_\varepsilon = \frac{\pi^2}{3} \approx 3.29$ (Merlo, J et al , 2005)using intercept only model, we define the intra-class correlation (ICC) by the equation;

$$\mathbf{ICC} = \boldsymbol{\rho} = \frac{\boldsymbol{\sigma^2_{u0}}}{\boldsymbol{\sigma^2_{u0}} + \boldsymbol{\sigma^2_\varepsilon}} \quad (3.26)$$

The intra-class correlation is the proportion of the group –level compared to the total variance, and also proportional change in variance (PCV) is estimated by using:

$$\boldsymbol{PCV} = \frac{\boldsymbol{\sigma^2_{uo|b}} - \boldsymbol{\delta^2_{u0|m}}}{\boldsymbol{\sigma^2_{uo|m}}} * \mathbf{100}\% \quad (3.27)$$

3.6.2.1.5. Parameter Estimation (Maximum likelihood via Quadrature)

Estimation of parameters (regression coefficient and variance components) in multilevel modeling is mostly done by maximum likelihood (ML) method. Maximum likelihood estimation is carried out by means of the *gllamm* procedure of Stata (Rabe-Hesketh, S. et al, 2004). The estimation algorithm implemented in *gllamm*, namely Newton-Raphson with adaptive Gaussian quadrature is well established. Maximum likelihood is approximated by either Laplace approximation or quadrature turns out to be sufficient for an accurate estimate. Many alternative estimation methods are possible, e.g. Maximum Simulated Likelihood and Bayesian MCMC (Train, K , 2009). The ML method is a general estimation procedure, which produces estimates for the population parameters that maximize the probability of observing the data given the model. It is the most commonly used estimation method in multilevel modeling by maximizing the likelihood function. Because it is generally robust, and produces estimates that are by maximizing the likelihood function. The individual–level covariance parameters are empirically not identified, as pointed out by the high condition number and the convergence difficulties of the estimation algorithm. Therefore the random errors are omitted from the models. As a consequence, the residual correlation among the responses categories is only due to community-level (cluster) factors.

3.6.2.1.6. Goodness of Fit Test for multilevel ordinal logistic regression model

To compare the ordinal logistics model the log-likelihoods were calculated. A model with a higher log-likelihood should be considered as a better fitting model. It is much better to compare models based on their results, reasonableness, and fit as measured e.g. by the Deviance, Akaike Information Criterion (AIC), and Bayes' Information Criterion. Models with the smaller absolute Deviance, AIC, and BIC values should be preferred.

In this study, we used the deviance value to assess the goodness of fit of the nested model. It is useful to be able to judge whether a model is a good fit to the data. The maximum likelihood procedure produces a statistic called the deviance, which indicates how well the model fits the data. The deviance is obtained by subtracting the two times log-likelihood (-2LL) for the final (full) model from the log- likelihood for the intercept only model (simpler model). This is log likelihood-ratio test that uses the ratio of maximized value of the likelihood function for the intercept only model L_0 over the maximized value of the likelihood function for the full model L_1. The difference between the log-likelihood functions for two models is a measure of how much one model improves the fit over the other. The deviance (the log-likelihood ratio test statistics) is given by:

$$\mathbf{D} = -2\log\left(\frac{\mathrm{L_0}}{\mathrm{L_1}}\right) = -2[\boldsymbol{log}\ (\boldsymbol{Lo})\ -\ \boldsymbol{log}\ (\boldsymbol{L1})]\ =\ -2[\boldsymbol{LL0} - (\boldsymbol{LL1})] \sim \boldsymbol{\chi^2}_{(\boldsymbol{q})} \qquad (3.\ 28)$$

Where: q is the difference in number of parameters between the full model and the simpler model, LL_0 is the log likelihood value of the model which has the intercept term only and LL_1 is the log likelihood value of the full model. The difference of the deviances for two nested models has a chi-square distribution, with degrees of freedom equal to the difference in the number of parameters estimated in the two models. This can be used to perform a formal chi-square test to test whether the more general model fits significantly better than the simpler model i.e. if $D > \chi^2_{(q)}$, then the full model improves significantly over the simpler.

In general, models with a lower deviance fit better than models with a higher deviance. In similarly fashion, the overall model evaluation is also examined using Akaike Information Criteria (AIC) (Akaike 1973) and Schwartz Information Criteria (BIC). The smaller the value, the better of the model was fitted. Akaike Information Criteria (AIC) is appropriate for non-nested model comparison. Akaike's Information Criterion and Bayesian Information Criterion are usually calculated with software. The basic formula is defined as:

$$\mathrm{AIC} = -2(\mathrm{Log}(\text{likelihood of the fitted model})) + 2\mathrm{K} = -2(\mathrm{Log} - \text{likelihood}) + 2\mathrm{K}$$

$$\begin{aligned}\mathrm{BIC} &= -2\big(\mathrm{Log}(\text{likelihood of the fitted model})\big) + \ln(\mathrm{N}) * \mathrm{K} \\ &= -2(\mathrm{Log} - \text{likelihood}) + \ln(\mathrm{N}) * \mathrm{K}\end{aligned}$$

Where: K is the number of model parameters (the number of variables in the model plus the intercept).

Log-likelihood is a measure of model fit. The higher the number results in the better the fit. This is usually obtained from statistical output.

N is the number of observations used in estimation or, more precisely, the number of independent terms in the likelihood.

For small sample size ($\frac{\mathrm{n}}{\mathrm{K}} <\approx 40$) use the second-order AIC

$$\mathrm{AIC} = -2(\mathrm{Log} - \text{likelihood}) + 2\mathrm{K} + \left(\frac{2\mathrm{K}(\mathrm{K}+1)}{\mathrm{n}-\mathrm{K}-1}\right) \qquad \text{Where:}$$

n = sample size, K= number of model parameters, and Log-likelihood is a measure of model fit

3.6.2.1.7. Test of Overall Model Fit

The null hypothesis for an overall model fit test is stated as "all the regression parameters are zero" and the alternative hypothesis is "at least one regression coefficient is not zero". To keep use of the selected model the null hypothesis must be rejected and possibility for examining the significance for the individual parameters.

In testing the goodness of fit of a model, the null and alternative hypotheses can also be stated as "the observed data are consistent with the fitted model" and alternative hypothesis "the observed data are not consistent with the fitted model". This can be tested by the deviance test with q degrees of freedom where q is the difference in number of parameters between the current model and the model with only intercept. The deviance statistic is the difference in -2*log (likelihood) values for the simpler model and fitted model. Then, the null hypothesis is rejected if Deviance (D) > χ^2(q). The logistic regression model can be generalized to handle cases when the outcome variable can take on more than two values and ordered in nature. To assess the fit of the ordinal model, few methods exist. (Lesaffre, E & Albert, A, 1989) have developed logistic regression diagnostics applicable for the ordinal and multinomial model. But they involve extensive calculations that render the method inaccessible to many researchers. Another approach, proposed by (Becg, C. B & Gray, R, 1984), is to replace the ordinal and multinomial model with a series of individual binary logistics regression models. Assessment of fit is performed by calculating any of the standard binary goodness of fit tests, such as the Hosmer-Lemeshow test, for each model and then integrating the results.

Let Y denote an outcome variable with c ordered categories, coded (0, 1, 2…c-1). Assume that the outcome Y = 0 is the reference outcome. Let x be a vector of p independent predictor variables,$x = (x_{1,} x_2, \dots, x_p)$ and we have a sample of n independent observations, (x_i , y_i), i=1, 2… n. Recode y_i into binary indicator variables$\widetilde{y_{ij}}$, such that $\widetilde{y_{ij}} = 1$ when $y_i = j$ and $\widetilde{y_{ij}} = 0$ otherwise$(i = 1, 2, \dots, n \ and \ j = 0, 1, \ \dots, c-1)$. Further, let $\pi_{ij} = P(Y = j/x_i)$.

3.8. Model Diagnostics checking

The researcher can use the estimated residuals to check on the assumptions of the single level model that includes the assumption of linearity, normality, homoscedasticity and independence observation. However, in higher multilevel model against the above particular assumptions when neglecting a nonlinear relationship may result in high estimates of slope variances and cross-level interaction effects. Ordinal logistic regression does have assumptions, such as the assumption of independence among the dependent variable choices. This assumption states that the choice of or membership in one category is not related to the choice or membership of another category of the dependent variable.

One of the main purposes of multilevel model is to deal with the cases where assumption of independence is violated; multilevel models can apply, but assume that: level-1 and level-2 residuals are uncorrelated, and the errors (as measured by residuals) at highest level are uncorrelated. It is of interest to obtain the residual values from the estimated multilevel model. Plots are a good way to examine the residuals. But in multilevel logistic regression, many different residual plots can be used to inspect model assumptions (Hox, J & Van de Schoot, 2017). I check the fitted model that may have outliers and influential values in the same way with standard logistic model.

In logistic regression analysis after fitting the model the adequacy of the model should be checked and it can measure based on diagnosing residuals and measure of influence. However, residual diagnostic techniques are not well developed for multinomial and ordinal logistic regression analyses. For multinomial regression models, Hosmer and Lemeshow suggested recombining outcomes according to the nominal structure of the data and applying residual strategies developed for binary logistic models (Perera, A. et al, 2016). There are comparable diagnostics that should be used to identify data problems. The logistic regression provides a variety of such statistics (Agresti, A, 2010).

Residuals

In logistic regression diagnostics Residuals are the basic building blocks and used to identify potential outliers (not well fitted by the model). The residuals for logistic regression model are typically defined as the difference between observed response, and the estimated probability of the response, conditional on the covariates. Pearson residual values fluctuate around zero, following approximately a normal distribution when n_i is large (Agresti, 2002). For a generalized linear model (GLM) with binomial random component, the Pearson residual (r_i) comparing y_i to its fit is (Agresti, A, 2019).

$$\boldsymbol{r_i} = \frac{\boldsymbol{Y_i - n_i \pi_i}}{\sqrt{\boldsymbol{n_i\, \pi_i (1-\pi_i)}}} \qquad (3.\ 29)$$

The Pearson residuals do not have unit variance since no allowance has been made for the inherent variation in the fitted value. A better procedure is to further adjust the Pearson residuals by their estimated standard deviation that contains variation due to leverage value is called standardized Pearson residual.

The standardized Pearson residual is similar with Pearson residual that it only uses the leverage from an estimated hat matrix that means for an observation i with leverage value $\hat{h_i}$. Observations with absolute standardized residual values in excess of three may indicate lack of fit. The standardized Pearson residual is given

$$\boldsymbol{Standardized\, residual} = \frac{\boldsymbol{Y_i - n_i\widehat{\pi_i}}}{\sqrt{\boldsymbol{n_i\widehat{\pi_i}(1-\widehat{\pi_i})(1-h_i)}}} \qquad (3.\ 30)$$

The term h_i in this formula is the observation’s leverage, the greater an observation’s leverage, the greater its potential influence on the model fit. The standardized residual equals $\frac{r_i}{\sqrt{(1-h_i)}}$, so it larger in absolute value than the Pearson residual. An absolute value larger than roughly two or three provides evidence of lack of fit (Agresti, A, 2019).

Deviance residuals are useful to determining individual points that are not well fitted by the (potential outliers or mis-specified cases) model. The deviance residual for the i^{th} observation is the signed square root of the contribution of the i^{th} case to the sum for the model deviance, for the i^{th}observation, and given by;

$$\boldsymbol{D_i = \pm\{-2[Y_i log\widehat{\pi_i} + (1 - Y_i)\log(1 - \widehat{\pi_i})]\}^{1/2}} \qquad (3.\ 31)$$

When$Y_i \geq \hat{\pi}_i$, D_i becomes positive otherwise it is negative. An observation with a residual greater than two or three in either direction is an indication of poor fit.

Standardized and deviance residuals are the most commonly used statistic in identifying points for which the model fits poorly. Presence of outliers is signaled if the standardized residuals lie outside the range of the interval (-3, 3). If the presence of outliers affects the inference, the decision whether or not to include them or revise the model must also require a close subjective examination of the data in addition to statistical grounds. Detecting outliers is common practice and it is important to distinguish between two types of outliers. Outliers in the response variable represent model failure. Such observations are called outliers. Outliers with respect to the predictors are called leverage points.

Leverage Values (Hat Diag) is a measure of how far an observation is from the others in terms of the levels of the independent variables (not the dependent variable). Observations with leverage values larger than one are considered to be potentially highly influential (Belsley, D. et al, 2005). The formula is: -

$$\mathbf{h}_{\mathbf{i}} = \frac{\mathbf{1}}{\mathbf{n}} + \frac{(\mathbf{x_i} - \bar{\mathbf{x}})^2}{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i} - \bar{\mathbf{x}})^2} \qquad (3.\ 32)$$

Where: $\boldsymbol{h}_{\boldsymbol{i}}$ = Leverage value, n= Number of observation

Multi-collinearity; multi-collinearity of both individual and community level variables were checked by correlation matrix to identify correlations between variables and determines the strength of the relationships. Variance inflation factor examined instability of effect size of predictors as the result of high collinearity among the factors (Schwarz, C, 2014). A simple but sometimes subjective technique is to inspect the magnitude of the standard error of each variable. The standard error is very large implying multi-collinearity exists and the model is not statistically stable. To solve this issue, start omitting the variable with high collinearity (Chan, 2022). There is no fixed criterion on how small the standard error should be but it is a matter of judgment. It is better way to identify correlated variables in the study. To confirm multi-collinearity diagnosis, it is better to drop relatively correlated variables from the analysis.

3.9. Ethical Consideration

Demographic Health Survey (DHS) dataset was obtained by registration on the DHS website (http://www.dhsprogram.com) and getting approval from the measure DHS. Data was used in this study only for the purpose of statistical analysis and reporting, for the proposed research project. The data downloaded from DHS website treated as congenital, and no effort should be made to identify any household or individual respondent interviewed in the survey. The data sets have been accessed through the web page of the international DHS program after subscription and being an authorized user.

CHAPTER FOUR

4. RESULTS AND STATISTICAL DATA ANALYSIS

The goal of this chapter is use the 2019 Ethiopian Mini Demographic Health Survey (EMDHS) to investigate several aspects that define the factors that influence vaccination incompletion for children aged 12 to 35 months in Ethiopia. There are three sections to the analysis. Results of descriptive statistics are shown in the first section. Using STATA software, we used ordinal logistic regression analysis to identify and investigate the factors that contribute to vaccine incompletion in the second half. Finally, with the aid of STATA, a multilevel ordinal logistic regression model was created to account for the variables and differences in vaccination coverage among regions of Ethiopia.

4.1 Results for Descriptive Statistics

Data analysis was presented in this study based on a total of 5753 children aged 12 to 35 months from the children's record data set of 2019 EMDHS. Table 4.1 presented below shows from a total of 5753 under-five children out of which 936(16.27%) were not received none of the recommended vaccines (not vaccinated), 3614(62.82%) were received one or more of the recommended vaccines (partially vaccinated) and 1203(20.91%) were received all the recommended vaccines (completely vaccinated).

Table 4.1: Proportion of childhood vaccination status, 2019EMDHS

Vaccination status	Frequency	Percentage
Not vaccinated	936	16.27
Partially vaccinated	3614	62.82
Completely vaccinated	1203	20.91

Source; own computing using 2019 EMDHS children record data.

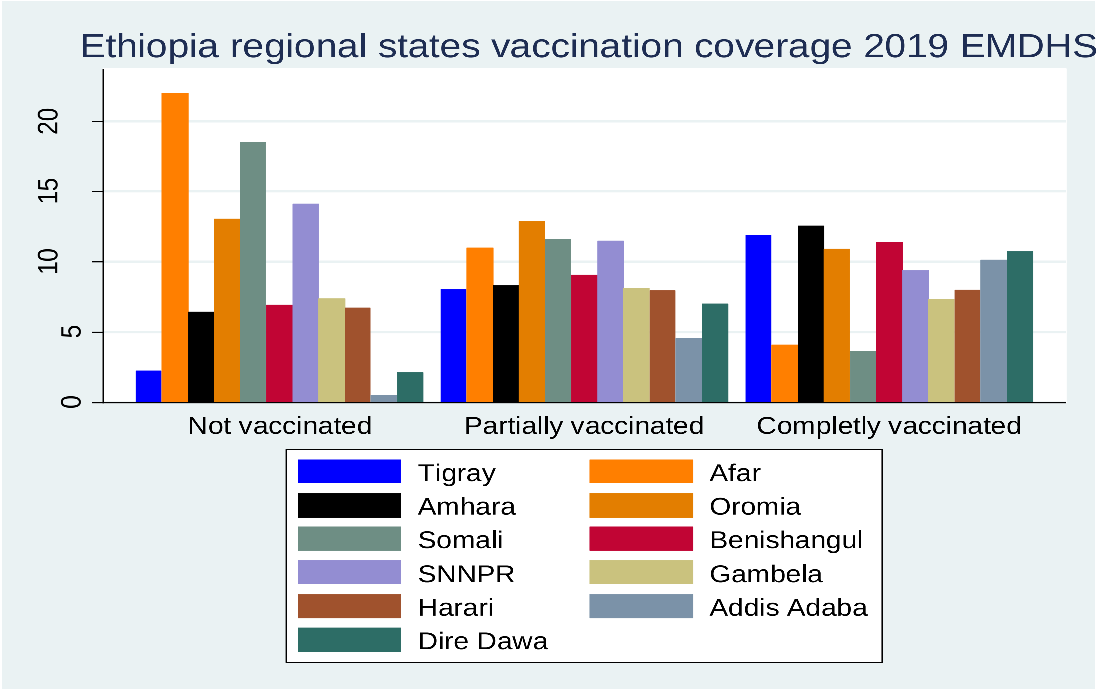

Figure 4.1: Childhood vaccination coverage in Ethiopia based on 2019 EMDHS

The multiple bar charts in Figure 4.1 given above showed that the highest number of children's aged 12 to 35 months who lived in Afar, Somalia and SNNPR were non-vaccinated, whereas children's aged 12 to 35 months who resided in Tigray, Amhara, and Benshangul were completely vaccinated. Majority of children's aged 12 to 35 months who resided in Addis Ababa and Dire Dawa specially fails in partially and completely vaccinated category.

In this study major socio-economic, demographic and reproductive health and service related characteristics were presented in Table 4.2 depicted that among under-five children 475(16%), 1874(63.12%) and 620(20.88%) were not vaccinated, partially vaccinated and completely vaccinated respectively for males. About 461(16.56%), 1740(62.5%), and 583(20.94%) were not vaccinated, partially vaccinated and completely vaccinated respectively for females. Lowest proportion of vaccination incompletion observed for under-five children were whose mother age is 15 - 19 (28.04%) compared to mother's whose age is above 45 years. However, the highest proportion of complete vaccination for under-five children's were 22.55% in 25-29 mothers age.

Regarding to highest educational level of women, majority of mothers had no formal education (3149), 1823 of them had primary, and the rest had secondary and higher education. Mothers who did not get a formal education they did not vaccinate their children's which is 19.85% as compared to mothers who had formal education. The higher proportion of complete vaccination observed for children aged 12 to 35 months were mothers have higher education level 34.88%.

With regard to sex of household head under-five children 739(16.07%), 2,882(62.68%) and 977(21.25%) were not vaccinated, partially vaccinated and completely vaccinated respectively for male household head. About 197(17.06%), 732(63.38%), and 226(19.57%) were not vaccinated, partially vaccinated and completely vaccinated respectively for females. In this regard the highest proportion of complete vaccination observed for 12 to 35 month children were 21.25% male household heads.

In case of marital status, the majority of children were born from marriage women 5397, and single mothers are more initiated for child vaccination 9(29.03%) compared to the rest status. In addition, marriage mothers had lowest proportion to non-vaccination status 883(16.36%).

Table 4.2: Demographic determinant factors of vaccination status

Child vaccination Status					Total	chisq	DF	P-value
Variables	Category	Not vaccinated Count (%)	Partially vaccinated Count (%)	Completely vaccinated Count (%)				
Sex of child	Male	475 (16%)	1,874 (63.12%)	620 (20.88%)	2,969	0.367	2	0.832
	Female	461 (16.56%)	1,740 (62.5%)	583 (20.94%)	2,784			
Mothers age	15 - 19	83 (28.04 %)	164 (55.41%)	49 (16.55%)	296	62.95	12	0.0000
	20 - 24	198 (17.32 %)	689 (60.28%)	256 (22.4%)	1,143			
	25 - 29	305 (16.42%)	1,134 (61.03%)	419 (22.55%)	1,858			
	30 - 34	196 (15.82%)	796 (64.25 %)	247 (19.94 %)	1,239			
	35 - 39	106 (13.86%)	519 (67.84 %)	140 (18.30 %)	765			
	40 - 44	43 (12.68%)	226 (66.67 %)	70 (20.65 %)	339			
	45 - 49	5 (4.42%)	86 (76.11%)	22 (19.47 %)	113			
Highest Educational	No education	625 (19.85%)	2,011 (63.86%)	513 (16.29%)	3,149			

level	Primary	260 (14.26 %)	1,119 (61.38%)	444 (24.36 %)	1,823	165.1	6	0.0000
	Secondary	39 (8.13 %)	300 (62.5%)	141 (29.37%)	480			
	Higher	12 (3.99 %)	184 (61.13%)	105 (34.88%)	301			
Sex of household head	Male	739 (16.07 %)	2,882 (62.68%)	977 (21.25%)	4,598	1.869	2	0.393
	Female	197 (17.06%)	732 (63.38 %)	226 (19.57 %)	1,155			
Current marital status	Single	4 (12.9%)	18 (58.06%)	9 (29.03%)	31	4.238	6	0.645
	Marriage	883 (16.36%)	3,378 (62.6%)	1,135 (21.03%)	5,396			
	Windowed	8 (13.11%)	43 (70.49%)	10 (16.39%)	61			
	Divorced	41 (15.47%)	175 (66.04%)	49 (18.49%)	265			

Source; own computing based on 2019 EMDHS

According to socio-economic determinant factors of children vaccination presented in Table 4.3 presented below the proportion of completely vaccinated children were higher among those women who were followers of the orthodox 483(29.96%) followed by catholic 7(29.96%). And hence, the lowest proportion of 8(9.76%) completely vaccinated children was observed among women who were followers of traditional/others by Muslim's women 511(17.18%). Likewise, wealth index has a greater impact on child vaccination the highest proportion of complete vaccination of child is recorded on the rich 29.3% and decreased to the poor 15.31%. Moreover, children's born first (23.24%) had greater chance to take all the recommended vaccines during their childhood followed by second to fourth place (21.71%). Similarly, (18.27%) and (63.46%) of the children that not vaccinated and partially vaccinated had birth order five and above. The proportion of children did not complete vaccination is also varied on the number of living children's in the family, in this instance, the highest percentage of not vaccinated were observed in four to six children (17.27%) and the highest proportion of completely vaccinated were 24.35% in 1 – 3 children in the family.

Table 4. 3: Socio-economic determinant factors of vaccination

Child vaccination Status				

Variable	Category	Not vaccinated Count[%]	Partially vaccinated Count[%]	Completely vaccinated Count[%]	Total	chisq	DF	p-value
Religion	Orthodox	145 (9%)	984 (61.04%)	483 (29.96%)	1,612	169.8	8	0.000
	Catholic	5 (15.63%)	20 (62.5%)	7 (21.88%)	32			
	Protestant	183 (17.38%)	676 (64.2%)	194 (18.42%)	1,053			
	Muslim	582 (19.57%)	1,881 (63.25%)	511 (17.18%)	2,974			
	Traditional/ others	21 (25.61%)	53 (64.63%)	8 (9.76%)	82			
Wealth index	Poor	670 (22.65%)	1,835 (62.04%)	453 (15.31%)	2,958	286.3	4	0.000
	Medium	122 (15.16%)	516 (64.1%)	167 (20.75%)	805			
	Rich	144 (7.24%)	1,263 (63.47%)	583 (29.3%)	1,990			
Birth order	1fst birth	159 (12.61%)	809 (64.16%)	293 (23.24%)	1,261	26.58	4	0.000
	2 – 4 birth	431 (16.59%)	1,603 (61.7%)	564 (21.71%)	2,598			
	5 and above	346 (18.27%)	1,202 (63.46%)	346 (18.27%)	1,894			
Number of living children	1 – 3 children	464 (15.56 %)	1,792 (60.09 %)	726 (24.35%)	2,982	51.52	4	0.000
	4 – 6 children	351 (17.27%)	1,308 (64.37%)	373 (18.36%)	2,032			
	7 and above	121 (16.37%)	514 (69.55%)	104 (14.07%)	739			

Source; own computing based on 2019 EMDHS data

According to reproductive health and service related associated factors of children vaccination presented in Table 4.4 below. Since today's world vaccination is promoted by the use of mass media like radio, TV and the like in this regard, the proportions of completely vaccinated children were higher among those women who had radio 866(19.91%) and TV 349(34.32%). While, the percentage of children's did not take any recommended vaccine were higher among those women who had not radio (17.98%) and TV (18.6%). Moreover, when we came to children's of place of delivery, the result presented in table 4,4 shows still higher numbers of children's were born at home which was 50% of the total children aged 12 to 35 months. And also the highest proportions of completely vaccinated children's were born at private health facility 35.03% followed by children's born at government health facility 29.07%. In addition, the highest proportions of not vaccinated and partially vaccinated children were born at home 23.64% and 63.42% respectively. Mothers' health care utilization plays a vital role in vaccination of their children. Respectively, mothers did and did not get assist from health professionals 63.2% and 61.16% of them partially vaccinated their children. But from completely vaccinated children respectively 24.84% and 20.01% were from mother of did not and did get assistant by health professional. When we see place of antenatal care more than 51% (2,936) women's follow antenatal care at health facility, and the proportion of not vaccinated, partially vaccinated, and completely vaccinated at health facility were 13.32%, 54.8% and 31.88% respectively. Furthermore, number of antenatal care visit play a vital role in child vaccination, and thus the highest percentage of children's did not take any recommended vaccine (not vaccinated) were 509(25.57%) mothers who had not any visit during her pregnancy. While the highest proportion of children's did take all the recommended vaccine (completely vaccinated) were 33.56% at the higher number of antenatal care visit followed by forth to fifth visit 30.11%.

Table 4.4: Reproductive health and service related determinant factors

Child vaccination Status					Total	chisq	DF	p-value
Variable	Category	Not vaccinated Count[%]	Partially vaccinated Count[%]	Completely vaccinated Count[%]				
Possession of Radio	No	782 (17.98%)	2,701 (62.11%)	866 (19.91%)	4,349	42.01	2	0.000
	Yes	154 (10.97%)	913 (65.03%)	337 (24%)	1,404			
Possession of TV	No	881 (18.6%)	3,001 (63.37%)	854 (18.03%)	4,736	197.028	2	0.000
	Yes	55 (5.41%)	613 (60.28%)	349 (34.32%)	1,017			
Place of delivery	At home	681 (23.64%)	1,827 (63.42%)	373 (12.95%)	2,881	386.872	8	0.000
	At GOV health facility	228 (8.9%)	1,590 (62.04%)	745 (29.07%)	2,563			
	At Private health facility	4 (2.55%)	98 (62.42%)	55 (35.03%)	157			

	At NGO health facility	12 (15.38%)	47 (60.26%)	19 (24.36%)	78			
	Other	11 (14.86%)	52 (70.27%)	11 (14.86%)	74			
Assistance; health professional	No	150 (14.01%)	655 (61.16%)	266 (24.84%)	1,071	14.42	2	0.001
	Yes	786 (16.79%)	2,959 (63.2%)	937 (20.01%)	4,682			
Place of antenatal care	At home	545 (19.35%)	2,005 (71.18%)	267 (9.48%)	2,817	438.49	2	0.000
	At health facility	391 (13.32%)	1,609 (54.8%)	936 (31.88%)	2,936			
Number of antenatal care visit	No visit	509 (25.57%)	1,304 (65.49%)	178 (8.94%)	1,991	438.04	6	0.000
	1 - 3	253 (15.51%)	1,010 (61.93%)	368 (22.56%)	1,631			
	4 – 5	146 (8.65%)	1,033 (61.23%)	508 (30.11%)	1,687			
	6 and above	28 (6.31%)	267 (60.14%)	149 (33.56%)	444			

Source; own computing using 2019 EMDHS data

The community level variable presented in Table 4.5 below shows percentage of partially vaccinated children aged 12 to 35 months of were varied from region to region. The first and the second minimum percentage were observed respectively in Addis Ababa (56.36%) and in Amhara (58.71%) while the first and the next maximum percentage were observed in Somali (65.93%) and in Gambela (65.11%) respectively. A significant number of respondents, 4,425(76.92%) were rural residents. In addition, taking into account residence most of rural (19.23%) and 6.4% of urban children were not vaccinated.

Table 4.5: Characteristics of the cluster variables

Child vaccination Status					Total	chisq	DF	p-value
Variable	Category	Not vaccinated	Partially vaccinated	Completely vaccinated				
		Count[%]	Count[%]	Count[%]				
Region	Tigray	21 (4.63%)	290 (63.88%)	143 (31.5%)	454			
	Afar	206 (31.6%)	397 (60.89%)	49 (7.52%)	652			

	Amhara	60 (11.74%)	300 (58.71%)	151 (29.55%)	511	530.0897	20	0.000
	Oromia	122 (16.97%)	466 (64.81%)	131 (18.22%)	719			
	Somali	173 (27.16%)	420 (65.93%)	44 (6.91%)	637			
	Benshangul	65 (12.26%)	328 (61.89%)	137 (25.85%)	530			
	SNNPR	132 (20%)	415 (62.88%)	113 (17.12%)	660			
	Gambela	69 (15.33%)	293 (65.11 %)	88 (19.56%)	450			
	Harari	63 (14.09%)	288 (64.43%)	96 (21.48%)	447			
	Addis Ababa	5 (1.72%)	164 (56.36%)	122 (41.92%)	291			
	Dire Dawa	20 (4.98%)	253 (62.94%)	129 (32.09%)	402			
Place of residence	Urban	85 (6.4%)	838 (63.1%)	405 (30.5%)	1,328	179.3	2	0.000
	Rural	851 (19.23%)	2,776 (62.73%)	798 (18.03%)	4,425			

Source; own computing based on 2019 EMDHS

A cross-tabulation or contingency table analysis is, often used to analyze categorical data, a two or more dimensional table that records the frequency (count) of respondents having the specific characteristics described in the cells of the table. The chi-square test is used to assess the relationship between two nominal or ordinal variables. It is a very general statistical test that can be used whenever we wish to evaluate whether frequencies that have been empirically obtained differ significantly from those that would be expected on the basis of chance or theoretical expectations. Chi-square statistic does not give any information about the strength of the relationship but it only conveys the existence or nonexistence of the relationships between the variables investigated. Based on the result of the cross tabulation analysis (presented in the previous descriptive tables) vaccination status of a child was found to be associated with all class variables except current marital status, sex of household head and sex of a child at 5% level of significance.

4.2. Ordinal logistics regression analysis

Before fitting the ordinal logistic regression model, a chi-square test for association was performed which is presented before in descriptive analysis, and variable selection were performed to select most significant explanatory factors to be added in the model at a specified significance level which is presented below.

4.2.1. Variable selection and Model building method

Stepwise variable selection method

Stepwise variable selection methods are a widely used variable selection technique, particularly in medical applications. This method is a combination of forward and backward selection procedures that allows moving in both directions, adding and removing variables at different steps. The process can start with both a backward elimination and forward selection approach. For example, if stepwise selection starts with forward selection, variables are added to the model one at a time based on statistical significance. Therefore, the significant covariates checked by stepwise variable selection procedures at 15% significance level then Table 4.6 below shows sex of household head, sex of child, highest education level of women, possession of Radio, possession of TV, marital status and assistant health professional are insignificant and the other variables are inter in to the ordinal and multilevel logistic regression.

Table 4. 6: Results of stepwise variable selection for ordinal logistic model

Variables	Coef.	Std.Error	Z	P >|z|	[95% Confi. Interval]
Age	0.266	0.039	6.74	0.000	(0.189, 0.344)
Region	0.051	0.015	3.49	0.000	(0.023, 0.080)
Residence	-0.434	0.136	-3.20	0.001	(-0.700, -0.168)
Religion	-0.188	0.035	-5.44	0.000	(-0.256, -0.120)
Wealth index	0.256	0.056	4.55	0.000	(0.146, 0.366)
Place delivery	0.269	0.064	4.18	0.000	(0.143, 0.396)
Place ANC	-0.644	0.111	-5.78	0.000	(-0.862, -0.426)
Num. ANCV	0.598	0.068	8.81	0.000	(0.465, 0.731)

Birth order	-0.449	0.079	-5.71	0.000	(-0.604, -0.295)
No living chil	0.174	0.076	2.30	0.021	(0.026, 0.323)

Source; own computing using 2019 EMDHS data

Checking the proportional odds assumption

One of the assumptions underlying ordered logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients (only one model). If this was not the case, we would need different models to describe the relationship between each pair of outcome groups. We need to test the proportional odds assumption, and there are two tests that can be used to do so. First, we need to download a user-written command called **omodel** (type **search omodel**) (stats.oarc). The null hypothesis is the logit surfaces are parallel/parallel line assumption is satisfied relative to the alternative hypothesis not satisfied. Test of parallel lines was designed to make judgment concerning the model adequacy.

The model null hypothesis states that the slope coefficients in the model are the same across the response categories. In Table 4.7 given below the significance p-value is 0.000<0.05 for all tests these indicated that there was significant difference for the corresponding slope coefficients across the response categories, suggesting that the assumption of parallel lines was violated in the model.

Table 4.7: Rests of parallel line regression (PO) assumption

Test	Chi-square	DF	Pr>chi2
Score	344.49	17	0.000
Brant	335.36	17	0.000
Likelihood ratio	682.41	17	0.000

Source; own computing using 2019EMDHS data

4.2.2. Model comparison for all ordinal logistics regression models

The proportional odds model was discarded due to violation of the parallelism assumption by the Brant test (chi-square = 335.36, p-value = 0.000), and the data were then fitted with partial proportional odds (PPOM), generalized ordered logit (GOM), constrained continuation ratio (CCRM), Unconstrained continuation ratio (UCRM) and adjacent category logit (ACM) models the model comparison is presented in Table 4.8 below. Finally, a model comparison was performed based on information criterion and log – likelihood values. Due to the lowest information criterion values, the PPOM was considered to the best-fitting model.

Table 4.8: AIC, BIC and -2LL values of all ordinal models

Selection Criteria	Ordinal models					
	POM	ACM	CCRM	UCRM	GOM	**PPOM**
AIC	9743.835	9532.769	9743.830	9429.381	9437.936	**9424.937**
BIC	9996.820	9796.606	9990.157	9908.719	9930.590	**9791.098**
-2LL	9,667.836	9,318.716	9,669.83	9,285.38	9,289.936	**9,314.938**

Source; own computing based on 2019 EMDHS

Then we used the generalized ordered logit model with AUTOFIT option also call partial proportional odds model (PPOM) can be fitted using STATA 14 by the command GOLOGIT2. From the Table A.1(see in the appendix) the global Wald test is performed for the PPOM with constrained versus the original unconstrained model. Wald test of parallel lines assumption for the PPOM: chi2 (17) = 20.86 Prob > chi2 = 0.2326. An insignificant test statistics indicates that the final model does not violate the proportional odds/ parallel lines assumption. The variable's region (Amhara, Oromia, Benshangul, SNNPR and Harari), age (except 20 – 24), wealth index, place of antenatal care, birth order and number of antenatal care visit violated the parallel lines' assumption in the partial proportional odds model. The model, therefore, allows the coefficients of these variables to vary across the response categories.

4.3. Results of Multilevel ordinal logistics regression model

The data used in this study have a hierarchical structure. Units at one level are nested within units at the next higher level. Here, the lower level (level-1) units are the individual child, and the higher level (level-2) units are the regions that constitute the groups into which the child are clustered or nested. The nesting structure is child within regions that resulted in a set of 11 regions with a total of 5753 child. The data used in this study consist of variables describing individuals as well as variables describing regions. Therefore, the statistical model used has to describe the data at both levels, in order to find the effect of childhood vaccination level both the individual and the regions level. The advantages of using a multilevel model include the ability to fully explore the variability at all levels of the data hierarchy, and estimation of correct standard errors in the presence of clustered data.

Test of heterogeneity proportion

Before starting to multilevel analysis, one has to test for the heterogeneity of under-five children childhood vaccination status among regions of Ethiopia. A chi-square test statistic was applied to assess heterogeneity in the proportion of childhood vaccination among the regions in Ethiopia. The test yields a Pearson ${X^2}_{(20)} = 530.1$ with $p - value\ =\ 0.000$ < 5% level of significance. Thus, there is evidence for heterogeneity among the regions with respect to childhood vaccination incompletion of under-five children in Ethiopia.

4.3.1. Multilevel ordinal logistics regression model comparison

In order to see whether the inclusion of level one covariates (child's/parent's level factor) with place of residence varying across regions significantly improved the random intercept only model, log likelihood ratio test, AIC and BIC values have been used.

The Table 4.9 presented below shows likelihood ratio test of random intercept with fixed effect model versus random coefficient model yields LR chi2(1) = 100 which is the difference between the two deviances and p =0.000 which implies that the null hypothesis of no difference between the two models is rejected.

From Table 4.9 given below both the AIC (9769.322) and BIC (9955.732) values of the random intercept with fixed effect model are greater than the AIC (9766.596) and BIC (9953.320) values of the random coefficient model indicating that the random coefficient model is preferred model.

Table 4.9: Results of multilevel ordinal logistics model selection criteria

	Multilevel ordinal logistics Model		
Criteria	Empty model (Random intercept only)	Random Intercept with fixed slope model	Random Intercept with random coeficient model
AIC	10054.21	9769.322	9766.596
BIC	10074.18	9955.732	9953.320
-2LL	10,050.206	9,713.3224	9,706.596
LR test	476.59(p–value= 0.000)	93.27(p – value = 0.000)	100(p – value = 0.000)

Source; own computing using 2019EMDHS children record data.

4.3.2. Random coefficient (slope) model

Multilevel random coefficient ordinal logistic regression model can allow the coefficient of level-one covariates to vary across regions instead of keeping them fixed across regions. Now we are going to see the effect of child's/parent's level covariates by allowing them to vary randomly across regions. This model contains fixed effects and random effects. The fixed effects are analogous to standard logistic regression coefficients and are estimated directly. The random effects are not directly estimated but are summarized in terms of their estimated variances and covariance. The random effects can take random intercepts (regional effects) and random coefficients (level-one covariates effect). In this section we investigate whether level-one covariates have random effects across regions or they have the same effects across regions.

Estimates of this model show that the random slope variances of all included variables except for place of delivery are approximately zero. This indicates that only the effects of place of delivery on childhood vaccination varied across regions, whereas the effect of other covariates for childhood vaccination remained fixed across regions. This may be due to the fact that place of delivery can be as community variables because in each region and federal states there is different health facility regarding to the presence of qualified government hospitals and health service institutions, the number of NGO and private health service sectors in each region. As usual it has specified an unstructured covariance matrix, since we have complete and balanced set of data.

From the Table 4.10 presented below the estimate of the fixed intercept is -1.064 and 2.316 the log-odds of the probability of childhood vaccination incompletion when all level one covariates are zero in region j is given by $-1.064 + \hat{u}_j$ and $2.316 + \hat{u}_j$ where $\hat{u}_j$ is a random intercept with variance of 0.146 (indicated in the table 4.11 as$\delta^2{}_{u0}$) which is the between-regions variance and its standard error is 0.067. The variance of intercept in the random slope model is still large; thus, there remains some regional level variance unaccounted for in the model. In the absence of level-one covariates, the status of each region on childhood vaccination incompletion as compared to the average childhood vaccination status measured with the cumulative log odds depends on the sign of the random intercept, $\hat{u}_j$. When $\hat{u}_j$is positive the cumulative log odds of childhood vaccination incompletion is higher than the average and when $\hat{u}_j$is negative the cumulative log odds of childhood vaccination incompletion is less than the average.

The value var(PDL) is the estimated variance of place of delivery. These estimated variances indicated that there is a significant variation in the effect of place of delivery on childhood vaccination incompletion of children's across regions in Ethiopia. The individual region slopes of place of delivery of children vary with a variance of 0.0177.

By adding level-1 predictors, the ICC increased and is estimated as$\hat{\rho} = 0.0425$, which was found to be significant at 5% level of significance, showed that 4.25% of the total variation in childhood vaccination status were explained or accounted by regions (level – 2 units) reveals that multilevel ordinal logistic regression is appropriate model and the remaining 95.75% of the variation in childhood vaccination status for 12 – 35 months children in Ethiopia were explained by lower level units (individual level variables) and other unknown factors within regions.

The random coefficient estimates for intercepts and the slopes vary significantly at 5% significance level, which implies that there is a considerable variation in the effects of children place of delivery; these variables differ significantly across the regions.

The unconditional model (intercept only model) considered as base line model which is compared with the best fitted model (random coefficient model: see model fit statistics value above Table 4.9) in order to determine the proportional change in variance (PCV) of the parsimonious model in this study. From the unconditional (empty) model, the researcher's added child's/parents' level predictor variables. As a result, the community level variance decreased from 0.412 to 0.146. Then, the proportion of variance explained at community level for childhood vaccination incompletion is calculated as follow:

$$R^2 = PCV = \frac{\delta^2_{u0[Empty]} - \delta^2_{u0[randomslope]}}{\delta^2_{u0[Empty]}} = \frac{0.412 - 0.147}{0.412} = 0.643 = 64.3\%$$

The proportion of explained variance at community level for childhood vaccination incompletion is about 64.3%. This implies that 64.3% of the variability for childhood vaccination incompletion between the regions is explained by women-level and cluster-level variables. As a result, the explained variances between the regions in the childhood vaccination incompletion were about 64.3% in Ethiopia since 2019. In general, the highest (64.3%) PCV in the random coefficient model revealed that 64.3% of the community-level variation on child vaccination has been explained by the combined factors at both the individual and community levels.

Table 4.10: Results of random coefficient model

Parameter	Coef.	Std.Error	Z	p>|Z|	95% CI
Fixed part					
Age (15 – 19 = reference)					
20 – 24	0.519	0.139	3.74	0.000*	[0.247 0.792]
25 – 29	0.628	0.142	4.42	0.000*	[0.349 0.906]
30 – 34	0.585	0.154	3.79	0.000*	[0.282 0.887]
35 – 39	0.582	0.165	3.52	0.000*	[0.258 0.906]
40 – 44	0.791	0.189	4.16	0.000*	[0.419 1.163]
45 – 49	1.052	0.244	4.31	0.000*	[0.573 1.531]
Residence (Urban = reference)					
Rural	-0.264	0.088	-2.99	0.003*	[-0.437 -0.091]
Religion (orthodox = reference)					
Catholic	0.157	0.386	0.41	0.684	[-0.599 0.913]
Protestant	-0.220	0.107	-2.06	0.040*	[-0.431 -0.010]
Muslim	-0.092	0.095	-0.97	0.331	[-0.277 0.093]
Traditional/others	-0.591	-2.39	0.248	0.017*	[-1.077 -0.106]
Wealth index(poor = reference)					
Middle	0.047	0.088	0.54	0.590	[-0.125 0.219]
Rich	0.214	0.083	2.56	0.010*	[0.050 0.377]
Place of delivery (at home = reference)					
GOV health facility	0.395	0.084	4.72	0.000*	[0.231 0.559]
Private health facility	0.414	0.208	1.99	0.047*	[0.006 0.823]
NGO health facility	0.326	0.300	1.09	0.277	[-0.262 0.914]
Other	0.016	0.316	0.05	0.960	[-0.604 0.636]
Place of ANC(at home = reference)					
At health facility	0.469	0.081	5.80	0.000*	[0.311 0.629]
NANCV(No visit = reference)					
1 – 3 visit	0.081	0.098	0.83	0.408	[-0.111 0.274]
4 – 5 visit	0.317	0.101	3.13	0.002*	[0.119 0.516]

6 and above	0.189	0.144	1.31	0.189	[-0.093 0.471]
Birth order (First birth = reference)					
2 – 4	-0.146	0.082	-1.79	0.074	[-0.306 0.014]
5 and above	-0.229	0.121	-1.89	0.058	[-0.467 0.009]
Number of living children (1 – 3 children = reference)					
4 – 6	0.113	0.083	1.36	0.173	[-0.049 0.277]
7 and above	0.094	0.120	0.79	0.432	[-0.141 0.329]
α_1, Intercept 1/cut 1	-1.064	0.208	-5.13	0.000	[-1.471 -0.658]
α_2, Intercept 2/cut 2	2.316	0.210	11.01	0.000	[1.903 2.728]
Random part					
Variance covariance estimate [Reg Unstructured]					
Var(cons) /$\delta^2{}_{u0}$	0.146	0.067			[0.059 0.359]
$Place\ delivery\ var(PDL)$	0.018	0.018			[0.002 0.128]
$Cov(\delta^2{}_{u0}, PDL)$	-0.018	0.0312	-0.58	0.563	[-0.079 0.043]
LR test vs. oligit. Chibar2(01) = 100.00 Prob >= chibar2 = 0.000					
Log likelihood = -4853.298 -2LL = 9706.596					
AIC = 9766.596 BIC = 9953.32 ICC = (0.146)/(0.146+3.29) = 0.042					

Source; own computing using 2019 EMDHS children record data.* imply significant at 5% level

Parameter interpretation of the multilevel random coefficient ordinal logistic model

The parameter interpretation in the multilevel ordinal logistic regression is similar to the classical ordinal logistic regression or the partial proportional odds model.

In this study mother's age is a statistically significant determinant factor for childhood vaccination status. When non-vaccinated status is compared to partially vaccinated and completely vaccinated status revealed that child born from mother's aged 20 – 24, 25 – 29, 30 – 34, 35 – 39, 40 – 44 and 45 – 49 are $exp(0.519) = 1.680$, 1.874, 1.795, 1.789, 2.206 and 2.863 times more likely of being in the partially vaccinated and completely vaccinated status as compared to child born from mother's aged 15 – 19 respectively. Given all other variables are held constant in the model.

The multilevel random coefficient ordinal logistic regression result revealed that children born from whose mother place of antenatal care visit at health facility is $exp(0.469) = 1.598$ times more likely of being in the partially and completely vaccinated status(as opposed to non-vaccinated status) as compared to children born from whose mother place of antenatal care visit at home. Similarly, children's from rich family member are $\exp(0.214) = 1.239$ times more likely of being in partially and completely vaccinated as compared to children's from poor family members, keeping all other variables in the model constant.

Number of antenatal care visit during pregnancy is also significantly associated with vaccination incompletion, the result revealed that children's born from whose mothers had four to five visit $OR = \exp(0.317) = 1.737$ times more likely of being in partially and completely vaccinated as (as opposed to non-vaccinated status) compared to children's born from whose mothers did not have any visit, keeping all other variables in the model constant.

According to the results of multilevel random coefficient ordinal logistic regression, children's from protestant follower and traditional/other follower mother are $OR = \exp(-0.220) = 0.803$ and $\exp(-0.591) = 0.554$ times less likely of being in the partially and completely vaccinated as compared to children's from orthodox follower mother respectively.

From the results of the multilevel random coefficient ordinal logistic regression revealed that place of residence of women is significant determinant factor of childhood vaccination incompletion. Children born from women's lived in rural areas are $exp(-0.264) = 0.768$ times less likely of being in the partially vaccinated and completely vaccinated (as opposed to non-vaccinated status) compared to children born from women's lived in urban areas, given the other variables are held constant in the model.

Place of delivery is also a significant determinate factor for childhood vaccination incompletion, the results of the random coefficient model showed children's born from government health facility and private health facility is 1.484 and 1.513 times more likely of being in partially and completely vaccinated (as opposed to non-vaccinated status) as compared to children's born from home respectively, given all other variables are held constant in the model.

4.3.3. Goodness-of-fit test of multilevel random coefficient model

The goodness of fit of the modeling present in Table 4.11 above the result showed that the multilevel random coefficient ordinal logistic regression model provides a significantly better result compared to the ordinal logistic modeling without considering the random effects, which can be seen from the results of the LR test (LR test vs. ordinal logistic regression chi-square

=100, P-value = 0.000) confirms that the difference is significant at 0.05 level, which verifies the better goodness of fit for the multilevel ordered logistic model.

4.4. Model Diagnosis checking

The assumption of level two residuals was checked by using scatter plot and normal probability plot and of standardized residuals and leverage values for higher level variables with the corresponding response category.

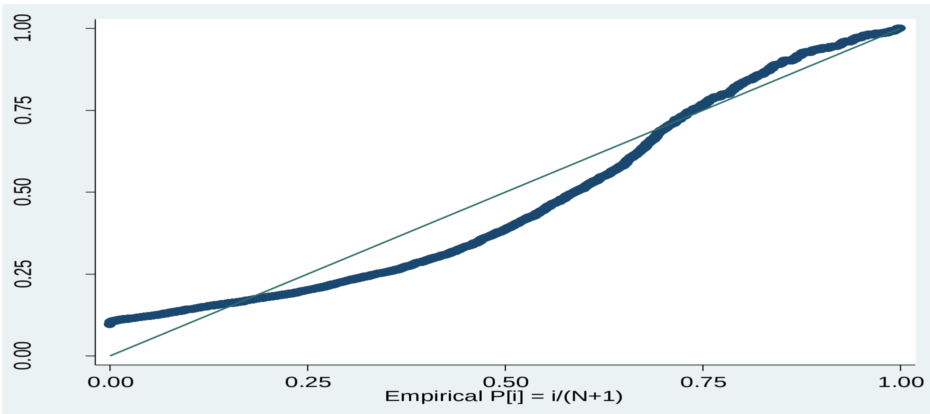

Figure 4.2: Normal probability plot of the standardized residual

Therefore, Figure 2 above shows the assumption of linearity and normality of level-2 residuals the researcher observed that the assumption of normality was resembles normal curve and we can say purely satisfied the linearity and normality assumption the residual.

Diagnosing residuals outliers, leverage values and influential points difficult to do for ordinal and multinomial logistic models in order to reduce difficulty, the ordinal response variable categories changed binary categories by collapsing two or more categories (Hosmer,D and Lemeshow, 2003).

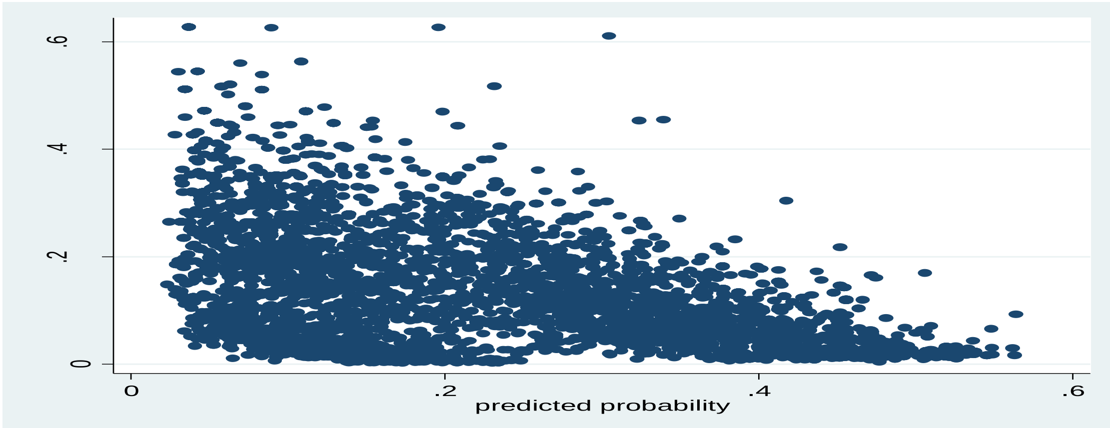

Figure 4.3: Plots of standardized residual against predicted probability

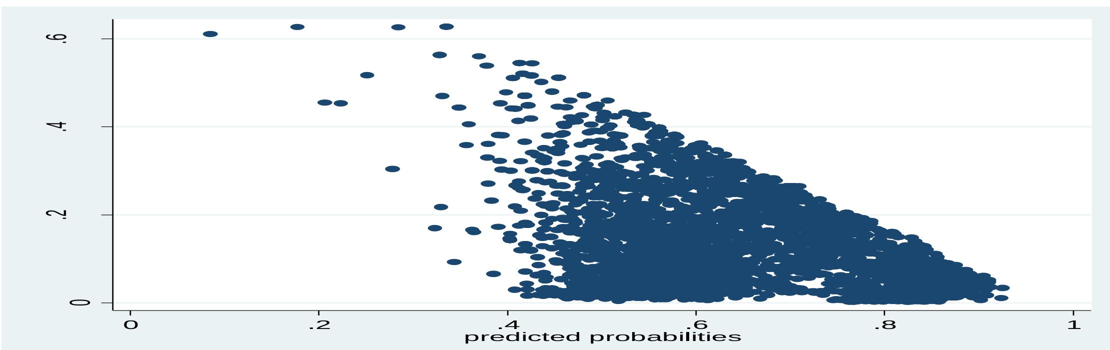

Figure 4. 4: Plots of Leverage Value against Predicted Probability

Due to collapsing of the two categories into one including partially and completely changed into vaccinated and not vaccinated. Plot of standardized residuals and leverage value versus predicted probabilities of all observations in Figure 4.3 Figure 4.4 above revealed that the model was adequate. However, the standardized residuals do not influencing the model, therefore, there are no outliers.

Test of multi-collinearity

Multi-collinearity is a problem because it can increase the variance of the regression coefficients, making them unstable and difficult to interpret. Due to this problem, the researcher checked the existence or absence of multi-collinearity among individual-level and community-level variables by using a correlation matrix displayed from table A1 (see Appendix), the correlation between all community-level and individual-level variables were less than 70%. As a result there is no multi-collinearity problem among the independent variables.

CHAPTER FIVE

5. DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1. Discussions

In this study, the vaccination status of children aged 12 to 35 months were assessed and ranked as an ordinal response depending on the number of vaccines the child received. Overall, 83.73% of children were vaccinated, with 20.91% receiving all recommended vaccines and 62.82% receiving only some. Only 16.27% of children's had received none of the recommended vaccinations this result is consistent with a result reported by (EMDHS, 2019).

Model comparisons were performed after fitting the data with proportional odds, partial proportional odds, generalized ordered logit, adjacent category logit, continuation ratio models and finally multilevel ordinal logistic regression model were fitted. As a result, based on information criteria and LR values the multilevel random coefficient model was found to the best fit our data and it used to find associated factors influencing children's vaccination status and explore variability of childhood vaccination across regions. Parameter estimates were presented and explained for the relevant predictors at 5% level of significance.

From the multilevel random coefficient ordinal logistic regression model fixed effect part was used to examine the demographic, socio-economic and health service related determinants of childhood vaccination incompletion among 12 – 35 month old children. Based on the findings of previous results, this study made a few comparative discussions as follow.

Results of random coefficient model revealed that age of mother had significant effect on status of childhood vaccination incompletion. A child's whose mother aged 20 – 24 $OR = 1.68$ [$\beta = 0.519;\ 95\%CI$ [0.247 0.792]] times more likely of being in partially and completely vaccinated as compared to children's whose mother aged 15 – 19. Similarly, a child's whose mother aged 25 – 29 $OR = 1.874$ [$\beta = 0.628, 95\%CI$ [0.349 0.906]] times more likely of being in partially and completely vaccinated as compared to children's whose mother aged 15 – 19. Although, a child's whose mother aged 30 – 34 $OR = 1.795$ [$\beta = 0.585, 95\%CI$ [0.282 0.887]] times more likely of being in partially and completely vaccinated as compared to children's whose mother aged 15 – 19. Also, a child's whose mother aged 35 – 39 $OR = 1.789$ [$\beta = 0.582$, $95\%CI$ [0.258 0.906]] times more likely of being in partially and completely vaccinated as compared to children's whose mother aged 15 – 19. This implies that old age mother's had higher odds to vaccinate their kids. This result was supported by (Adokiya et al, 2017) from Ghana, and (Metkie, K. A. et al, 2023) from Ethiopia. But contradict with (Mekonnen ZA, et al, 2020) a study done from Ethiopia.

In addition, religion of the children's mother had a significant effect on childhood vaccination incompletion, in this study the result of random coefficient model revealed that children's from protestant follower mother are $OR = 0.803$ [$\beta = -0.220, 95\%CI$ [-0.431 -0.010]] times less likely of being in the partially and completely vaccinated as compared to children's from orthodox follower mother. While, children's from traditional/other follower mother are $OR = 0.554$ [$\beta = -0.591, 95\%CI$ [-1.077 -0.106]] times less likely of being in the partially and completely vaccinated as compared to children's from orthodox follower mother this finding is line up with the study conducting in Ghana (Adokiya et al, 2017).

Place of residence of a child also had a significant effect on childhood vaccination status children's from rural women $OR = 0.768$ [$\beta = -0.264, 95\%CI$ [-0.437 -0.091]] times less likely of being in partially and completely vaccinated as compared to children's from urban women this finding is supported by the research from Ethiopia (EMDHS, 2019), (Gizachew W, 2020) and (Metkie, K. A. et al, 2023).

With respect to wealth index of the family, children's from rich family member are $OR = 1.239$ [$\beta = 0.214, 95\%CI$ [0.050 0.377]] times more likely of being in partially and completely vaccinated as compared to children's from poor family members this may be due to the fact that rich family has ability to pay and desire good balanced life this finding is supported by (EMDHS, 2019) and a research conducted from Ethiopia (Alemu, K, 2014), (Fenta S, et al, 2021), (Lakew et al , 2015). (Mekonnen ZA, et al, 2020), (Gizachew W, 2020) and (Melaku et al, 2020).

Similarly, place of antenatal care is also strongly correlated with childhood vaccination in this study random coefficient model result revealed that children's born from mothers follow antenatal care at health facility is $OR = 1.598$ [$\beta = 0.469, 95\%CI$ [0.311 0.629]] times more likely of being in partially and completely vaccinated as compared to children's born from mothers follow antenatal care at home. This may be due to hospitals are suitable for antenatal purpose and mothers prefer health institution for antenatal care service this finding is supported by a research conducted in Ethiopia by (Alemu, K, 2014). Although, children's born from government health facility is $OR = 1.484$ [$\hat{\beta} = 0.395$ $95\%CI$ [0.231 0.559]] times more likely of being in partially and completely vaccinated as compared to children's born from home. While, children's born from private health facility is

$OR = 1.513$ [$\hat{\beta} = 0.414, 95\%CI$ [0.006 0.823]] times more likely of being in partially and completely vaccinated as compared to children's born from home. This result is supported by a research done in Ethiopia by (Fenta S, et al, 2021), (Meleko et al , 2017), (Gizachew W, 2020) and (Metkie, K. A. et al, 2023).

According to this study number of antenatal care visit had a significant effect on childhood vaccination incompletion the result of random coefficient model showed that children's born from whose mothers had four to five visit $OR = 1.373$ [$\beta = 0.317$ $95\%CI$ [0.119 0.516]] times more likely of being in partially and completely vaccinated as compared to children's born from whose mothers did not have any visit. This implies that number of antenatal care visit during pregnancy increases the odds of complete childhood vaccination also increase this finding is consistent with a study done in Ethiopia by (Alemu, K, 2014), (Fenta S, et al, 2021), (Legesse & Dechasa, 2015), (Mekonnen ZA, et al, 2020), (Gizachew W, 2020) and (Metkie, K. A. et al, 2023).

From the results of the multilevel random coefficient ordinal logistic regression the random effect part showed that there is a clear variation on childhood vaccination coverage across regions in Ethiopia this finding was line up with (Abadura et al., 2015), (Fenta S, et al, 2021) and (Gizachew W, 2020) a study done in Ethiopia.

5.2. Conclusions

In this study the researcher considered the explanatory variables such as age of mother, mother's highest education level, place of residence, region, religion, wealth index, place of antenatal care visit, number of antenatal care visit during pregnancy, birth order of children, place of delivery of child, number of living children, sex of child, sex of household head, assistant health professional, possession of radio, possession of TV and current marital status which are available in 2019 EMDHS data. From the results of this study, it was found that various factors associated with childhood vaccination incompletion..

Due to the hierarchical nature of DHS data both single level (ordinal logistics modeling) and multilevel analyses were used for statistical data analysis. And hence the multilevel fixed effect part was used to investigate the determinant factors of childhood vaccination incompletion, while the multilevel random effect part was also used to explore the variability of childhood vaccination across regions in Ethiopia.

Multilevel logistic regression model allows for comparison of childhood vaccination variations between regions. Before data analysis for multilevel approach, heterogeneity of the childhood vaccination status with regard to regions was checked first using chi-square test [$\chi 2 = 530.0897, p-value = 0.000$] and it was statistically significant.

In multilevel models random coefficient with fixed effects of the explanatory variables had similar interpretation as that of the partial proportional odds model, whereas the random parts of the intercept and the coefficients provided additional information. Results obtained based on the empty model the overall variance of the constant term suggest that children status of vaccination differed across regions. In addition to null model, two other models were fitted, one with random intercept and fixed effect model and another with random coefficient (slope) model. Based on Log Likelihood, AIC and BIC, two level random intercept and random slope model was fitted the data well. The overall variance constant term random intercept and random coefficient models was found to be statistically significant implying that status of childhood vaccination differ across regions in Ethiopia. However, from among the above listed explanatory variables age of mother, place of residence, region, religion, wealth index, place of antenatal care visit, number of antenatal care visit during pregnancy, and place of delivery of child were found to be significantly associated with childhood vaccination incompletion among children's of age 12 – 35 month old in Ethiopia. Moreover, place of delivery of children is also found differ across regions. The plots of outlier residuals and influential points confirmed that observations are well fitted by the model.

5.3. Recommendations

The study revealed that the proportion of non-vaccinated children was about 16.27%. And the prevalence of non-vaccinated status among children's from rural women was 19.23% and about 22.65% among children from poor mothers. Based on the findings of the study the following recommendations had been suggested. Treatments are better to focus on targeting essential childhood vaccinations by promoting rural maternal access to health care service that are included the low economic family, because the finding showed low vaccination coverage in the rural resident and poor economic level. Minster of health and other allied organizations need to increase child vaccination coverage in rural areas of Ethiopia by promoting place of delivery at health institution and create awareness on antenatal care visits during pregnancy. As there is low coverage of childhood vaccination on children's from poor economic level family, public initiatives needed to improve child vaccination coverage, and further advancement of health care services for poor women, As there is a clear regional variation on childhood vaccination incompletion in Ethiopia, policies and programs should be aim at addressing cluster variations in child vaccination need to be formulated and their implementation must be strongly followed up. Further research should look at spatial data analysis to deal with the geographical distribution in the vaccination status among under-five children in Ethiopia.

5.4. Strength and Limitation of the Study

This study use secondary data so it has its own limitation. The study's strength might be in the high response rate and the way it took into account the ordinal aspect of vaccination status by incorporating various variables. Due to unavailability of the data, numerous missing values, many important explanatory variables, such as the HIV status of women, media exposure, Wise to have a balanced family life, distance to health facility and husband/household head education level were not included in this study. Since the EDHS are a questionnaire-based survey that relies on respondents' recollections, recall bias in the data could be a weakness in this study. In addition, the tangibility of the data set was depends on the study subjects, data collectors and editors. In general, there was a convergence problem to fit complex models in the STATA software. Due to this reason, interaction effect of multilevel ordinal logistic regression model was not fitted in this study.

REFERENCE

Abadura et al. (2015). Individual and community level determinants of childhood full immunization in Ethiopia: a multilevel analysis. *BMC Public Health*, 15:972 .

Adebiyi, F. (2013). Determinants of full child immunization among 12-23 months old in Nigeria

Adokiya et al. (2017). Evaluation of immunization coverage andits associated factors among children 12–23months of age in Techiman Municipality,Ghana. *Archives of Public Health*, 75-28.

Agresti, A. (2010). *Analysis of ordinal categorical data (Vol. 656). John Wiley & Sons.*

Agresti, A. (2012). *Categorical data analysis (Vol. 792). John Wiley & Sons.*

Agresti, A. (2019). *An introduction to categorical data analysis.* John Wiley & Sons.

Alemu, K. (2014). Factors affecting childhood vaccination incompletion in children aged 12 to 59 months in Ethiopia. *Journal of the Ethiopian Statistical Association*.

Anokye et al. (2018). Socio-demographic determinants of childhood immunization incompletionin Koforidua, Ghana. *BMC Research Note*.

Aragaw K. (2014). Vaccination Incompletion of Children Aged 12 to 59 months in Ethiopa. *Addis Ababa university*, 49-63.

Becg, C. B & Gray, R. (1984). *Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika.*

Belsley, D. et al. (2005). *(2005). Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons.*

Borba RC, et al. (2015). The reemergency and persistence of vaccine preventable diseases. Anais da Academia Brasileira de Ciências. 311-22.

Bosker & Snijders. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling. Multilevel Analysis* (Vols. 1-368.).

Byström, E et al. (2020). Confidence in the National Immunization Program among parents in Sweden 2016–A cross-sectional survey. Vaccine. 38(22), 3909-3917.

Chan. (2022). *Mitigating the multicollinearity problem and its machine learning approach: a review. Mathematics.*

Clogg & Shihadeh. (1994). *Statistical models for ordinal variables. (No Tittle.*

Dessie & Negeri. (2018). Determining factors of full immunization of children among 12-23 months old in rural Ethiopia. *Am J Public Health Res*, 6(3), 160-5.

EDHS. (2011). *Ethiopian Demographic and health survey (EDHS).* Addis Ababa, Mary land, USA: Ethiopia and Carverton Mary land, USA.

EMDHS. (2019). *Mini Demographic and health survey ke indicators report.* Addis Abeba: EPHI.

EMOH. (2018). *The Federal Democratic Republic of Ethiopia Ministry of Health HSTP Health Sector Transformation Plan.* Addis Abeba: MOH.

Fenta S, et al. (2021). (2021). Individual and community-level determinants of childhood vaccination in Ethiopia. *Archives of Public Health, 79(1)*, 1-11.

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge university press.

Gizachew W. (2020). Determinants of vaccination incompletion among childrens aged 12 - 35 months in Ethiopia; an application of ordinal and multilevel logistics regression models. *Debre Berhan University* .

GVAP. (2020). *THE GLOBAL VACCINE ACTION PLAN 2011 - 2020: Strategic Advisory Group of Experts on Immunization.*

Habyarimana, F, et al. (2014). A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under five years in Rwanda. *Mediterranean Journal of Social Sciences*, 5(23), 1642.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (Vol. 398). John Wiley & Sons.*

Hosmer, D & Lemeshow, S. (1988). Goodness-of-fit testing for the logistic regression model when the estimated probabilities are small. *Biometrical Journal*, 30(8), 911-924.

Hosmer,D and Lemeshow. (2003). Hosmer and Lemeshow type goodness-of-fit statistics for the Cox proportional hazards model. Handbook of statistics. 23, 383-394.

Hox, J & Van de Schoot. (2017). Multilevel analysis: Techniques and applications. Routledge.

Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications. Routledge.*

Hox, J.et al. (2017). *Multilevel analysis: Techniques and applications. Routledge.*

JAMA. (2015). Vaccine Preventable Deaths and the Global Immunization Vision and Strategy, 2006-2015. *JAMA 2006;* , 295: 2840-2.

Khazaei Z, et al. (2018). Globocan estimates of incidence and mortality worldwide prostate cancers and their relationship with the human development index. *Global cancer statistics, Adv Hum Biol. 2019;9(3)*, 245.

Kidane et al. (2008). National EPI coverage survey report in Ethiopia, 2008 . *Ethiopian Journal of Health Development*, 22(2), 148-157.

Klugman, J. (2011). *Human Development Report 2011. Sustainability and Equity: A better future for all. Sustainability and Equity: A Better Future for All (November 2, 2011).* UNDP-HDRO Human Development Reports.

Lakew et al . (2015). Factors influencing full immunization coverage among 12–23 months of age children in Ethiopia: evidence from the national demographic and health survey in 2011. BMC public health. *BMC Public Health*, 15, 1-8.

Legesse & Dechasa. (2015). An assessment of child immunization coverage and its determinants in Sinana District, Southeast Ethiopia. *BMC Pediatric Health*, 15(1), 1-14.

Lesaffre, E & Albert, A. (1989). Partial separation in logistic discrimination. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 51(1), 109-116.

Luman, E. et al. (2008). Compliance with vaccination recommendations for US children. *American journal of preventive medicine*, 34(6), 463-470.

Maman, K et al. (2005). The value of childhood combination vaccines: from beliefs to evidence. Human vaccines & immunotherapeutics. *11(9),* 2132-2141.

McCullagh. (1980). *Regression models for ordinal data. Journal of the Royal Statistical Society: Series B (Methodological).*

McCullagh,P & Nelder, J . (1989). *Binary data. In Generalized linear models (pp. 98-148). Springer US.*

McHugh, M. (2009). The odds ratio: calculation, usage, and interpretation. Biochemia medica. 19(2), 120-126.

McHugh, M. (2009). The odds ratio: calculation, usage, and interpretation. Biochemia medica. 19(2), 120-126.

Mekonnen ZA, et al. (2020). Timely completion of vaccination and its determinants among children in northwest, Ethiopia: a multilevel analysis. *BMC Public Health, 20(1),* 1–13.

Melaku et al. (2020). Spatial distribution of incomplete immunization among under-five children in Ethiopia: evidence from 2005, 2011, and 2016 Ethiopian demographic and health survey data. BMC Public Health, 20, 1-22. . *BMC PublicHealth*.

Meleko et al . (2017). Assessment of child immunization coverage and associated factors with full vaccination among children aged 12–23 months at Mizan Aman town, bench Maji zone, Southwest Ethiopia. *International journal of pediatrics*.

Merlo, J et al . (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journalf Epidemiology & Community Health*, 59(12), 1022-1029.

Metkie, K. A. et al. (2023). Determinants of immunization status among 12–24 months old children in Ethiopia: Using 2019 Ethiopian mini demographic and health survey data. *PloS one*, 18(3), e0283629.

Mutua, M et al. (2011). Childhood vaccination in informal urban settlements in Nairobi, Kenya: who gets vaccinated? *BMC public health*, 11, 1-11.

Ngunyi, A.et al . (2014). *On the estimation and properties of logistic regression parameters.*

Nour TY, et al. (2020). Predictors of immunization coverage among 12–23 month old children in Ethiopia: systematic review and meta-analysis. *BMC Public Health, 20(1)*, 1–19.

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables (Vol. 146).*

Perera, A. et al. (2016). *A goodness of fit test for the multilevel logistic model. Communications in Statistics-Simulation and Computation.*

Rabe-Hesketh, S. et al. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167-190.

Rutherford et al. (2010). How access to health care relates to under-five mortality in sub-Saharan Africa: systematic review. *Tropical medicine & international health*, 15(5), 508-519.

Schermelleh-Engel,et al . (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online* , 8(2), 23-74.

Schwarz, C. (2014). *Examining the impact of multicollinearity in discovering higher-order factor models. Communications of the Association for Information Systems.*

Sisay, H. (2018). A Research Thesis Submitted to Department of Public Health, College of Medicine and Health Sciences, School of Graduate Studies, Haramaya University.

Smith, T & McKenna, C. M. (2013). A comparison of logistic regression pseudo R2 indices. Multiple Linear Regression Viewpoints. 39(2), 17-26.

Snijders and Bosker . (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage.*

Solis-Soto, M. et al. (2020). Relationship between vaccination and nutritional status in children. *Demographic Research,* 42, 1-14.

stats.oarc. (n.d.). *https://stats.oarc.ucla.edu/stata/dae/ordered-logistic-regression/*.

Tadesse et al. (2009). Predictors of defaulting from completion of child immunization in south Ethiopia, May 2008 – A case control study. *BMC Public health*.

Tadesse, et al. (2009). Predictors of defaulting from completion of child immunization in south Ethiopia, May 2008–A case control study. *BMC public health*, 9, 1-6.

Tom, A. et al. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage.*

Train, K . (2009). *Discrete choice methods with simulation.* Cambridge : Cambridge university p

UHC/UCN/WHO. (2022). *status of immunization coverage in Africa as of the end of 2022.*

Umoke et al. (2021). Investigating Factors Associated with Immunization Incompletion of Children Under Five in Ebonyi State, Southeast Nigeria: Implication for Policy Dialogue. *Global Pediatric Health, volume-8,* 1-18.

UN. (2013). The millennium development goals report 2013. United Nations Pubns. 365, 366.

UNICEF & WHO. (2009). *Immunization Facts and Figures.*

Waisbord & Larson. (2015). Why invest in communication for immunization: Evidence and lessons learned. 2005. Retrieved from.

Watkins, K. (2016). *The State of the World's Children 2016: A Fair Chance for Every Child. UNICEF. 3 United Nations Plaza, New York, NY 10017.*

WHO. (2009). *State of the World's Vaccines and Immunization.* Geneva: World Health Organization, UNICEF and the World Bank 3rd edn.

WHO. (2012). *World Health Organization. Weekly Epidemiological Record.*

WHO. (2015). *World Health Organization. (2015). Expanded programme on immunization prototype curriculum for medical schools in the WHO African Region:*

WHO. (2022). *Global immunization coverage reported in 2022.*

Williams. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The stata journal*, 6(1), 58-82.

Williams, D. (2001). *Weighing the odds: a course in probability and statistics.* Cambridge

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The stata journal*, 6(1), 58-82.

Xu et al. (2022). Bayesian approaches to assessing the parallel lines assumption in cumulative ordered Logit Models. Sociological Methods & Research. 51(2), 667-698.

Yismaw et al. (2019). Incomplete childhood vaccination and associated factors among children aged 12–23 months in Gondar city administration,Northwest, Ethiopia 2018. *BMC Research Notes* , 12-241.

APPENDIX

Table A.1: Maximum likelihood parameter estimation of PPOM

	Not vaccinated				Partially vaccinated			
Variables	Coef.	Std.Err	P >|z|	95%CI	Coef.	Std.Err	P >|z|	95% CI
Age (15 – 19 = reference)								
20 – 24	0.468	0.138	0.001*	[.198 .738]	0.468	0.138	0.001*	[.199 .739]
25 – 29	0.765	0.156	0.000*	[.459 1.07]	0.436	0.156	0.005	[.130 .741]
30 – 34	0.936	0.177	0.000*	[.589 1.283	0.258	0.173	0.134	[-.0799 .596]
35 – 39	1.218	0.196	0.000*	[.834 1.603]	0.015	0.191	0.938	[-.359 .388]
40 – 44	1.349	0.239	0.000*	[.879 1.819]	0.317	0.222	0.153	[-.118 .752]
45 – 49	2.582	0.493	0.000*	[1.615 3.548]	0.220	0.304	0.469	[-.379 .817]
Region (Tigray = reference)								
Afar	-1.399	0.170	0.000*	[-1.73 -1.066]	-1.399	0.170	0.000*	[-1.73 -1.066]
Amhara	-0.628	0.187	0.001*	[-.996 -.261]	-0.011	0.143	0.938	[-.291 .269]
Oromia	-0.761	0.176	0.000*	[-1.107 -.416]	-0.394	0.159	0.014*	[-.707 -.080]
Somali	-1.085	0.175	0.000*	[-1.428 -.742]	-1.085	0.175	0.000*	[-1.428 -.742]
Benshangul	-0.535	0.195	0.006*	[-.916 -.153]	-0.155	0.158	0.324	[-.464 .153]
SNNPR	-1.013	0.177	0.000*	[-1.361 -.665]	-0.464	0.165	0.005*	[-.788 -.139]
Gambela	-0.487	0.170	0.004*	[-.821 -.154]	-0.487	0.170	0.004*	[-.821 -.154]
Harari	-1.044	0.2095	0.000*	[-1.454 -.633]	-0.469	0.181	0.009*	[-.823 -.115]
Addis Ababa	0.082	0.180	0.648	[-.271 .436]	0.082	0.180	0.648	[-.271 .436]
Dire Dawa	0.092	0.169	0.585	[-.239 .425]	0.092	0.169	0.585	[-.239 .425]
Residence (urban = reference)								
Rural	-0.289	0.092	0.002*	[-.471 -.109]	-0.289	.092	0.002*	[-.471 -.109]
Religion (orthodox = reference)								
Catholic	0.158	0.3880	0.683	[-.602 .919]	0.158	0.388	0.683	[-.602 .919]
Protestant	-0.225	0.110	0.041*	[-.442 -.009]	-0.225	0.110	0.041*	[-.442 -.009]
Muslim	-0.065	0.097	0.505	[-.259 .126]	-0.065	0.097	0.505	[-.256 .126]
Traditional/othe	-0.602	0.252	0.017*	[-1.096 -.108]	-0.602	0.252	0.017*	[-1.096 -.108]
Wealth index(poor = reference)								

Middle	0.174	0.117	0.138	[-.056 .404]	-0.112	0.109	0.305	[-.325 .102]
Rich	0.571	0.118	0.000*	[.341 .801]	-0.058	0.098	0.554	[-.251 .134]
Place of delivery(at home = reference)								
At GOV health	0.399	0.073	0.000*	[.256 .543]	0.399	0.073	0.000*	[.256 .543]
At private health	0.458	0.195	0.019*	[.076 .841]	0.458	0.195	0.019*	[.076 .842]
At NGO health	0.332	0.253	0.190	[-.164 .828]	0.332	0.253	0.190	[-.164 .828]
Others	0.087	0.251	0.728	[-.405 .579]	0.087	0.251	0.728	[-.405 .579]
Place ANC(at home = reference)								
At health facility	-1.071	0.152	0.000*	[-1.368 -.774]	1.306	0.113	0.000*	[1.085 1.527]
Num. ANCV(no visit = reference)								
1 – 3	1.069	0.159	0.000*	[.756 1.382]	-0.351	0.141	0.013*	[-.628 -.074]
4 – 5	1.359	0.170	0.000*	[1.025 1.692]	-0.156	0.141	0.267	[-.433 .119]
6 and above	1.518	0.260	0.000*	[1.007 2.028]	-0.380	0.179	0.034	[-.732 -.029]
Birth order(first birth = reference)								
2 – 4 birth	-0.436	0.119	0.000*	[-.668 -.203]	0.060	0.099	0.545	[-.135 .255]
5 and above	-0.745	0.159	0.000*	[-1.058 -.433]	0.159	0.146	0.276	[-.127 .446]
Number of living children(1 – 3 child = reference)								
4 – 6 child	0.099	0.086	0.250	[-.069 .268]	0.099	0.086	0.250	[-.069 .268]
7 and above	0.057	0.124	0.647	[-.186 .299]	0.057	0.124	0.647	[-.186 .299]
Cons. (α_1 & α_2)	1.880	0.214	0.000*	[1.457 2.294]	-1.915	0.216	0.000*	[-2.34 -1.492]

Source; own computing based on 2019 EMDHS data

Note * implies variables that are significant at 5% level of significance

Table A. 2: Results of the empty model with random intercept

Parameter	Coef.	Std.Error	Z	p>|Z|	95% CI
Fixed part					
$\alpha 1(\leq notvacinated)$	-1.889	0.198	-9.57	0.000	[-2.277 -1.503]
$\alpha 2(\leq partially\ vaccinated)$	1.321	0.196	6.72	0.000	[0.936 1.705]
Random part					
Between region variance	0.412	0.179	2.29	0.000	[0.175 0.968]
LR test vs. oligit. Chibar2(01) = 476.59 Prob >= chibar2 = 0.0000					
Log likelihood = -5024.103					
AIC = 10054.21 BIC = 10074.18 ICC = (0.412)/(0.412+3.29) = 0.111					

Source; own computing using 2019 EMDHS children record data.

Table A.3: Test of multi-collinearity [Correlation Matrix]

	Age	Reg	Res	EDL	Rel	WI	PDL	PANC	NANCV	No_LC	BORDR
Age	1.0000										
Reg	-0.0203	1.0000									
Res	0.0171	-0.3130	1.0000								
EDL	-0.2531	0.2621	-0.3743	1.0000							
Rel	-0.0600	0.1723	0.0648	-0.1694	1.0000						
WI	-0.0548	0.2842	-0.5498	0.4426	-0.1963	1.0000					
PDL	-0.0297	0.1962	-0.3211	0.3342	-0.1090	0.3474	1.0000				
PANC	-0.0500	0.0710	-0.1842	0.2453	-0.2373	0.2928	0.2843	1.0000			
NANCV	-0.0240	0.1185	-0.3200	0.3542	-0.2643	0.4090	0.3670	0.6320	1.0000		
No_LC	0.5623	-0.0859	0.1679	-0.3593	0.1456	-0.2196	-0.1703	-0.1954	-0.1922	1.0000	
BORDR	0.6315	-0.0705	0.1736	-0.3855	0.1189	-0.2229	-0.1601	-0.0874	-0.1481	0.6866	1.0000