GEOSPATIAL ANALYSIS OF ZERO DOSE AND UNDER IMMUNIZED CHILDREN IN MALAWI: 2015-2016 MALAWI DEMOGRAPHIC HEALTH SURVEY (MDHS)

MSc (BIOSTATISTICS) THESIS

JOACHIM NYIRONGO

UNIVERSITY OF MALAWI

MARCH, 2025

GEOSPATIAL ANALYSIS OF ZERO DOSE AND UNDER IMMUNISATION CHILDREN IN MALAWI: 2015-2016 MALAWI DEMOGRAPHIC HEALTH SURVEY (MDHS)

MSc (BIOSTATISTICS) THESIS

By

JOACHIM NYIRONGO
BSc (Mathematical Sciences Education-Statistics and Computing)- University of Malawi, The Polytechnic

Submitted to the Department of Mathematical Sciences, School of Natural & Applied Sciences in partial fulfillment of the requirements for the Master of Science (Biostatistics)

UNIVERSITY OF MALAWI
MARCH, 2025

DECLARATION

I, the undersigned hereby declare that this thesis is my original work and has not been submitted to any other institution for the award of degree. Where works of others have been used, due acknowledgement have been used. Every effort has been made to be as accurate as possible and any errors contained herein are my sole responsibility.

JOACHIM NYIRONGO____

Full Legal Name

____

Signature

____

Date

CERTIFICATE OF APPROVAL

I, the undersigned certify that this thesis is student’s own work and effort, and due acknowledgement have been made for any other source of information used. This thesis is submitted with my approval.

Signature:_______________________________**Date:**________________________

Lawrence Kazembe, PhD (Professor)

Supervisor

DEDICATION

This thesis is dedicated to my beloved parents, brothers and sisters who have always been there for me and loved me throughout my work.

To my late mother, Getrude Nzima, I know you could have been proud to see the milestone I have reached, may your soul continue resting in peace.

To Joachim Nyirongo you have always thrived, continue working hard and aiming higher to achieve your goals

ACKNOWLEDGEMENTS

I am truly grateful to the almighty God, the Giver and Sustainer of life, who has seen me through all the time during my study period.

I am also greatly indebted to my supervisor Professor Lawrence Kazembe, whose constructive and insightful has been a core ingredient to successful completion of this study.

Due credit should go to my Dad, Mr Levison Nyirongo who has been the source of encouragement to further my studies. His desire for education is unmatched. It would not be complete if I fail to thank my support system beginning with my mother Queen Lungu, my brothers (Gift, Austin, Joseph, Wisdom and Peter), my sisters (Staphel, Ireen and Stayness), my classmate: Innocent Gondwe, Jeromot Masoambeta, Tony Matimati, Boswell Munthali, Potiphar Damiano and my friends: Twika Mwalwanda, Tryness Chaula, John Billon, Stalin Shaba, Judith Ngwala, Saina Muyaka, Neuman Munthali, Japheth Chimaliro, Mwawi Chirwa and Tawonga Nyasulu for various support they have always rendered towards my study. You were always there for me.

In special way I would like to thank Mr Gentry Kamanga, the Executive Director for Good Health Organisation who was my previous boss for allowing me to do school while working. It could have been difficult to complete this journey without his support.

Thanks to all my classmates and friends too numerous to list, for encouragement and warmth. May God richly bless you.

ABSTRACT

Despite Malawi's progress in immunization, with 89% of children receiving DPT1 and 84% receiving DPT3, gaps persist due to geographical disparities, leaving many children unprotected against vaccine-preventable diseases. Zero-dose and under-immunized children remain highly vulnerable, contributing to preventable disease burdens and hindering progress toward Immunization Agenda 2030 and SDG target 3.3. This study models the spatial variability of these children. The study adopted a quantitative cross-sectional design using secondary data from the 2015-2016 Malawi Demographic and Health Survey (MDHS). It focused on children aged 12 to 35 months, with zero-dose and under-immunized status as the outcome variables. Key covariates included maternal age, ethnicity, religion, region, residence type, birth order. Exploratory spatial data analysis (ESDA) was conducted to map the spatial distribution of zero-dose and under-immunized children and assess spatial autocorrelation. A copula-based spatial modeling approach using Generalized Joint Regression Modeling (GJRM) was applied to examine the joint spatial dependency of immunization indicators and associated risk factors. Markov Random Field (MRF) smoothing was used for district-level analysis. Of the 5668 children aged 12-35 months, 2.32% were zero dose and 6.72% were under immunized. The joint modelling revealed that children from poorest families, those children from homes where parents have no education background were more likely to be zero dose and under immunized. On contrary children from the northern region and from rural type of residence are more likely to be non-zero dose and non under immunized children. The spatially data revealed districts with more proportions of zero dose and under immunized children which included: Nsanje, Chikwawa, Thyolo, and Phalombe. The joint modelling further revealed strong positive correlation between zero dose and under immunized. Applying copula provided a flexible approach to the study in identifying areas that need major intervention for scaling up immunization status. The presented maps and the analysis approach demonstrated a mechanism for monitoring zero dose and under immunized children in Malawi..

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBRIVIATIONS

AIC	Akaike Information Criterion
ANC	Antenatal care
BCG	Bacillus Calmette-Guerin
CAR	Conditional Autoregressive Model
CDF	Cummulative Density Function
CI	Confidence Interval
COVID-19	Corona virus 19
DPT	Diphtheria Pertusis Tetanus
EDA	Exploratory Data Analysis
EPI	Expanded Programme Immunisation
ESDA	Exploratory Spatial Data Analysis
GLM	Generalised Linear Model
GLMM	Generalised Linear Mixed Model
GJRM	Generalised Joint Regression Model
IA2030	Immunisation Agenda (2030)
INLA	Integrated Nested Laplace Approximation
LISA	Local Indicator Spatial Association
LMM	Linear Mixed Model
MCV	Measles Containing Vaccine
MCMC	Markov Conte Monte Carlo
MDHS	Malawi Demographic Health Survey
MPHC	Malawi Population Housing Census
MRF	Markov Random Field
OPV	Oral Polio Vaccine
SDG	Sustainable Development Goals
SEA	Standard Enumeration Area
TB	Tuberculosis
UNICEF	United Nationa Children Education Fund
WHO	World Health Organisation

CHAPTER ONE

INTRODUCTION

1.1 Background

Vaccination is the action of receiving the vaccine whereas the immunization is the process by which an individual acquires immunity from infectious diseases following vaccination. Vaccine reduces risks of getting a disease by working with the body’s natural defenses to build protection (WHO, 2024). Vaccine have been instrumental in eradicating and controlling many infectious diseases that pose great threat in the global health of human kind. For instance, vaccines have played a crucial role in the eradication of polio and the control of measles, significantly reducing global morbidity and mortality rates (UNICEF, 2024). Immunization also helps to protect vulnerable populations such as infants and the elderly who are at higher risk for severe disease outcomes. Lack of vaccination exposes these groups of people into deadly diseases such as polio, measles and pneumonia.

Immunization is a global health success story, saving 3.5-4 million lives every year from vaccine preventable diseases (WHO, 2024; CDC, 2024). Immunization services are core part of primary health care and the most cost effective public health intervention that significantly contribute to the reduction childhood morbidity and disability from vaccine preventable diseases (WHO, 2024). This follows the introduction of the Expanded Programme on Immunization (EPI), in 1974, by World Health Organisation (WHO). The EPI programme was introduced in Malawi by the WHO, in 1979, with the aim of reducing infant morbidity and mortality rates due to vaccine preventable diseases by providing quality immunisation (WHO, 2024). The current Malawi EPI defines basic vaccination as one dose of BCG, three doses of DTP-HemB-Hib, three doses of oral polio, two doses of rotavirus vaccine, and three doses of pneumococcal vaccine provided at 6, 10 and 14 weeks of age and one dose of measles vaccine at 9 months of age (UNICEF, 2022). The government of Malawi and its partners have played a significant role in funding the immunisation programme in which immunisation services are provided through both outreach and static clinics.

This is emphasized in the global Immunization Agenda 2030 (IA2030). IA2030 aims to make a world where everyone, everywhere, at every age fully benefits from the vaccines for good health and well-being (WHO, 2020). IA2030 operationalizes the Sustainable Development Goals (SGD) target 3.3, which puts a great emphasis on equity with shared commitment of leaving no one behind in terms of immunization (WHO, 2017).

Despite these successes, global immunization coverage has remained at 84% with no significant changes over the years despite the uptake of new and under-used vaccines increasing over the past few years (WHO, 2024). In 2023 alone, worldwide, 22 million infants did not receive full immunization (WHO, 2024). Since 2019 there has been a drastic increase in the number of zero dose children and under-immunised children globally due to: increased economic crisis, political crisis, conflicts and decline in vaccine confidence. WHO has defined zero dose children as those children who did not receive DPT1 and under immunized as infants who have not received the third dose of DTP-containing vaccine (DTP3) (WHO, 2024, GAVI, 2024). Corona Virus-19 (COVID-19) disruptions did exacerbate the already stalled condition (Basu, et al., 2023). According to UNICEF, COVID-19 pandemic resulted into the disruption of the third dose of Diphtheria Tetanus Pertusis (DPT3) vaccine often used to assess how countries are doing in providing routine immunization services to children. The global pandemic negatively affected the uptake of immunization in many countries. Statistical estimates on DPT3 coverage indicates the decrease from 86% in 2015-2019 to 83% in 2021 (UNICEF, 2023). DPT is usually used as a tracker for routine vaccination uptake. The increase in number of zero dose children and under-immunised is worrisome as these children are at an increased risk of infections from vaccine preventable diseases such as polio, measles and pneumonia during their first 5 years of life. They are also part of the world’s vulnerable population more likely to suffer from disability and mortality that could be easily prevented through immunization (Ozigbu, Olatosi, Li, Hardin, & Hair, 2022). Several studies have found that routine vaccination remains a problem globally. For instance, Alegana, Ticha et al (2024) found that there are factors contributing to under immunization which include mothers primary education, level of the education of the mother and house hold wealth (Alagana, Ticha, & Mwenda, 2024). Bile et al (2024) on Assessing Vaccination Delivery Strategies for Zero-Dose and Under-Immunized Children in the Fragile Context of Somalia found that despite the contextual heterogeneity of the population groups, the lack of targeted, population-specific strategies and meaningful engagement of local communities in the planning and implementation of immunization services is problematic in effectively reaching zero-dose and under immunized children (Bile, et al., 2024).

Malawi is one the countries that is burdened by the inadequate of the vaccine coverage in the world. The lack of timely and complete immunization exposes children to preventable diseases which results into putting the children's health at risk which in turn hinder the national progress in achieving Sustainable Development Goals (SDG3.3). Furthermore, there is inequity in immunization coverage in Malawi. Equity is a measure of difference in outcomes or impacts due to a difference based on social and economic exposures like gender, wealth, location and education status. Huge inequities exist between regions with highest prevalence of zero dose children and under immunized in hard to reach areas but inequities also exist within regions. For instance, Alegana et al (2024) on their study on modelling the spatial variability and uncertainty for under-vaccination and zero-dose children in fragile settings found that there was a variation at fine spatial scale in the population 12–23 months a national mean with the highest proportion in the South-East district of Laremanaye. Therefore, identifying vulnerable populations spatially detailed data are required beyond large area of summaries reported from household survey.

1.2 Problem Statement

Malawi has made notable progress in immunization, with one of the most successful Expanded Programs on Immunization (EPI) in the African region. However, despite these efforts, gaps in vaccine coverage persist. The 2022 vaccination report indicates that while 89% of children aged 12 to 23 months received DPT1, only 84% received DPT3 (WHO, 2023). This means that 11% of children did not receive any vaccines (zero-dose), and an additional 16% did not complete the vaccination schedule (under-immunized), leaving them vulnerable to vaccine-preventable diseases and deaths**.**

The presence of these missed children signals geographical and socio-demographic disparities in immunization coverage, yet limited studies have examined where these children are located and the factors contributing to their exclusion. Without targeted strategies, Malawi risks falling short of its Immunization Agenda (IA) 2030 and

Sustainable Development Goal (SDG) target 3.3, which aim to reduce child mortality through universal immunization coverage. Identifying the spatial variability and risk factors associated with zero-dose and under-immunized children is crucial to designing evidence-based interventions that bridge the immunization gap and strengthen Malawi's immunization efforts.

1.3 The main objective of the study

The main objective of the study is to model the spatial variability of the zero dose and under-immunized children in Malawi, where are they and their profile.

1.3.1 Specific objectives

- To model the spatial distribution of the zero dose and the under immunized in Malawi
- To assess the risk factors associated with zero dose and under immunized in Malawi
- To quantify spatial correlation or dependency between zero-dose and under-immunized in Malawi

1.4 Significance of the study

The findings of the research will not only aid in identifying the most affected areas of the zero dose children and under-immunized children in Malawi but will also contribute to the growing body of knowledge on immunization practices in resource limited settings. Ultimately, the results will inform policy makers and public health officials to enable them to formulate evidence based strategies to enhance vaccine uptake and protect the health and well-being of children in Malawi.

This research is also paramount in achieving an equitable and inclusive immunization coverage in the country where no child is left vulnerable to vaccine preventable diseases in line with Immunisation Agenda 2030 and also Sustainable Development Goal number 3.3. By working towards eliminating zero dose phenomenon, Malawi can take a significant step forward in securing a healthier future for the youngest citizen.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview of childhood vaccination in Malawi

Child immunization remains a key preventive health priority area for the Government of Malawi. The history of child immunisation in Malawi started in 1979 when the country launched and began to administer vaccines against diseases (WHO, 2024). At the onset, the target was against six diseases called a dose of Bacillus Calmette–Guérin (BCG) vaccine against tuberculosis (TB), DTP1, DTP2 and DTP3 vaccine against diphtheria, tetanus and pertussis, three doses of oral polio vaccine (OPV) against polio, and one dose of measles-containing vaccine (MCV) against measles (UNICEF MALAWI, 2020). Numerous other vaccines have been introduced and administered in the country including pneumococcal conjugative vaccine, monovalent rotavirus vaccine, malaria vaccine, oral cholera vaccines (UNICEF MALAWI, 2020). For years, Malawi has sustained a high coverage of immunization, with well above 80% of the population having received the six basic vaccinations. This has led to reduced burden of infectious diseases, especially in under five children (Mmanga, Mwenyenkulu, Nkoka, & Ntenda, 2021). In addition to the routine immunization, supplemental immunization activities, popularly known as mass vaccination campaigns, have contributed to high immunization rate of Malawian children, especially against measles. Since 1996, periodic immunization campaigns targeting polio and measles have been successfully conducted in accordance with WHO recommendations (UNICEF MALAWI, 2020). These campaigns relished massive gain in an average vaccination coverage of 95% (UNICEF MALAWI, 2020). Despite Malawi sustaining a high vaccination coverage, the zero dose children still exist. These zero-dose children are subjected to life threatening diseases that are preventable through vaccination.

2.2 Overview of the zero dose children

Existing literature on zero dose children globally indicates the prevalence of zero dose children exist mostly in middle and low income countries in the world regardless of the massive campaign on universal immunisation. In a study that was conducted by Edson Utazi et al (2023) on assessing the characteristics of un- and under-vaccinated children in low- and middle income countries in a cross sectional study where they fitted Bayesian multi-level logistic regression in the geospatial dataset found that remoteness increases the odds of un or under vaccinated in all study areas (Utazi, et al., 2023). According to Baptiste et al (2023) found that most zero dose children and under vaccinated children are more prevalent in deprived socially group (Baptiste, et al., 2024). Another study conducted by Wigley et al (2022) on Estimates on the number and distribution of the zero dose and under immunized children across remote rural urban and conflict affected settings in low and middle income countries found that coverage varies widely among and within countries with substantial heterogeneity existing in the context and spatial distribution of those un vaccinated (Wigley, et al., 2022) The study further found that over 40% of the zero dose children are living in remote-rural and urban settings and nearly 20% in the conflict setting.

Geospatial analyses have now gained traction as vital tool for creating high resolution and district level maps of health and demographic indicators (Jabbour & Attal, 2020). In the case of vaccination coverage, these maps are integrated with relevant gridded population data to produce estimates of numbers of zero dose and under-vaccinated at various spatial scales thus helping in identifying and delineating clusters of vulnerabilities within the country and better allocation of the resources (Tamir, Workneh, Mekonen, & Zegeye, 2024).

2.3 Types of Spatial Data

Spatial data also known as geospatial data refers to the information that identifies the geographical location and the characteristics of natural or constructed features and boundaries of the earth (Moraga, 2023). The application of this data is often represented in terms of the Cartesian coordinates (x, y). The concept of spatial data is integral to a variety of applications that require an understanding of how different elements relate to each other within the geographical space. Spatial data can be thought of as a resulting from observation of stochastic process $\{Z(s): s \epsilon D \complement R^d\}$ Where $Z(s)$ denotes the observed attributes at location s, R^d represent the d-dimensional of the Euclidean space and D is the spatial domain.

2.3.1 Lattice Data (Areal)

In Lattice or Areal data the domain D is fixed countable collection of regular or irregular areal units at which variables are observed. Areal or Lattice data usually arise when the number of events corresponding to some variable of interest are aggregated in areas. This data is used to understand geographical pattern and in identifying factors of risk taking into account the neighborhood configuration.

Lattice data may also arise in remote sensing application where satellites provides information on the number of variables such as social and economic characteristics of the population as well as household amenities and housing circumstances.

2.3.2 Geostatistical Data

In geostatistical data, the domain D for a given problem is continuous. This means that $Z(s)$ can be observed at any point $s \epsilon D$ and infinite points can be observed between two given points $s_i \ and \ s_j$. The continuity and discontinuity of Z depends on the nature of Z and not on the nature of the domain D. Since the nature of the domain is continuous, it cannot be sampled for an infinite time.

2.3.3 Point Pattern Data

A point data set is a discrete unit of information that is typically derived from measurement and can be represented numerically or graphically. This means that the spatial point process is a d-dimensional random pattern of points (where $d = 2$ or $d = 3$ in applications) (Dimaggio, 2012).

Point processes are assumed to follow a probability distribution and they are frequently described as Poisson processes. Spatial points can be associated with covariates where analyses may include assessing the role of covariates in finding intensity or controlling for covariates effects when assessing the interaction between points (Morris, et al., 2019). As a result, a point process is used for the analysis of observed point patterns, where the points denote the locations of some research object, such as disease cases.

2.4 Exploratory Spatial Data Analysis

Exploratory Spatial Data Analysis (ESDA) is the sub-discipline of the general Exploratory Data Analysis (EDA) with primary focus on the presence of the spatial effects (Guo, 2023). ESDA can be defined as a set of techniques that describe and visualize spatial distribution, identify outliers, find distributions patterns, clusters and hotspots and suggest regimes or other forms of spatial heterogeneity (Abelairas-Etxebarria & Astorkiza, 2020). This section outlines the common methods used in ESDA.

2.4.1 Choropleth Map

Choropleth map is a visual representation of the distribution of the variables. These kind of map divide an event, space generally a geographical area into smaller regions or divisions to provide data distribution per region (Schabenberger & Gotway, 2022). These smaller divisions maybe regular or irregular. The range of variables in particular divisions is generally thematically represented using color codes or variations.

2.4.2 Visualising Outliers and Extreme values

Outliers are points in the dataset that have values that stray too far from the rest of the data. Visualising and analyzing outlier points is absolutely essential to any statistical analysis for instance in determining the range of data and some anomalies or misinformed data which needs to be re-evaluated or discarded.

2.4.3 Global Spatial Autocorrelation

2.4.3.1 Joint Count Statistics

This is the simplest form of spatial autocorrelation because it uses binary labels to categorize quantitative values (Anselin & Li, 2019). At a nominal or ordinal levels, presence or absence of characteristics or range of features determine whether the joint corresponds to a similar or different values from neighboring values to determine the spatial relation.

The spatial dependence of data can be expressed in terms of the type of distribution which can be either grouped, random or dispersed.

2.4.3.2 Global Moran's I and Geary's C

This is the common and reliable method that is commonly used in statistics to measure global spatial autocorrelation. These methods they measure the linear correlation between a variable at one point in a division coordinate system a domain and the weighted average of the divisions. The test of I and C statistics are based on the Z statistics and related distributions. Between I and C statistics Moran's I statistics attest more to global spatial relations while C statistics are more relevant for local spatial relations.

Moran's I can be expressed as:

$$I = \frac{n}{W} \frac{\sum_{i=0}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Geary's C is given by:

$$C = \frac{n-1}{W} \frac{\sum_{i}^{n} \sum_{j}^{n} w_{ij}(x_i - \bar{x})^2}{\sum_{i}(x_i - \bar{x})^2}$$

Where n the number of spatial units is, x_i is the value of the variable at location i, $\bar{x}$ is the mean of the variable and W_{ij} is the spatial weight between locations i and j. The values of Moran's I typically fall between -1 and +1. -1 means negative autocorrelation which implies that nearby locations tend to have dissimilar values; +1 means positive autocorrelation which indicated that similar values which tend to occur in adjacent areas. A value of Geary's c less than 1 indicates positive spatial autocorrelation, while a value larger than 1 points to negative spatial autocorrelation. To assess the statistical significance of these measures, Z-scores are computed to compare observed values against expected values under spatial randomness. The Z-statistic for Moran's I is given by:

$$Z_I = \frac{I-E(I)}{\sqrt{Var(I)}}$$

where $E(I) = -\frac{1}{n-1}$ represents the expected Moran's I under the null hypothesis of spatial randomness, and $Var(I)$ is its variance. Similarly, the Z-statistic for Geary's C is:

$$Z_C = \frac{C - E(C)}{\sqrt{Var(C)}}$$

where $E(C) = 1$ under spatial randomness. Large positive or negative Z-scores indicate significant deviation from the null hypothesis, suggesting strong spatial patterns in the data.

2.5 Spatial Regression Models

Apart from spatial autocorrelation, spatial regression models accounts for spatial dependence in the data to examine relationship between variables while considering spatial proximity.

2.5.1 Spatial Lag Model

This is a type of spatial regression model that accounts for spatial dependence by incorporating a spatially lagged dependent variable as an additional predictor. It assumes that the value of the dependent variable in a given location is influenced by the values of the same variable in neighboring locations. Mathematically this can be expressed as:

$$Y = \beta_0 + \beta_1 x_i + \rho w_i y + \varepsilon_i \qquad i = 1, \dots, n$$

where Y is the value of the dependent variable for case i, β_0 is the intercept, x_i is the predictor variable, w_i is a 1xn row vector representing the proximity of other cases to case i (coefficients sum to 1), y is a nx1 column vector of values of the dependent variable for all cases, and ρ is the lag coefficient. The model shows that the value of y_i at each location i is determined not only by x_i at that location, but also by the value of y for neighboring units (Louzada, Nascimento, & Egbon, 2021)

Combining geographical information on populations, locations of health care sites, and the movement of vaccinators offer insights into how efficient and equitable vaccination coverage is, and has great potential to improve immunisation delivery. For instance, Joseph et al (2020) used spatial modelling and logistic to quantify for accessibility to immunizing health facilities and identify its influence on immunization uptake in Kenya, found that travel time of more than 1 hour to the health facility were significantly associated with low immunization coverage in the univariate analysis for both fully immunized status and DPT3 vaccine. Children living more than 2 hours were significantly less likely to be fully immunized and DPT3 after adjusting for other confounding variables such house hold wealth, mother's highest education level, parity and urban/rural residence (Joseph, et al., 2020).

2.5.2 Spatial Error regression model

The model emerges from the presence of spatial dependence in the error term of the spatial unit and corresponding neighboring units. It occurs in the case that some variables influencing dependent variable value but excluding in the model correlate among spatial units.

$$y_i = \beta_0 + \beta_1 x_i + u$$

$$u = \lambda W u + \epsilon$$

Where y_i is the dependent variable, x_i represents the independent variables, β is the vector of coefficients, u is the spatially correlated error term, W is the spatial weights matrix, λ is the spatial autoregressive coefficient, indicating the extent of spatial correlation, ϵ is the independent and identically distributed error term.

2.5.3 Local Indicator Spatial Association (LISA)

At each observation local indicator of spatial association provides a signal of strong spatial clustering or related values surrounding that observation. The sum of LISA is proportional to global indicators. The following is the definition of local Moran's statistics.

$$I_i(d) = (x_i - \bar{x}) \sum_{j=1}^{n} W_{ij}(d)(x_j - \bar{x}), j \neq i$$

Where x_i and x_j denotes the proportions of cases at a district i and j respectively and W_{ij} denotes the spatial weight matrix based on the defined distance lags between districts i and j.

*2.5.4 Getis-Ord Gi**

The Getis-Ord Gi* statistic is a local spatial autocorrelation measure used to identify clusters of high (hot spots) and low (cold spots) values in spatial data (Songchitruksa & Zeng, 2010). Unlike global spatial statistics such as Moran's I, Gi* focuses on localized spatial dependence.

Mathematically, the Getis-Ord Gi* statistic is defined as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j}$$

Where x_j is the attribute value at location j and w_{ij} represents spatial weights. The standardized Z-score form is given by:

$$Z(G_i^*) = \frac{\sum_{j=1}^{n} w_{ij} x_j - \bar{X} W}{S_X \sqrt{\frac{W_i^2 - W_{ii}}{n-1}}}$$

A high positive Z-score indicates significant clustering of high values (hot spots), while a negative Z-score suggests clustering of low values (cold spots).

2.6 Spatial Mixture Copula Model

2.6.1 Copula

Copulas, are function that are used to describe the dependence between variables. They are a distribution functions such that the cumulative density function (CDF) can be written in terms of the variables and the copula itself (Abraj, Thompson, Mery, Wang, & You-Gan., 2021). They are joint cumulative distribution functions (CDFs) of two or more random variables with uniformly distributed margins on the interval of [0, 1]. Let Y_1 and Y_2 be continuous random variables with marginal CDFs F_1 andF_2, then there exist a unique copula C:$[0,1]^2$ such that,

$$C(u,v) = P[F_1(Y_1) \leq u, F_2(Y_2) \leq v],$$
$$= C\big(F_1(y_1), F_2(y_2)\big),$$

Where u and v are any selected quantiles on the interval [0,1]

There are various types of copula families that describes a variety of dependence structures. Commonly used bivariate copulas include Gaussian, student-t and Joe copulas (Nelsen, 2007).

2.6.2 Spatial Copulas of univariate spatial random fields

While a non-spatial copula models the dependence between two or more non-spatial continuous random variables, spatial copula describes the joint dependence of a univariate spatial variable at any two spatial locations.

Let $Z(x)$ denote an univariate second-order stationary spatial random field Z that is observed at dimensional location $x \epsilon X$ and let $X = (x_{1,}, x_2, \dots, x_n)$ be a set of the existing observed locations in the given spatial domain X. The spatial copula that describes the dependence of a variable at any two locations separated by distance h is,

$$C_h(u,v) = P[F(Z(x)) \leq u, F(Z(x+h)) \leq v],$$
$$= C_h\big(F(Z(x)), F(Z(x+h))\big),$$

Where F is the CDF of Z and is assumed to be the same at each location x.

2.6.3 *Mixture copula*

The mixture copula is a weighted linear combination of two or more copulas on the interval [0, 1]. Let C_1 and C_2 be bivariate copulas, then there exists a bivariate mixture copula $C^m : [0,1]^2 \rightarrow [0,1]$ such that,

$$C^m(u,v) = (1-w)C_1 + wC_2$$

where $0 \leq w \leq 1$.

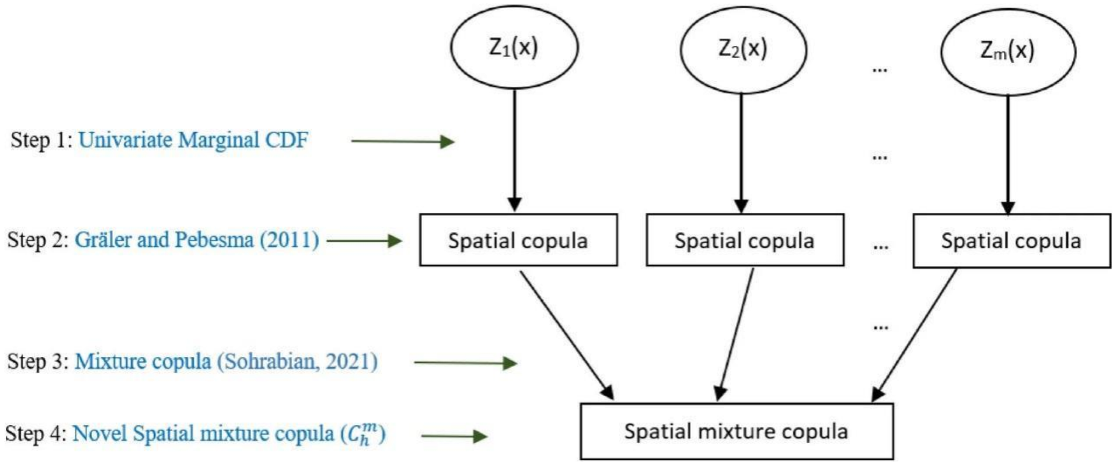
2.6.4 *Spatial mixture copula model*

The methodology for modelling spatially correlated multiple variables consists of two essential components: modelling each spatial variable separately using Graler and Pebesma's univariate spatial copula, then joining the univariate spatial copulas using the idea of mixture copula.

Let $Z(x) = [Z_1(x), Z_2(x), \dots, Z_m(x)$ be the second-order stationary multivariate spatial random field Z with m spatial variables that are sampled at the same two dimensional location $x \in X$ and let $X = (x_1, x_2, \dots, x_n)$ be the set of the existing locations in the spatial domain X.

A spatial copula describes the joint spatial dependence of a univariate spatial variable at any two spatial locations x and $x + h$, where h is the separation distance between two locations. Hence, spatial copula model dependence of one spatial location relative to another spatial location, rather than modelling dependence using absolute locations.

Fig 1: shows the basic steps for modelling:



Source: (Abraj, Wang, & Thompson, 2022)

Figure 1: A diagram for the spatial mixture copula construction

The detailed procedure for the model development is provided in the following stages

Stage 1: For each spatial variable $Z_l, l = 1, 2, \dots, m$, the marginal cumulative distribution functions (CDFs) such as Gamma, Weibull, Normal, Log-normal and obtain the best fitted CDFs.

Stage 2: Based on the distance between pairs, place each sample pair $\{F_l\big(Z_l(x), F_i(x + h)\big)\}$ into K equally spaced spatial bins as follows; $[0, h_1), [h_1, h_2), \dots, [h_{K-1}, h_K)$, where h_k is the cut off distance.

Given the pair of points for each spatial bin, spatial copula that describes the dependence of the spatial variable at any two locations can be calculated as:

$$C_{l,k,h}(u, v) = P\big[F_l\big(Z_l(x)\big) \le u, F_l\big(Z_l(x+h)\big) \le v\big],$$
$$= C_k(F_l(Z_l(x), F_l\big(Z_l(x+h)\big))$$

Where $k = 1, 2, \dots., K$ is the index of spatial bin, u and v are any selected quantiles of the corresponding univariate CDF of Z_l at location x and $x + h$.

The copula for each bin are selected using maximum log-likelihood values of competing copulas such as Gaussian, Student's t, Clayton, Frank, Gumbel and Joe. The maximum likelihood estimation (MLE) method is used to determine the parameter θ of each candidate copula by maximizing the log-likelihood function.

$$L(\theta) = \sum_{i=1}^{n} \log C_\theta\,(u_i, v_i)$$

Where $C_\theta(u,v)$ is the copula density function and u_i, v_i are transformed uniform marginal of the data. The copula with highest log-likelihood value is selected for each bin, ensuring that the best fit to the dependence structure in the data.

2.7 Generalised Linear Models (GLM)

Generalised Linear Models (GLMs) are class of regression models that are used to model a wide range of relationships between the response variable and one or more predictor variables (Agresti, 2015). The GLM improves on linear regression by allowing the linear model to be connected through the link function and the variance of each measurement is a function of its projected value (McCullagh, 1989)

In GLM, each dependent variable y is considered to be generated by specific distribution in an exponential family such as Normal, Binomial, Exponential and Gamma distribution where both mean μ of the distribution depends on the independent factors (Hardin, Hilbe, & Hilbe, 2007)

GLM extends standard linear regression models to encompass non-normal response distribution and possibly non-linear function of the mean. They have three components:

i. Random component: This specifies the response variable y and it probability distribution. The observation $y = (y_1, \dots, y_n)^T$ on that distribution are related as independent.

ii. Linear/Systematic predictor: For parameter vector $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ and nxp model matrix X that contains the value of p explanatory variables for n observations, the linear predictor is $X\beta$

iii. Link function: This is the function g applied to each component of the $E(y)$ that relates to the linear predictor.

GLM are traditionally formulated within the framework of the exponential family of distribution. The exponential family is written as:

$$f_y(y\colon \theta, \emptyset) = exp\left\{\frac{y\theta - b(\theta)}{a(\emptyset)} + C(y, \emptyset)\right\}$$

Where θ is the canonical (natural) parameter of the location and $\emptyset$ is the dispersion parameter (scale).

2.7.1 Generalised Linear Mixed Model (GLMM)

Generalized Linear Mixed model (GLMM) are extension of linear mixed model to allow a response variable from a different distribution such as binary responses. Alternatively, GLMM as an extension of GLM (e.g Logistic regression) to include both fixed and random effects (hence mixed model). The general form of the model (in matrix notation) is as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

Where $\boldsymbol{y}$ is an $Nx1$column vector of the outcome variable, $\boldsymbol{X}$ is an NxP matrix of the predictor variables; $\boldsymbol{\beta}$ is a $Px1$ column vector of the fixed effect of the regression coefficient (the βs**);** $\boldsymbol{Z}$ is the Nxq designed matrix of the random effects (the random complement of the fixed $\boldsymbol{X}$**),** $\boldsymbol{u}$ is a $qx1$ vector of the random effects (a random complement to the fixed) and $\boldsymbol{\varepsilon}$ is $Nx1$ column vector of the residuals the part of y that is not explained by $\boldsymbol{X\beta} + \boldsymbol{Zu}$**.**

The main difference with Linear Mixed Models (LMMs) is that GLMM response variables can come from different distributions besides Gaussian. In addition, rather than modelling the responses directly some link function is often applied as log link.

Let's consider a linear predictor:

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma}$$

The generic link function called $g(.)$. The link function relates the outcome y to the linear predictor η. Hence the conditional expectation of y (conditional because it is expected value depending on the level of predictors) as denoted as:

$$g(E(y) = \eta$$

The expectation of y can also be modelled as:

$$E(y) = g(\eta) = \mu$$

With $\quad y = g(\eta) + \varepsilon$

2.7.2 Bivariate Probit Models

Bivariate probit model is a joint model for two binary dependent variables whose distribution are assumed to be correlated. It generalises the index function model from one latent variable to two latent variables that maybe correlated.

Bivariate probit model maybe derived from a pair of regression or reduced from equations as:

$$y_1^* = X_1\beta_1 + \varepsilon_1.$$

$$y_2^* = X_2\beta_2 + \varepsilon_2.$$

Where y_1^* and y_2^* are latent variables. A latent variable is a variable that is incompletely observed (Seyoum, 2018). ε_1 and ε_2 are joint normal with mean zero, variance one and correlation ρ.

The bivariate probit model specifies the outcomes as:

$$y_1 = \begin{cases} 1 \; if \; y_1^* > 0 \\ 0 \; if \; y_1^* \leq 0 \end{cases}$$

$$y_2 = \begin{cases} 1 \; if \; y_2^* > 0 \\ 0 \; if \; y_2^* \leq 0 \end{cases}$$

The bivariate probit model can be written as

$$P(y_1 = i, y_2 = j) = \varphi_2(X_1'\beta_1, X_2'\beta_2, \rho)$$

2.7.3 Conditional Autoregressive Models (CAR)

Conditional Autoregressive Models (CAR) are often used in spatial data analysis. They are useful to obtain multivariate joint distribution of the random vector based on univariate conditional specification (Nobre, Alexandra, & Widemberg, 2018). These conditional specifications are based on Markovian properties such that conditional distribution of component of random vector depends only on a set of neighbors.

Typically, Gaussian CAR specification used latent structures in hierarchical models for real level data. Where the region of interest is divided into a set of disjoint areas and CAR random effect is used to account for possible correlation among observations made across the different areas.

Let y_i denote the observation that is associated with spatial units s_i for $i = 1,2,\dots,n$ and let x_i be a $px1$ vector that denotes values of p regressors recorded at unit s_i.

The conditional autoregressive model is defined as:

$$y(s_i)|Y_{(i)} \sim N(x(s_i)'\beta + \sum_{j=1}^{n} C(s_i, s_j)\left(y(s_j) - x(s_j)'\beta\right)\sigma^2(s_i))$$

Where $y(s_i)$ represents the observation at spatial unit s_i, $x(s_i)$ denotes the regressors at s_i, $C(s_i, s_j)$ represents the spatial dependency between locations s_i and s_j, $\sigma^2(s_i)$ is the variance associated with unit s_i.

The models are often stated in a hierarchical Bayesian framework with Markov Chain Monte Carlo (MCMC) simulation used for inference or Integrated Nested Laplace Approximation (INLA) (Lee & Mitchell, 2013).

2.8 Bayesian Hierarchical Model

The Bayesian approach provides samples of the entire posterior distribution of incidence rates or relative rates for each area by providing more information than a single point estimate where all parameters are allocated to cope with their likely volatility prior to distribution and this is achieved through MCMC and INLA.

Bayesian Analysis method of statistical inference that allows one to combine prior information about population parameter with evidence from information contained in a sample to guide the statistical inference processes.

This can be given by the following equation:

$$P(\theta|Y) \propto P(Y)P(Y|\theta)$$

Where $P(\theta|Y)$ is the posterior, $P(Y)$ is prior and $P(Y|\theta)$ is the likelihood.

Bayesian hierarchical models can be used to analyze data that arise when outcome variable is aggregated into areas that form a partition of the study regions. Models can be specified to describe the variability in the response variable as a function of a number of the covariates known to affect the outcome as well as random effects to model residual variations not explained by covariates.

CHAPTER THREE

METHODOLOGY

This chapter expands on the methods that has been used in the study. It expounds the statistical models for the analysis that has been used in the study.

3.1 Study Population

The study population of the participants consists of children aged 12 to 35 months which represents the youngest cohort of children who are supposed to take all vaccines recommended in the first year of life.

3.2 Data Source

The study uses secondary data from the 2015/16 Malawi's Demographic Health Survey (MDHS) as part of national surveys implemented by the National Statistical Office for Malawi over a 4-month period, from 19 October 2015 through 17 February 2016. The sampling frame used for the 2015/16 MDHS was the frame of the Malawi Population and Housing Census (MPHC), conducted in Malawi in 2008. The 2015/16 MDHS sample was stratified and selected in two stages. Each district was stratified into urban and rural areas. In the first stage, 850 standard enumeration areas (SEAs), including 173 SEAs in urban areas and 677 in rural areas, were selected with probability proportional to the SEA size and with independent selection in each sampling stratum. In the second stage of selection, a fixed number of 30 households per urban cluster and 33 per rural cluster were selected with an equal probability systematic selection from the newly created household listing.

3.3 Study Variables

3.3.1 Dependent Variables

In order to model the spatial distribution of the zero dose children, a binary variable was created. Zero dose was created from a variable "ever received DPT1" where response “No” and “I don’t know” were categorized as 1 whereas “vaccination date on the card”, “Reported by mother”, “Vaccination marked on the card” were categorized as 0. Similarly, under immunized was created from a variable “ever received DPT3” where response “No” and “I don’t know” were categorized as 1 whereas “vaccination date on the card”, “Reported by mother”, “Vaccination marked on the card” were categorized as 0.

3.3.2 Independent Variables

Explanatory variables were selected based on the findings from the literature and the availability of data at individual and at a community level. These variables include: region, ethnicity, religion, education level, age of the respondents, wealth index and the type of residence.

3.4 Data management and analysis

After the data was accessed, data extraction, cleaning, weighting, and recoding were done before any statistical analysis. Descriptive statistics was done to describe the characteristics of the study population using R. R software were used to execute the spatial data analysis. To determine the presence of spatial autocorrelation, Global Moran’s Index (Moran’s I) was used. Moran’s I is a statistic that measures whether zero dose and under immunized patterns were clustered, dispersed, or randomly distributed in the study area by producing an output ranging from -1 to +1. Moran’s I value close to -1 indicated zero dose or under immunized children are dispersed, whereas Moran’s I +1 indicated zero dose or under immunized children are clustered, and zero dose or under immunized children are distributed randomly if Moran’s I value was zero.

3.4.1 Logistics regression

Logistics regression model is a statistical model used to estimate the effect of the factors when there are categorical responses.

Let Y_1 and Y_2 be binary response variables as Zero Dose and Under Immunised respectively. For binary response variables Y_1 and vector of explanatory variables X, the logistics regression model is given by:

$$log\frac{p(x)}{1-p(x)} = \beta_0 + x.\beta$$

Where $p(x)$ is the binomial proportion, x is the matrix of the covariates. Logistic distribution constrains the estimated probabilities that lie between 0 and 1. The estimated probability is given by:

$$\pi(X) = \frac{e^{\beta_{j0}+\beta_{j1}X_1+\cdots+\beta_{jp}X_p}}{1 + e^{\beta_{j0}+\beta_{j1}X_1+\cdots+\beta_{jp}X_p}} = \frac{e^{\beta_{jp}X_p}}{1 + e^{\beta_{jp}X_p}}$$

Where $\pi_j(X) = P(Y_{ji} = 1|X)$ is the probability is of either Zero Dose or Under-vaccinated given other covariates X.

The log odds that manifest linear relationship with explanatory variables can be expressed as:

$$\begin{aligned} logit(\pi_j|x) &= logit[P(Y_{ji} = 1|X)] \\ &= \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \cdots + \beta_{jp}X_p \\ &= X_P\beta_{jp} \end{aligned}$$

In logistics regression model, the most common types of measuring association between categorical variables is odds ratio. It is the ratio of odds defines as:

$$OR_j = \frac{odds\, J_1}{odds\, J_2} = \frac{\pi_j(X_1) \big/ 1 - \pi_j(X_1)}{\pi_j(X_2) \big/ 1 - \pi_j(X_2)}$$

An odds ratio that departs or varies from 1 refers as a less dependency between two categorical variables.

3.4.2 Goodness of fit

Prior to the fitting the model, checking the adequacy or goodness of fit to the model is needed. Wald test statistics could be used to measure the model quality of fit (named after Abraham Wald). The Wald test assesses statistical parameter constraints by calculating the weighted distance between the unrestricted estimate and its expected value under the null hypothesis, where the weight represents estimate precision.

The Wald test is denoted as follows:

$$W = \frac{\hat{\beta}_i - \hat{\beta}_0}{s\hat{e}\beta_i}$$

3.4.3 Akaike Information Criterion (AIC)

One of the most commonly used information criteria is AIC. The concept behind AIC is to select the model that minimizes the negative likelihood penalized by number of parameters. AIC is commonly used for model selection. In model selection, the AIC proposes the following criterion:

$$AIC = -2\log p(L) + 2p$$

Where L is the likelihood under the fitted model and p is the number of parameters in the models. Specifically AIC is used to identify the best approximating model to the unknown data generating data and its processes (Akaike, 1998)

3.5 Bivariate Probit Model

The bivariate probit model is a joint model for two binary dependent variables. These variables may be correlated with correlation ρ. Let y_1^* and y_2^* represent zero dose and under immunized respectively . The bivariate model can be written as:

$$P(y_1 = i, y_2 = j) = \varphi_2(X_1'\beta_1, X_2'\beta_2, \rho)$$

Where the coefficients (β_1, β_2 and ρ) can be estimated using the maximum likelihood estimation.

3.5.1 Marginal Effect

After the parameter coefficients were estimated, the marginal effects of the covariates in the conditional distribution were estimated (Green, 1996) Marginal effects determine the magnitude of change of the conditional probability of the outcome variable when one change the value of the regressor, holding all the regressor constant at some value (Njoroge, Mwalili, & Wanjoya, 2015). . The marginal effect for the bivariate probit model is given by:

$$\frac{\partial \Phi_2(X_1'\beta_1, X_2'\beta_2)}{\partial X_i} = \varphi(X_i'\beta_i)\Phi_2\left(\frac{X_2'\beta_2 - \rho X_1'\beta_1}{\sqrt{1-\rho^2}}\right)\beta_i, i = 1,2$$

The marginal effects for categorical variables shows how conditional probability changes as the categorical variables changes from 0 to 1, after controlling in some way for other variables in the model (Seyoum, 2018).

3.6 Spatial Autocorrelation testing

Several R packages were used in the analysis. Queen contiguity *nb* neighborhood was defined as adjacent district with adjacent weights $W_{ij} = 1$ if district i and j share a common boundary and $W_{ij} = 0$ otherwise. Moran's I was used to test the data for global Moran spatial autocorrelation. Zero dose and under immunized cluster were then identified using local moran and Getis and Ord spatial statistics. The proportion of zero dose and under immunized for each district were computed. Moran's I was calculated to determine whether there is spatial autocorrelation in the data for zero dose and under immunized clusters in Malawi.

3.7 Copula Spatial Models

To perform flexible modelling at the district level copula modelling was applied to account for spatial dependence and spatial distribution for zero dose and under immunized indicators.

Joint analysis of the immunization indicators was carried out to account for spatial dependence and three models were fitted and the best fitting model was selected based on the lowest AIC. The joint regression model used in this study will contribute to the methodology of joining discrete variables.

3.7.1 Kendall's Tau and Spearman Rho

Spearman rho and Kendall tau are two important correlation measures. For random variable X_1 and X_2, spearman rho is defined as:

$$\rho_s(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)$$

Spearmans is therefore simply the linear correlation of probabilities transformed random variables.

For random variables X_1 and X_2 kendall tau is defined as:

$$\rho_\tau(X_1, X_2) = E((X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$$

Where $(\bar{X}_1, \bar{X}_2)$ is independent of (X_1, X_2) but has the same joint distribution as (X_1, X_2)

CHAPTER FOUR

RESULTS AND FINDINGS

4.1 Social Demographic characteristics of the respondents

Table 1 gives a summary of the characteristics of the data used for the study. The study targeted children of ages 12 months to 35 months. A total of 5668 children were selected.

Table 1: Baseline table of the social demographic of the data.

	Total	**Zero dose**	**Under immunised**
	N=5,668	N=132	N=381
Region			
Northern	970	13 (1.34%)	47 (4.85%)
Central	2,015	33 (1.64%)	132 (6.55%)
Southern	2,683	86 (3.20%)	202 (7.53%)
Religion			
Catholic	835	15 (1.80%)	60 (7.19%)
CCAP	747	14 (1.87%)	40 (5.35%)
Anglican	248	3 (1.20%)	11 (4.44%)
Seventh Day	403	6 (1.49%)	21 (5.21%)
Other Christians	2,596	74 (2.85%)	197 (7.59%)
Muslims	805	19 (2.36%)	48 (5.96%)
No Religion	34	1 (2.94%)	4 (11.76%)
Ethnicity			
Chewa	1,742	28 (1.60%)	105 (6.03%)
Tumbuka	604	11 (1.82%)	42 (6.95%)
Lhomwe	972	29 (2.98%)	61 (6.28%)

Tonga	218	2 (0.92%)	13 (5.96%)
Yao	748	16 (2.14%)	40 (5.34%)
Sena	331	23 (6.95%)	51 (15.41%)
Nkhonde	83	3 (3.61%)	5 (6.02%)
Ngoni	726	13 (1.79%)	47 (6.47%)
Mang'anja	137	6 (4.38%)	14 (10.22%)
Nyanja	107	1 (0.93%)	3 (2.80%)
Education Level			
No Education	648	30 (4.63%)	68 (10.49%)
Primary	3,803	79 (2.08%)	250 (6.57%)
Secondary	1,118	21 (1.88%)	61 (5.46%)
Higher	99	2 (2.02%)	2 (2.02%)
Wealth Index			
Poorest	1,310	45 (3.44%)	118 (9.01%)
Poorer	1,239	22 (1.78%)	72 (5.81%)
Middle	1,124	21 (1.87%)	69 (6.14%)
Richer	1,004	22 (2.19%)	67 (6.67%)
Richest	991	22 (2.22%)	55 (5.55%)
Birth Order Number			
1	2,888	59 (2.04%)	171 (5.92%)
2-4	1,117	27 (2.42%)	76 (6.80%)
5-7	1,406	40 (2.84%)	113 (8.04%)
Above/equal to 8	257	6 (2.33%)	21 (8.17%)
Number ANC Visit			
Adequate visit	1,288	25 (1.94%)	91 (7.07%)
Inadequate visit	4,380	107 (2.44%)	290 (6.62%)
Maternal Age			
15-24	2,218	60 (2.71%)	182 (8.20%)
25-34	2,426	49 (2.02%)	127 (5.23%)
35-44	948	20 (2.11%)	62 (6.54%)
Above/equal 45	76	3 (3.95%)	10 (13.16%)
Exposure to Media			

No	3,309	90 (2.72%)	251 (7.59%)
Yes	2,324	41 (1.76%)	127 (5.46%)
Not a dejure Resident	35	1 (2.86%)	3 (8.57%)
Type of Residence			
Urban	901	28(3.11%)	67(7.44%)
Rural	4767	194(4.07)	314(6.59%)

Table 1 presents the distribution of the immunization status among children in Malawi based on various social-demographic factors. Of 132 children classified as zero dose, 13(1.34%) reside in the Northern Region while 86(3.20%) are from Southern Region indicating higher prevalence of zero dose status in the Southern Region compared to the Northern Region and Central Region. Among those classified as under immunized (N=381), Southern region shows higher percentage 202(7.53 %) compared to the Northern Region 47 (4.85 %) and Central Region 132(6.55 %). In terms of religion, those who were not affiliated to any religious beliefs showed a relatively higher percentage of the under immunized compared to the other religious groups. According to ethnic groups, the Sena showed a higher percentage of being zero dose (2.94 %) and of being under immunized (11.76 %) than any other ethnic groups. Mothers who had no education background were seen to have higher percentage of being zero dose (4.63%) and under immunized (10.49 %).

4.2 Mapping the spatial distribution of the zero- dose and under- immunised children in Malawi.

The first objective of the study was to model the spatial distribution of the zero dose and under immunised children. In order to model the distribution of the zero dose and under immunised, proportions were computed for each outcome variable in each district in R software package. Figure 2 presents the proportions of the zero dose and under immunized children in Malawi.

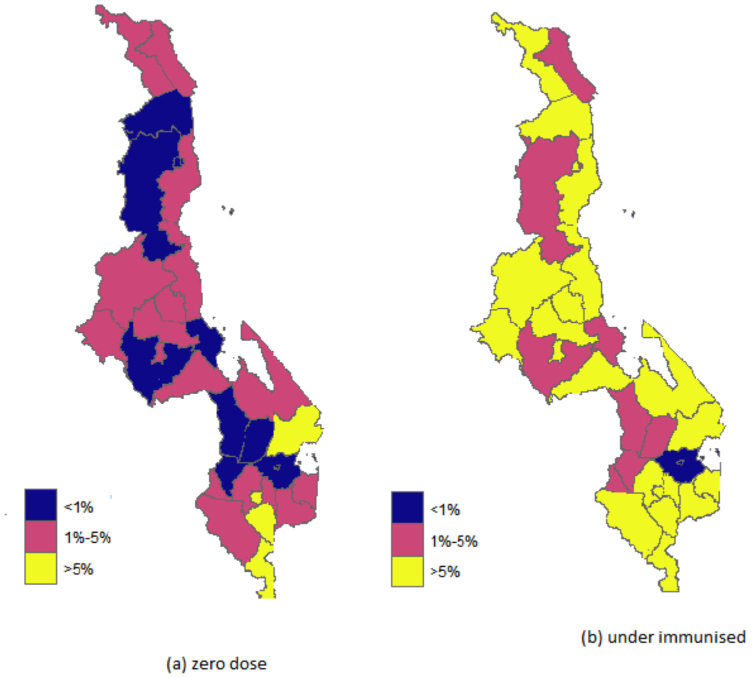


Figure 2: Proportions of the zero dose and under immunised Children in Malawi

The results (Figure 2) shows that Nsanje, Chikwawa, Thyolo, part of Blantyre and Machinga as district with higher proportions of zero dose in the southern region where as Kasungu and Chitipa districts have higher proportions of zero dose children in the central and northern region respectively.

On the other hand, the results (Figure 2b) shows that Nsanje, part of Blantyre, Mangochi have higher proportions of under immunized children in the southern part of Malawi, whereas as Nkhota-kota and Ntchisi have a higher proportion of under Immunised children in the central part of Malawi and Chitipa district in the Northern part of Malawi.

4.3 Spatial Autocorrelation Analysis

4.3.1 Global Moran's Analysis

Global Moran's is the most powerful and widely used spatial autocorrelation tool for determining the overall strength of the spatial dependence in data. At $\alpha = 0.05$ Global Moran's was used to test the null hypothesis H_0 of no significance clustering of zero dose and under immunized within the districts. The Moran's I statistics for the 2015-2016 MDHS data indicate a positive spatial autocorrelation for both zero-dose and under-immunized children, suggesting that these immunization statuses are not randomly distributed across Malawi. The Moran's I value for zero-dose was 0.285 (p =

0.001, variance = 0.012), indicating a moderate and statistically significant clustering, where districts with high zero-dose rates tend to be surrounded by similar districts. For under-immunized children, the Moran's I value was 0.167 (p = 0.04, variance = 0.013), showing a weaker but still significant spatial clustering. As a result, the findings encourage one to continue using Local Moran's to determine where high or low clusters occurs. Figure 3 presents the Global Moran's I for both zero dose and under immunized children.

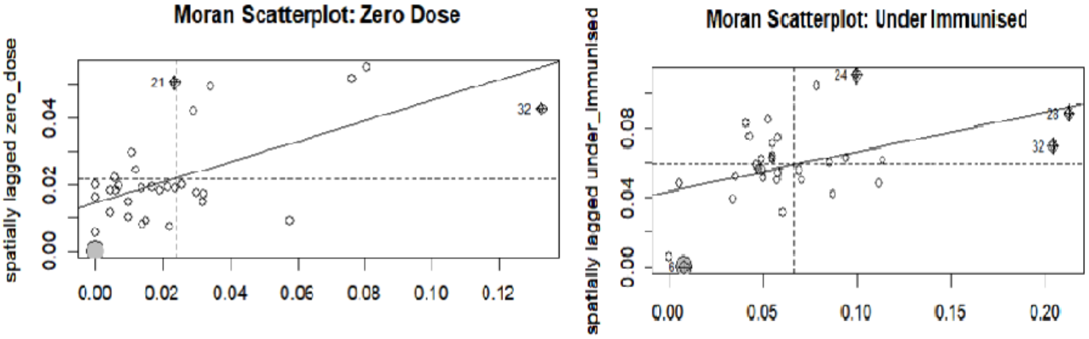


Figure 3: Moran I (zero-dose and under immunised)

Fig 3: The Moran I scatterplots for zero-dose and under-immunized children illustrate the spatial autocorrelation of these variables. Both plots display a positive slope, indicating a positive spatial correlation where areas with high zero-dose or under-immunized rates tend to be surrounded by similar areas. The presence of quadrant-based clustering suggests spatial heterogeneity, with some districts showing higher influence (e.g., districts 21 and 32). The scatterplots confirm that vaccine uptake disparities exhibit spatial patterns, highlighting potential clusters of low immunization coverage that may require targeted interventions.

4.3.2 Local Moran I Analysis

A Local Moran's I analysis was conducted to identify and assess the spatial autocorrelation of the zero dose and under-immunised proportions across the districts in Malawi. The local Moran's I statistics provide an insight into clustering similar values across the space which can reveal significant hotspots and cold spots within the data. Figure 4 presents the local Moran's I of the zero dose and under immunized children in Malawi.

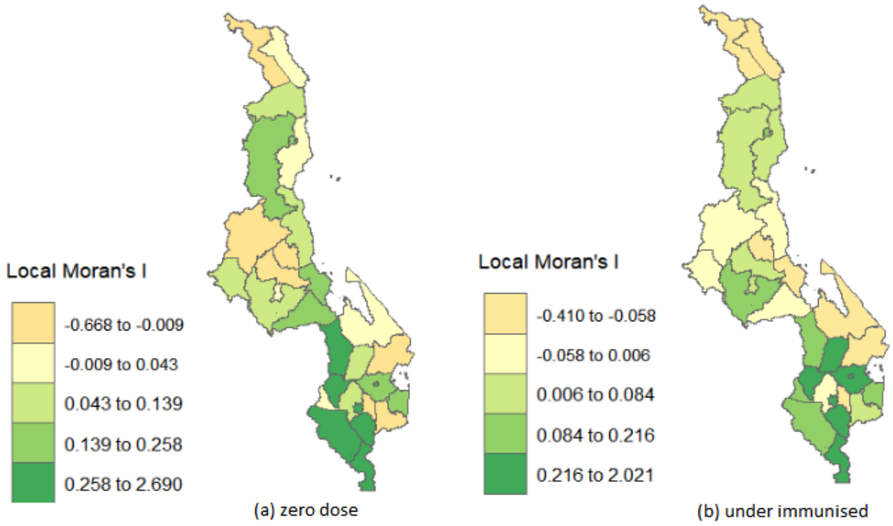
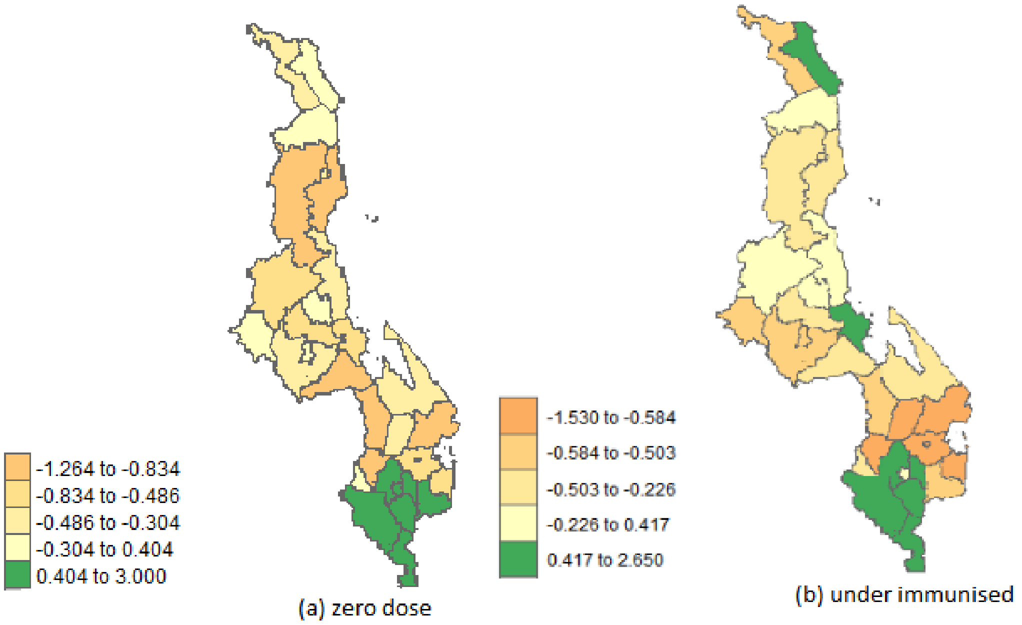


Figure 4: Map of Local Moran's Statistics for zero dose and under immunised

The Local Moran's map (Figure 4(a) and 4(b)) indicates distinct spatial patterns of the zero dose proportions and under immunised proportions across the districts in Malawi. Districts shaded dark green indicate districts with significant high local moran's values. These districts exhibit high levels of the zero dose and under-immunised surrounded by other districts with similarly values. The concentration of high values suggests the presence of spatial clusters "hotspots districts" for both zero dose and under-immunised mostly located in the southern region of Malawi.

4.3.3 Getis ord-Gi Statistics*

In the study, Getis ord-Gi* statistics was employed to identify spatial clusters of high and low values of zero dose and under immunized across district in Malawi. Getis ord-Gi* is widely used method for detecting spatial clusters (hotspots and coldspots) of high and low values within geographic data. This allows the researcher to identify districts with significantly higher and lower concentration of zero dose and under immunized. Figure 6 shows Getis ord-Gi* Statistics for both zero dose and under immunized children in Malawi.



(a) zero dose

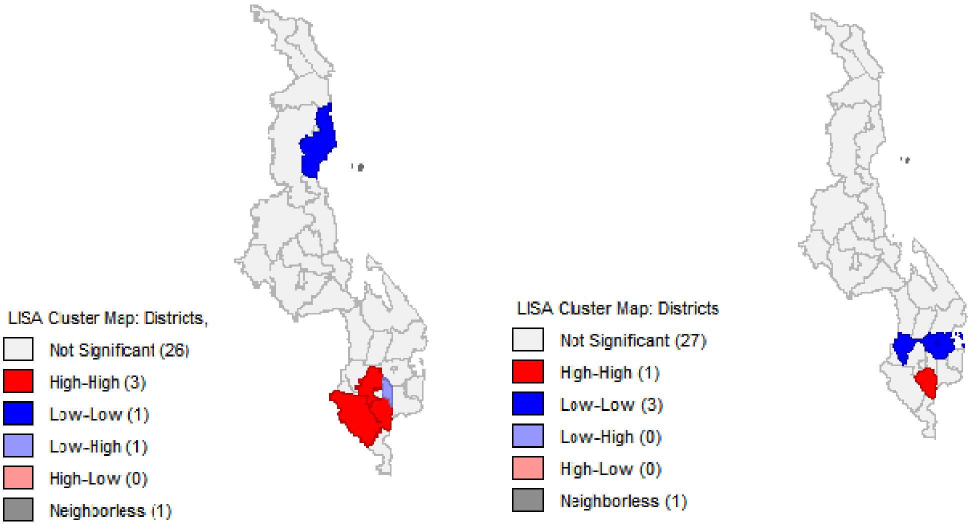
(b) under immunised

Figure 5: Map of the Getis ord Gi* for zero dose and under-immunised Proportions

The results (Figure 5(a) and 5(b)) shows the existence of spatial clustering of zero dose and under immunized children across districts in Malawi. The maps are color coded based on Gi* which quantify the degree of clustering. District with high positive Gi* are shaded in green indicating “hot spots” where there is significant clustering of high percentages of zero dose and under immunized children. The results therefore show that Nsanje, Chikwawa, Blantyre, Thyolo, Phalombe and Chiradzulu as hotspots districts of zero dose children whereas Nsanje, Chikwawa, Blantyre, Salima, Thyolo and Chiradzulu as the hotspot districts for under immunized.

4.3.4 Spatial Analysis of zero dose and under immunised using LISA cluster map

Local Indicator Spatial Autocorrelation (LISA) cluster maps were generated to identify and visualize spatial autocorrelation in the distribution of the zero dose and under immunized. In figure7 indicates LISA maps for zero dose and under-immunised children in Malawi.



(a) zero dose

(b) under immunised

Figure 6: LISA cluster map for zero dose and under-immunised Children

LISA analysis revealed significant spatial clustering of the zero dose children across Malawi (Figure 6a). Three district Thyolo, Chikwawa and part of Blantyre in the southern region exhibited high high clustering indicating areas where the prevalence of zero dose children is significantly higher than expected. In contrast Nkhatabay district in the northern region exhibited low low clustering suggesting success vaccination efforts. LISA map for under immunized children (figure 6b) reveal Thyolo as a high high clusters. However, Zomba district was identified with low low indicating low levels of under immunized.

4.4 Assessing the risk factors associated with zero dose and under immunised

The second objective of the study was to assess risk factors associated with zero dose and under immunized in Malawi. In order to assess the risk factors, the use of copula was adopted in the research due to their ability to model the dependence structure between two random variables. Copulas allows the separation of marginal distributions from dependence structure, making them suitable for modelling complex, non-linear relationships (Durante & Sempi, 2010). Several copula models were fitted in R statistical software to capture different types of dependencies and the following models were considered: Gaussian copula was chosen due to its simplicity and ability to model symmetric dependence, Clayton copula was selected due to the lower tail dependence to capture strong relationships between covariates when both are experiencing low outcomes and Gumbel copula was selected due to upper tail dependence to capture the tendency of extreme high outcome (Haugh, 2016).

4.4.1 Generalised Joint Regression Model

4.4.1.1 Model Fitting

The study applied copula regression to joint two immunization outcomes, quantify the determinants of zero dose and under immunized and capture spatial dependency. The Markov Random Field (MRF) smoother was applied to model the spatial dependency based on geographical location of the survey respondents. The variable DIST_CODE was fitted as an unstructured random effect. Based on the 2015-2016 MDHS there were 32 geographic districts. Spatial pattern suggests that areal observation close to each other have similar values than those that are far from each other. Thus

$$N_{ij} = \begin{cases} 1 & if\ i\ and\ j\ are\ neighbors \\ 0 & otherwise \end{cases}$$

The continuous variable Maternal Age applied the smooth function using the regression spline approach by Eilen and Marx (1996). These splines are used to model non-linear effects for variables. In under immunized the variable Maternal Age was fitted as categorical variable because of the linear effect of the variable.

Model fitting used a semiParBIV function in the GJRM package by Marra and Radice (2024) in R to fit the model (Radice & Marra, 2024). A semiParBIV fits a flexible bivariate binary model with several types of covariates effects. One joint model was fitted for the immunization outcomes. The model was with spatial effects to capture district spatial variation using penalized likelihood estimation. The non-linear effect was included in the model to capture non-linearity for Maternal Age in zero dose outcome.

The immunization outcomes in the bivariate were assumed to be correlated as unrelated regression model. Consider a bivariate model and let y be binary indicator at location i. The bivariate joint model is given by:

$$y_i = Z_\beta + s(nonlinear) + s(spatial)$$

where $Z_\beta$the fixed effects fitted and s is the smooth function applied.

The following equations of GJRM were fitted in the study:

$$\log(zero\ dose) = y_{1i}^T\alpha_{11} + s(Maternal\_Age) + s(DIST\_CODE) + \varepsilon$$

$$\log(under\ immunised) = y_{2i}^T\alpha_{21} + s(DIST\_CODE) + \varepsilon$$

Where y_{1i} and y_{2i} are categorical variables, α_{11} and α_{21} represents the effects of the variables in y_i while s_u for $u = 1,2,3\ and\ n$ as smooth function of Maternal Age in zero dose and DIST CODE for districts and ε is the error term.

The variables included in y_{1i} and y_{2i} were wealth index, maternal age, education level, region, type of residence, birth order number, number of the ANC visits, religion and ethnicity. Upon fitting the three copula models, the best fitted model was selected that had lower AIC value as shown in Table 2.

Table 2: Comparison of copula models for zero dose and Under Immunised

Model	**AIC**
Gaussian Copula	3334.17
Gumbel Copula	3347.36
Clayton Copula	3347.36

Model fitted aggregated data for individual and context factors. The resulting models is illustrated in the table 2 found that the Bivariate Spatial Copula model for Gaussian to be the best model since it had lower AIC (3334.17).

4.4.1.2 Joint Model Results

This section presents results of joint spatial model of the immunization outcomes that were fitted to capture variation and assess dependency.

Table 3: Results from GJRM with spatial effects fixed and non-linear effect of district to capture spatial variation and spline to model non-linear effects of maternal age in zero dose

	Zero dose			Under immunised		
Covariates	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
(Intercept)	-1.9	-2.6, -1.2	<0.001	-0.67	-1.4, 0.03	0.06
Ethnicity						
Chewa (Ref)						
Tumbuka	0.25	-0.16, 0.66	0.2	0.52	0.25, 0.80	<0.001

Lhomwe	0.15	-0.12, 0.42	0.3	-0.01	-0.21, 0.19	>0.9
Tonga	0.14	-0.37, 0.66	0.6	0.38	0.04, 0.72	0.029
Yao	0.04	-0.31, 0.39	0.8	-0.11	-0.37, 0.14	0.4
Sena	0.49	0.19, 0.80	0.001	0.48	0.25, 0.70	<0.001
Nkhonde	0.43	-0.22, 1.1	0.2	0.44	-0.06, 0.95	0.087
Ngoni	-0.05	-0.33, 0.24	0.7	0.05	-0.13, 0.23	0.6
Mang'anja	0.29	-0.15, 0.73	0.2	0.25	-0.08, 0.58	0.14
Nyanja	-0.14	-0.90, 0.62	0.7	-0.24	-0.79, 0.31	0.4
Religion						
Catholic (Ref)						
CCAP	0.08	-0.22, 0.38	0.6	-0.08	-0.28, 0.12	0.4
Anglican	0.09	-0.35, 0.54	0.7	-0.12	-0.44, 0.20	0.5
Seventh Day	-0.06	-0.42, 0.30	0.7	-0.16	-0.41, 0.08	0.2
Other Christians	0.16	-0.07, 0.39	0.2	-0.03	-0.18, 0.12	0.7
Muslims	0.05	-0.30, 0.41	0.8	-0.06	-0.31, 0.19	0.7
No Religion	0.11	-0.77, 0.98	0.8	0.12	-0.46, 0.69	0.7
Education Level						
No Education (Ref)						
Primary	-0.3	-0.50, -0.09	0.005	-0.23	-0.39, -0.07	0.004
Secondary	-0.36	-0.65, -0.07	0.015	-0.32	-0.54, -0.11	0.003
Higher	-0.38	-1.1, 0.35	0.3	-0.92	-1.6, -0.22	0.01
Wealth Index						
Poorest (Ref)						
Poorer	-0.21	-0.42, 0.01	0.059	-0.18	-0.33, -0.03	0.022
Middle	-0.25	-0.48, -0.02	0.034	-0.14	-0.30, 0.02	0.083
Richer	-0.13	-0.36, 0.11	0.3	-0.09	-0.26, 0.08	0.3
Richest	-0.15	-0.47, 0.17	0.3	-0.15	-0.38, 0.08	0.2
Region						
Northern (Ref)						

Central	0.24	-0.14, 0.63	0.2	0.48	0.21, 0.75	<0.001
Southern	0.05	-0.46, 0.56	0.8	0.48	0.19, 0.76	<0.001
Type of Residence						
Urban (Ref)						
Rural	-0.24	-0.47, -0.01	0.043	-0.2	-0.37, -0.03	0.022
Birth Order number						
>=8 (Ref)						
1	0.1	-0.41, 0.61	0.7	-0.21	-0.57, 0.15	0.3
2-4	0.06	-0.38, 0.50	0.8	-0.21	-0.51, 0.10	0.2
5-7	0.13	-0.27, 0.53	0.5	-0.04	-0.31, 0.23	0.8
Number ANC Visit						
Adequate visit (Ref)						
Inadequate visit	0.11	-0.07, 0.29	0.2	-0.05	-0.18, 0.07	0.4
Exposure to Media						
No (Ref)						
Yes	-0.11	-0.28, 0.07	0.2	-0.1	-0.22, 0.02	0.11
Not a dejure Resident	-0.08	-1.0, 0.86	0.9	0.08	-0.55, 0.72	0.8
Maternal Age						
15-24				-0.59	-1.2, 0.02	0.058
25-34				-0.65	-1.2, -0.14	0.013
35-44				-0.43	-0.84, -0.01	0.045
Above/Equal 45 (Ref)						

The copula model indicates significant association between ethnicity, maternal education level, region and they type of residence for both vaccination outcomes (zero dose and under immunised). On ethnicity, Tumbuka ethnic group had higher likelihood of being under immunized (Beta=0.52, p-value=0.001) compared to the reference group Chewa. Similarly, Sena ethnic group showed significantly higher probability of being both zero dose (Beta=0.49, p-value=0.001) and under immunized (Beta=0.48, p-value<0.001) than the reference group Chewa. Additionally, Tonga ethnic group indicates significantly higher of being under immunized (Beta=0.38, p-value=0.029) than the Chewa ethnic group. The results further indicate that lower education levels were associated with higher probabilities of zero dose children. Children whose mothers had primary education had lower likelihood of being zero dose (Beta=-0.33, p-value=0.0005) compared to those with no education background. In terms of region, children from Southern regions were more likely to be both zero dose (Beta=0.05, p-value=0.8) and under immunized (Beta=0.48, p-value<0.0001) compared to those from the Northern Region. In the figure 7 shows the spatial distribution of the zero dose vaccination probabilities by districts.

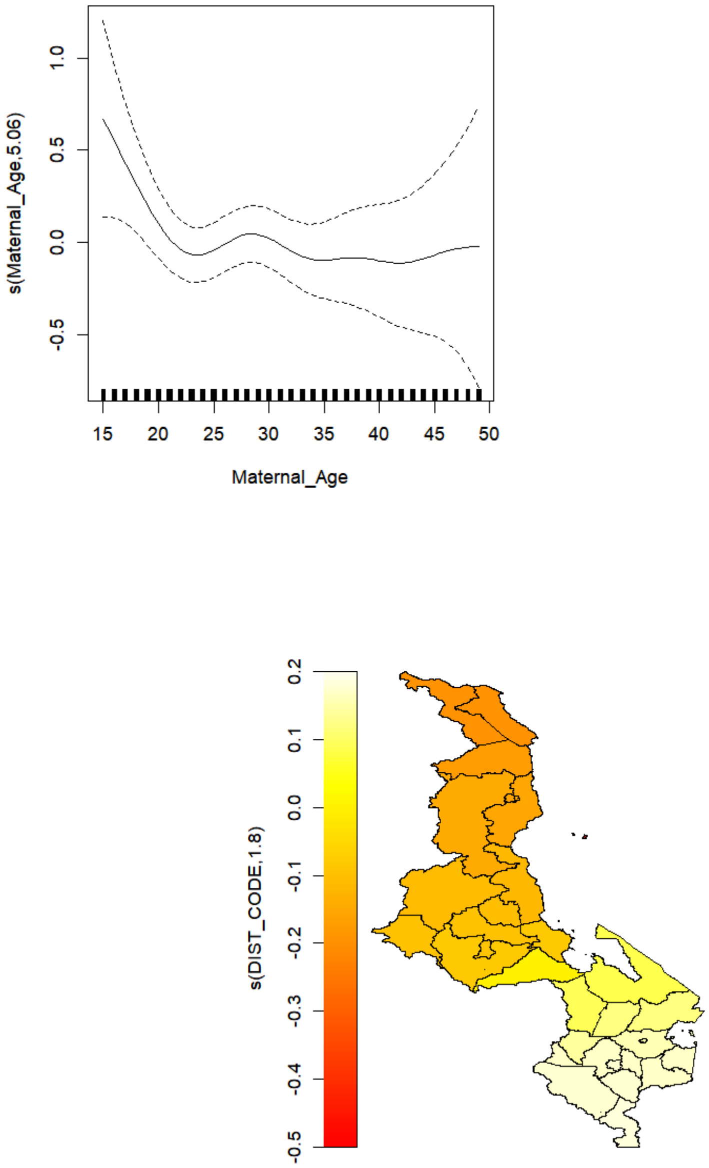


Figure 7: The smooth function Maternal Age and the [1]district spatial variation of zero dose. See Appendix for names of the districts.

The spatial map in figure 7 reveals a higher concentration of the Zero Dose children in the Southern Region. This aligns with significant spatial variance term from the model. In contrast, Northern Region show lower probabilities of zero dose children. In terms of age, the figure shows that mothers who were aged below 25 had higher chances of having zero dose children, the decrease from age 25 to 40 indicate the decrease in the

---

[1] DIST_CODE represents the district in Malawi

likelihood of having zero dose children and mothers above 40 years were less likely to have zero dose children. Figure 8 illustrate the predicted estimates of the under immunized children in Malawi.

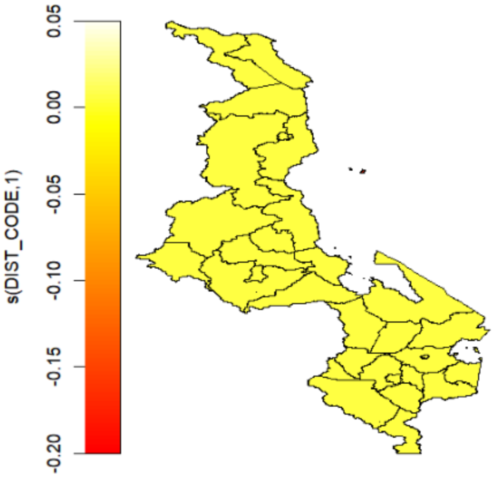


Figure 8: District spatial variation of under immunized. See Annex for District names.

Figure 8 indicates that there are small spatial variations on under immunized across the districts in Malawi, but the effects are relatively minor. Certain districts slightly influence under immunized negatively.

The estimated correlations and confidence interval CIs for the fitted model with spatial, fixed and non-linear effects are reported in the table 4. The results show that there exists strong positive dependency between zero dose and under immunized as shown in the table.

Table 4: Theta and Kendall's Tau results from model fitted with spatial, fixed and non-linear effects

Parameter	**Estimates**	**95% CI**
Theta(θ)	0.931	(0.895, 0.954)
Kendall's Tau (τ)	0.762	(0.706, 0.807)

The estimated correlation between zero dose and under immunized children has Theta(θ) value of 0.931(95% CI: 0.895, 0.954). This indicates a strong positive spatial dependence between zero dose and under immunized. This implies that areas with high levels of zero dose are likely to have high cases of under immunized. The Kendall's Tau (τ) value was 0.762(95% CI: 0.706, 0.807) indicating a strong positive spatial correlation between zero dose and under immunized children in Malawi.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATION

Immunization Agenda 2030 has the target of reducing the number of zero dose children substantially by reaching high and equitable coverage levels by 2030 (IA2030, 2020). Despite many continued efforts both national and international, zero dose and under immunized remains a concern in many developing countries. This study aimed at applying joint copula to model zero dose and under immunized to identify spatial variations at district level in Malawi. The specific objectives of the study included: modelling the spatial distribution of the zero dose and under immunized, assessing the risk factors of the zero dose and under immunized and quantifying for spatial dependence between zero dose and under immunized children in Malawi. The subsections in this chapter covers the results in detail.

5.1 Discussion of the Results

The study investigated the determinants and spatial patterns of zero dose and under immnised children in Malawi. The distribution of percentages of zero dose and under immunized, local morans maps, global morans, Getis and LISA maps and the estimates from the joint model to account for spatial dependence were the method used in the study. With regards with zero dose, the result of the study indicated that northern region had the lowest percentages of zero dose children compared to central and southern region. A similar pattern was also observed for under immunized where northern region had the lowest percentage of under immunized children.

The distribution of the immunization indicators is important as it help to understand the relationship between indicators and the covariates. This relationship is also essential to the government in designing effective policies and intervention to tackle the problem. The findings further demonstrate that immunization status is not solely a function of geography but is also influenced by socioeconomic and demographic factors. The model fitted with fixed effect showed that children from Sena ethnic groups, those with mothers having only primary or secondary education, and those from rural areas were significantly more likely to be zero-dose. Similarly, the Tumbuka, Tonga, and Sena ethnic groups were significantly associated with under-immunization, as were families in the central and southern regions. These patterns suggest that cultural beliefs, education levels and regional disparities play a crucial role in immunization uptake. Studies by Taneja, Das, Ray and Lasesson (2023), Mohamoud (2024) pointed out that maternal education and care givers is strongly associated with immunization coverage, which is similar to this study (Taneja, et al., 2023). Mohamoud (2024) also found that parents who are rich and richest were less likely to have zero dose children which is in line to this study (Mohamoud, et al., 2024). Studies by Biks, Dawit, Tsegaye and Wondie (2024) found that mothers and care givers from younger age (15-24) were more likely to have zero dose children which is consistent to the findings of this study (Biks, et al., 2024). The current study has found that mothers who had inadequate visit to the ANC follow up services were more likely to have zero dose children which attest to the study findings by Ozigbu (2023). On contrary to previous studies, this study found that zero dose children in Malawi were more in urban setting as compared to rural. A study by Ntenda (2019) on factors associated with non- and under- vaccination among children aged 12-23 months in Malawi using multinomial analysis of the population based sample found that geographical regions was the determinants of the immunization (Ntenda, 2019). Notably, this study found that Region emerged as the major determinant of the zero dose and under immunized as Southern Region specifically in Nsanje, Chikwawa, Thyolo and Chiradzulo as hotspot districts which were being disproportionately affected with zero dose and under immunised which aligned with the study by Ntenda (2019).

5.2 Spatial dependence and correlation of zero dose and under immunized-its implications

The estimated value of the theta θ is 0.931 (95%CI: 0.895, 0.954) (Table 4) suggesting a strong dependence between zero dose and under immunized children in Malawi. This implies that the district with high zero dose tend to have higher cases of under immunized children in Malawi. In addition, the value of Kendall tau is estimated to be at 0.762(95% CI: 0.706, 0.807) (Table 4) which measures the rank correlation between zero dose and under immunized. This value demonstrates moderately strong positive correlation between zero dose and under immunized children suggesting that zero dose cases are highly correlated with under immunized cases. The findings of the study aligns with the a study by Jean-Baptiste, Wagai, Hahne, Adeniaran (2024) and Vidal Fuertes, Johns, Goodman, Munro and Hosseinpoor (2022) who also found that there exists a strong correlation between zero dose and under immunized (Vidal, et al., 2022). These findings demonstrate the need for targeted intervention and policies which would help to scale up immunization status in Malawi as the intervention for one outcome is likely to affect the other outcome.

5.3 Policy Implications

The findings of this study have significant implications for public health policy in Malawi. Firstly, improving access to immunization services in underserved areas is critical for reducing the number of zero dose children and under immunized children. This can be achieved by strengthening health care and increasing outreach programs. There is also a need for health education programs that target communities with low literacy rates, emphasizing the importance of vaccination and dispelling the myths in ensuring that all children regardless of the background have equal access to immunization services.

Lastly the spatial analysis suggests that a one size fits all approach may not be effective in addressing regional disparities instead a localized strategy that focus on districts with highest concentration of zero dose and under immunized children are necessary to ensure equitable health outcomes across Malawi.

5.4 Limitation of the study

The limitation of this study was using secondary data as the researcher had no control over the variables in the dataset. Another limitation was that the data might not be of quality representation because the data was collected with a different purpose. The sample from the urban setting were significantly lower compared to those in the rural setting it is therefore important to be vigilant when interpreting the predicted estimates. For example, the model predicted that families from rural setting had lower chances of being zero dose than those from urban which may not be true on the ground.

5.5 Conclusion

The study highlights social demographic and economic factors such as region, religion, education level, wealth index and type of residence as critical factors for influencing whether a child is zero dose or under immunized. The findings suggest that immunization coverage is not only the function of geographical disparities but is also linked with social and economic factors. Furthermore, the study demonstrates significant spatial dependence between zero dose and under-immunised in Malawi. The findings from copula models suggest that improving immunization coverage should consider geographical clustering of these issues.

Overall the research highlights the importance of addressing immunization gap in geographically targeted manner, recognizing areas with high rates of zero dose children are likely to have broader immunization challenges.

5.6 Recommendations

Based on the findings, the following recommendations are made:

1. Develop targeted health communication strategies that are sensitive to regional, education level and religion. For instances in areas where some religious belief may hinder immunization coverage, engaging with traditional leaders and religious leaders using culturally appropriate messaging may increase vaccine uptake
2. Invest in education programs that raise awareness on the importance of immunization particularly among women and care givers with lower levels of education. Empowering them with knowledge can lead to better decision for their children
3. Public health interventions should focus on the regions with high spatial dependence of zero dose as the impact is directly going to affect under immunized aswell.

5.7 Suggested further studies

i. Assess how health care infrastructure and service delivery influence vaccination coverage

ii. Examine disparities in vaccination access among marginalized vulnerable groups, using geospatial approach.

REFERENCES

Abelairas-Etxebarria, P., & Astorkiza, I. (2020). From Exploratory Data Analysis to Exploratory Spatial Data Analysis. . *Mathematics and Statistics, 8(2), 82 - 86.* http://doi.org/:10.13189/ms.2020.080202.

Abraj, M., Thompson, Mery, Wang, & You-Gan. (2021). Spatial Mixture Copula Model For Multiple Correlated Responses With An Environmental Application. http://doi.org/*10.21203/rs.3.rs-1169352/v1.*

Abraj, M., Wang, Y., & Thompson, M. (2022). A new mixture copula model for spatially correlated multiple variables with an environmental application. *Sci Rep.,* 12(13867). https://doi.org/10.1038/s41598-022-18007-z

Adelle, W, L. J. (2022). Estimates of the number and distribution of the zero dose and under immunised children across remote rural, urban and conflict affected setting in low and middel income countries. *PLOS Glob Public Health*, 2(10), e0001126. https://doi.org/10.1371/journal.pgph.0001126

Agresti, H. (2015). *Foundations of Linear and Generalised Linear Models.* John Wiley & Sons.

Akaike, H. (1998). *Information theory and an extension of the maximum likelihood principle. In Selected papers of hirotugu akaike* . Springer.

Alagana, V., Ticha, J., & Mwenda, J. (2024). Modelling the Spatial variability and uncertainty for under vaccination and zero dose children in fragile setting. *Sci Rep 14, 24405*. doi:10.1038/541598-024-74982-5

Anselin, L., & Li, X. (2019). Operational local join count statistics for cluster detection. . *Journal of Geographical Systems, 21(1), 189-210.*

Baptiste, J., Wagai, A., Hahné, J., Adeniran, Koko, A., de Vos, R. I., . . . Hak, E. (2024). High resolution Geospatial mapping of zero dose and underimmunised children following Nigeria 2021 multiple idicator cluster/national immunisation coverage survey. *The Journal of Infectious Diseases, 230(1), e131-e138.*

Basu, S., Ashok, G., R, D., S, R., P, L., & A, A. (2023, Dec 31). *Impact of COVID-19 pademic on routine vaccine landscape: Aglobal perspective.* http://doi.org/10.1080/21645515.2023.2199656

Biks, G., Shiferie, F., Tsegaye, D., Asefa, W., Alemayehu, L., Wondie, T., . . . Gebremedhin, S. (2024). *High prevalence of zero-dose children in underserved and special setting populations in Ethiopia using a generalize estimating equation and concentration index analysis.* http://doi.org/10.21203/rs.3.rs-3034517/v1

Bile, A. S., Ali-Salad, M. A., Mahmoud, A. J., Singh, N. S., Abdelmagid, N., Sabahelzain, M. M., . . . Nor, B. (2024). Assessing Vaccination Delivery Strategies for Zero-Dose and Under-Immunized Children in the Fragile Context of Somalia. *Vaccines, 12(2).* http://doi.org/10.3390/vaccines12020154

CDC. (2024). *Global Immunization*. https://www.cdc.gov/global-immunization/index.html

Dimaggio, C. (2012). Spatial Epidemiology notes. *Appl Vignettes R, 20(2), 242-244.*

Durante, F., & Sempi, C. (2010, May). Copula theory: an introduction. In *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009 (pp. 3-31).* Springer

Eimias BE, A. A. (2022). *Vaccination status and factors associated among children aged 12 to 23 months in Ethiopia based on 2016 EDHS. Logit based multinomial logistic regression analysis.* https://doi.org/10.1371/journal.pone.0264004.

GAVI. (2021). *Reaching zero dose children* https://www.gavi.org/vaccineswork/reaching-zero-dose-children.

Green. (1996). Marginal effect in bivariate probit model. *NYU working paper #87.*

Guo, D. (2023). Exploratory Spatial Data Analysis. http://doi.org/10.1002/9781118786352.wbieg1001

Hardin, J. W., Hilbe, J. M., & Hilbe, J. (2007). *Generalized linear models and extensions.* . Stata press. College Station.

Haugh, M. (2016). *Quantitative Risk Management*. http://www.columbia.edu/~mh2078/QRM/copulas.pdf

IA2030. (2020). *A global Strategy of leaving no one behind.* https://doi.org/10.1016/j.vaccine.2022.11.042

Jabbour, S., & Attal, B. (2020). Geospatial analysis: a new frontier in humanitarian health research?. *The Lancet Global Health, 8(11), e1353-e1354.*

Joseph, N. K., Macharia, P. M., Ouma, P. O., Mumo, J., Jalang'o, R., Wagacha, P. W., . . . Okiro, E. A. (2020). Spatial Access inequities and childhood immunisation uptake in Kenya. *BMC Public Health, 20(1), 1-12.*

Lee, D., & Mitchell, R. (2013). Locally adaptive spatial smoothing using conditional auto-regressive models. *Applied Statistics, 62(4), 593-608.*

Louzada, F., Nascimento, D., & Egbon, O. (2021). Spatial Statistical Model; An overview under Bayesian Approach. *Axioms, 10(4).* https://doi.org/10.3390/axioms10040307

McCullagh, P. (1989). *Generalized Linear Models (2nd ed.). Routledge.* https://doi.org/10.1201/9780203753736

Mmanga, K., Mwenyenkulu, T. E., Nkoka, O., & Ntenda, P. A. (2021). Tracking Immunisation coverage, dropout and equity gaps among children ages 12-23 months in malawi-bottleneck analysis of the Malawi Demographic and Health Survey. *International Health, 14(3), 250-259.*

Mohamoud, S. A., Ali-Salad, M. A., Bile, A. S., Singh, N. S., Mahmud, A. J., & Nor, B. (2024). Determinants and prevalence of zero-dose children in Somalia: Analysis of the 2020 Health Demographic Survey data. *PLOS Global Public Health,* 4(7), e0002612.

Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R.* CRC Press.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., & DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. . *Spatial and spatio-temporal epidemiology, 31(1), 100301.*

National Statistical Office, ICF Macro. (2017). *Malawi Demographic and Health Survey 2015-2016*. Author

Nelsen, G. (2007). *An Introduction to copulas*. Springer.

Njoroge, T., Mwalili, S., & Wanjoya, A. (2015). *Bivariate probit model for correlated binary data with application to HIV and Male circumcision.* http://doi.org/10.11648/j.ajtas.20150406.27

Nobre, Alexandra, S., & Widemberg. (2018). *Conditional Autoregressive (CAR) Model*. https://doi.org/10.1002/9781118445112.stat08048

Ntenda, P. A. (2019). Factors associated with non-and under-vaccination among children aged 12–23 months in Malawi. A multinomial analysis of the population-based sample. *Pediatrics & Neonatology,* 60(6), 623-633.

Ozigbu, C. E., Olatosi, B., Li, Z., Hardin, J. W., & Hair, N. L. (2022). Correlates of zero dose vaccination status amaong children aged 12-59 months in sub saharan africa. *Vaccines, 10(7), 1052-67.*

Radice, G., & Marra, A. (2024). GJRM: *Generalised Joint Regression Modelling.* https://cran.r-project.org/web/packages/GJRM/GJRM.pdf

Schabenberger, O., & Gotway, C. A. (2022). *Statistical methods for spatial data analysis.* Chapman and Hall/CRC.y.

Seyoum, S. (2018). Analysis of prevalence of Malaria and Anemia using Bivariate Probit Model. *Ann. Data. Sci.*, 5(1), 301–312. https://doi.org/10.1007/s40745-018-0138-3

Songchitruksa, P., & Zeng, X. (2010). Getis-Ord Spatial Statistics to identify Hot spots by Using Incident Management Data. *Transportation Research Record Journal of the Transportation Research*, 6(2), 42-51.

Tamir, T. T., Workneh, B. S., Mekonen, E. G., & Zegeye, A. F. (2024). Coverage, spatial distribution and determinants of childhood inactivated poliovirus vaccine immunization in Ethiopia. . *Plos one, 19(5), e0301933.*

Taneja, G., Datta, E., Sapru, M., Johri, M., Singh, K., Jandu, H. S., . . . Dhawan, V. (2023). An equity analysis of zero-dose children in India using the national family health survey data: status, challenges, and next steps. *Cureus, 15(2), 123-34.*

UNICEF. (2022). *Expanded Programme on Immunisation (EPI) & COVID-19 Vaccination.* Author

UNICEF. (2023). *Zero dose: Children missing out life saving vaccines* https://data.unicef.org/resources/zero-dose-the-children-missing-out-on-life-saving-vaccine.

UNICEF. (2023, April). *Zero dose: The children missing out on life saving vaccines*. http://data.unicef.org/resources/zero-dose-the-children-missing-out-on-life-saving-vaccines/

UNICEF. (2024). *Eradicating Polio-UNICEF*. https://www.unicef.org/immunization/polio

UNICEF MALAWI. (2020). *2019/20 child immunisation budget brief upscaling investment in child immunisation: A step towards achieving immunisation target by 2021*. http://www.unicef.org/esa/media/6191/file/UNICEF-Malawi-2019-2020-Immunisation-Budget-Brief.pdf

Utazi, C. E., Aheto, J. M., Wigley, A., Tejedor-Garavito, N., Bonnie, A., Nnanatu, C. C., . . . Cutts, F. T. (2023). Mapping the distribution of zero-dose children to assess the performance of vaccine delivery strategies and their relationships with measles incidence in Nigeria. *Vaccine, 41(1), 170-181.*

Vidal, F. C., Johns, N., Goodman, T., Heidari, S., Munro, J., & Hosseinpoor, A. (2022). The Association between Childhood Immunization and Gender Inequality: A Multi-Country Ecological Analysis of Zero-Dose DTP

Prevalence and DTP3 Immunization Coverage. *Vaccines (Basel),* 10(7). http://doi.org/10.3390/vaccines10071032.

WHO. (2017). *Question and Answer Booklet on Routine Immunisation.* Author

WHO. (2017). *Question and Answers on Routine Immunisation.* Author

WHO. (2017). *Regional Office of Europe, Immunisation ststems: fact sheet on sustainable Development Goals (SDGs): Healtrh targets*. https://iris.who.int/handle/10665/340835

WHO. (2020). *Immunisation Agenda 2030.* Author

WHO. (2023). *Immunization Data Portal*. Author

WHO. (2024, July 20). Retrieved from Vaccine and Immunisation: https://www.who.int/health-topics/vaccine-and-immunisation#tab=tab_1

WHO. (2024). *50th Anniversary of the Expanded Program on immunisation.* Author

WHO. (2024, j). *Vaccine and Immunisation*. Retrieved from http://www.who.int/health-topics/vaccines-and-immunisation#tab=tab_1.

WHO. (2024, July 15). *Vaccines and Immunisation*. https://www.who.int/health-topics/vaccines-and-immunization

Wigley, A., Lorin, J., Hogan, D., Utazi, E., Hagedorn, E., & Dansereau, E. (2022). Estimates of the number and distribution of zero-dose and under-immunised children across remote-rural, urban, and conflict-affected settings in low and middle-income countries. *PLOS Glob Public Health 2(10): e0001126.* https://doi.org/10.1371/journal.pgph.0001126

Yu Hu, Q. L. (2013). Determinants of childhood immunisation uptake among social economically disadvantaged Migrants in East China. *Int J Environ Res Public Health,* 10(7), 2845–2856. http://doi.org/10.3390/ijerph10072845

Zhang, Y. (2011). *Introduction to Geostatistics* . University of Wyoming

APPENDICES

APPENDIX 1: r-code

```
#installing the packages that will be used for analysis
getwd()
setwd("C:/Users/HP/Documents/Documents/Shapefiles")
install.packages(c("sf", "sp", "raster", "rgdal", "spatialEco",
"gstat", "geoR", "automap", "spdep", "leaflet", "tmap", "maptools",
"rgeos", "sfnetworks", "sp","sf","raster","ggplot2","spatialreg"
,"survey","dplyr","plyr", "SDMTools", "stats", "cluster", "dbscan",
"dendextend", "fpc", "rnaturalearth"))
install.packages("spgwr")
install.packages("writexl")
install.packages (c("arsenal", "gtsummary", "pwr","lme4", "metafor"))
install.packages("car")
install.packages("copula")
install.packages("stats4")
install.packages("MASS")
install.packages("GJRM")
install.packages("lmtest")
install.packages("officer")
install.packages("flextable")
update.packages(ask = FALSE)
remove.packages("flextable")    # Remove the package first
install.packages("flextable")   # Reinstall the package
install.packages("systemfonts")
devtools::install_version("systemfonts", version = "1.0.4")
install.packages("aod")
install.packages("INLA", repos = "https://inla.r-inla-
download.org/R/stable")
##Loading the installed packages in readiness of the analysis
library(INLA)
library(stats4)
library(aod)
library(writexl)
library(officer)
library(GJRM)
library(MASS)
library(copula)
library(car)
```

```r
library(tmap)
library(tidyverse)
library(stats)
library(cluster)
library(survey)
library(spdep)
library(spatialEco)
library(dplyr)
library(spatialreg)
library(sp)
library(sf)
library(raster)
library(lmtest)
library(ggplot2)
library(dplyr)
library(foreign)
library(plyr)
library(psych)
library(geoR)
library(spatial)
library(metafor)
library(lme4)
library(gtsummary)
library(mgcv)
set.seed(123)
#Importing data for 2016 MDHS
data <- read.dta("MWKR7AFL.dta", convert.factors=FALSE)
mydata <- c("caseid","v001", "v002", "v012","v024", "v025",
"v106","v120", "v130", "v131","v151", "v152", "v190", "v716", "h1",
"h3","h7", "b19", "m14", "bord", "sdist")
mydata_subset <- subset(data, select = mydata)
#Defining the age group of the dataset that will be used for analysis
df <- mydata_subset[mydata_subset$b19 >= 12 & mydata_subset$b19 <= 23,
]
##Renaming the variables
names(df)[names(df)=="v012"]<-"Maternal_Age"
names(df)[names(df)=="v024"]<-"Region"
names(df)[names(df)=="v025"]<-"Type_of_Residence"
names(df)[names(df)=="v106"]<-"Education_Level"
names(df)[names(df)=="v130"]<-"Religion"
names(df)[names(df)=="v131"]<-"Ethnicity"
```

```
names(df)[names(df)=="v190"]<-"Wealth_Index"
names(df)[names(df)=="h1"]<-"Has_health_card"
names(df)[names(df)=="h3"]<-"Received_DPT1"
names(df)[names(df)=="h7"]<-"Received_DPT3"
names(df)[names(df)=="b19"]<-"Age_of_child_in_months"
names(df)[names(df)=="sdist"]<-"DIST_CODE"
names(df)[names(df)=="v120"]<-"Exposure_to_Media"
names(df)[names(df)=="bord"]<-"Birth_Order_Number"
names(df)[names(df)=="m14"]<-"Number_of_ANC_visits"
##Creating a variable Zero dose to have value 1 if variable received
DTP1 is 0 or 8 which is no received DTP1 or i dont know
df$Zero_Dose <- ifelse(df$Received_DPT1 %in% c(0,8), 1,
                       ifelse(df$Received_DPT1 %in% c(1,2,3), 0, NA))
##Creating a variable DTP3_uptake to have value 1 if variable received
DTP1 is 0 or 8 which is no received DTP1 or i dont know
df$DPT3_uptake <- ifelse(df$Received_DPT3 %in% c(0,8), 1,
                         ifelse(df$Received_DPT3 %in% c(1,2,3), 0,
NA))
#Creating a variable under immunised if the child received first dose
of DPT and did not receive DPT3
df <- df %>%
  mutate(Under_Immunized = ifelse(DPT3_uptake == 1, 1, 0))
df <- df %>%
  mutate(Under_Immunized1 = ifelse((Zero_Dose == 0 & DPT3_uptake ==
1), 1, 0))
df <- df %>%
  mutate(Number_ANC_Visit_Cat = case_when(
    Number_of_ANC_visits >= 1 & Number_of_ANC_visits <= 4 ~ "Inadequate
visit",
    Number_of_ANC_visits >= 5 ~ "Adequate visit",
  ))
df <- df %>%
  mutate(Birth_Order_Number_Cat = case_when(
    Birth_Order_Number ==1 ~ "1",
    Birth_Order_Number >= 2 & Birth_Order_Number <= 4 ~ "2-4",
    Birth_Order_Number >= 5 & Birth_Order_Number <= 7 ~ "5-7",
    Birth_Order_Number >= 8 ~ ">= 8",
  ))
df <- df %>%
  mutate(Maternal_Cat = case_when(
    Maternal_Age >=15& Maternal_Age <= 24 ~ "15-24",
```

```
    Maternal_Age >= 25 & Maternal_Age <= 34 ~ "25-34",
    Maternal_Age >= 35 & Maternal_Age <= 44 ~ "35-44",
    Maternal_Age >= 45 ~ ">= 45",
  ))
#cleaning the dataset by removing NAs
View(df)
df$Under_Immunized[is.na(df$Under_Immunized)] <-
mean(df$Under_Immunized, na.rm = TRUE)  # Mean
df$Zero_Dose[is.na(df$Zero_Dose)] <- median(df$Zero_Dose, na.rm =
TRUE)  # Median
sum(df$Under_Immunized1)
sum(df$Under_Immunized)
sum(df$Zero_Dose)
clean_data<-na.omit(clean_data)
sum(clean_data$Under_Immunized)
sum(clean_data$Zero_Dose)
write.csv(clean_data, file = "mydata.csv", row.names = TRUE)
clean_data<-df
library(GJRM)
#importing spatial data for Malawian district
districts = st_read(dsn = "Districts.shp", stringsAsFactors = FALSE)
#Merging spatial data with the ditrict proportions for zero dose and
under immunised
View(districts)
districts_sp <- as(districts, "Spatial")
districts_sp
pol<-polys.setup(districts_sp)
##merged_data2 <- merge(districts, clean_data, by.x = c("DIST_CODE"),
by.y = c("DIST_CODE"), all.x=TRUE)
dist<-pol$polys
dist
name <- cbind(names(dist), pol$names1)
name
xt <- list(polys = dist)
xt
str(pol)
dis <- NULL
for(i in 1:dim(clean_data)[1]){
  if(clean_data$DIST_CODE[i]=="101") dis[i] <- 1
  if(clean_data$DIST_CODE[i]=="102") dis[i] <- 2
  if(clean_data$DIST_CODE[i]=="103") dis[i] <- 3
```

```
  if(clean_data$DIST_CODE[i]=="104") dis[i] <- 4
  if(clean_data$DIST_CODE[i]=="105") dis[i] <- 5
  if(clean_data$DIST_CODE[i]=="106") dis[i] <- 6
  if(clean_data$DIST_CODE[i]=="107") dis[i] <- 7
  if(clean_data$DIST_CODE[i]=="201") dis[i] <- 8
  if(clean_data$DIST_CODE[i]=="202") dis[i] <- 9
  if(clean_data$DIST_CODE[i]=="203") dis[i] <- 10
  if(clean_data$DIST_CODE[i]=="204") dis[i] <- 11
  if(clean_data$DIST_CODE[i]=="205") dis[i] <- 12
  if(clean_data$DIST_CODE[i]=="206") dis[i] <- 13
  if(clean_data$DIST_CODE[i]=="207") dis[i] <- 14
  if(clean_data$DIST_CODE[i]=="208") dis[i] <- 15
  if(clean_data$DIST_CODE[i]=="209") dis[i] <- 16
  if(clean_data$DIST_CODE[i]=="210") dis[i] <- 17
  if(clean_data$DIST_CODE[i]=="301") dis[i] <- 18
  if(clean_data$DIST_CODE[i]=="302") dis[i] <- 19
  if(clean_data$DIST_CODE[i]=="303") dis[i] <- 20
  if(clean_data$DIST_CODE[i]=="304") dis[i] <- 21
  if(clean_data$DIST_CODE[i]=="305") dis[i] <- 22
  if(clean_data$DIST_CODE[i]=="306") dis[i] <- 23
  if(clean_data$DIST_CODE[i]=="307") dis[i] <- 24
  if(clean_data$DIST_CODE[i]=="308") dis[i] <- 25
  if(clean_data$DIST_CODE[i]=="309") dis[i] <- 26
  if(clean_data$DIST_CODE[i]=="310") dis[i] <- 27
  if(clean_data$DIST_CODE[i]=="311") dis[i] <- 28
  if(clean_data$DIST_CODE[i]=="312") dis[i] <- 29
  if(clean_data$DIST_CODE[i]=="313") dis[i] <- 30
  if(clean_data$DIST_CODE[i]=="314") dis[i] <- 31
  if(clean_data$DIST_CODE[i]=="315") dis[i] <- 32
}
clean_data$DIST_CODE <- as.factor(dis)
clean_data$Religion<-factor(clean_data$Religion)
clean_data$Education_Level<-factor(clean_data$Education_Level)
clean_data$Type_of_Residence<-factor(clean_data$Type_of_Residence)
clean_data$Region<-factor(clean_data$Region)
clean_data$Wealth_Index<-factor(clean_data$Wealth_Index)
clean_data$Ethnicity<-factor(clean_data$Ethnicity)
clean_data$Number_ANC_Visit_Cat<-
factor(clean_data$Number_ANC_Visit_Cat)
clean_data$Exposure_to_Media<-factor(clean_data$Exposure_to_Media)
```

```
clean_data$Birth_Order_Number_Cat<-
factor(clean_data$Birth_Order_Number_Cat)
clean_data$Maternal_Cat<-factor(clean_data$Maternal_Cat)
formula_list <- list(
  Zero_Dose ~ Ethnicity + Religion  + Education_Level + Wealth_Index
+Region +Type_of_Residence +s(Maternal_Age) + s(DIST_CODE, bs = "mrf",
xt = xt, k=3),
  Under_Immunized ~ Ethnicity + Religion  + Education_Level +
Wealth_Index +Region +Type_of_Residence + s(Maternal_Age) +
s(DIST_CODE, bs = "mrf", xt = xt, k=3)
)
formula_list1 <- list(
  Zero_Dose ~ Ethnicity + Religion  + Education_Level + Wealth_Index
+Region +Type_of_Residence + Birth_Order_Number_Cat+s(Maternal_Age) +
Number_ANC_Visit_Cat + Exposure_to_Media + s(DIST_CODE, bs = "mrf", xt
= xt, k=3),
  Under_Immunized ~ Ethnicity + Religion  + Education_Level +
Wealth_Index +Region +Type_of_Residence +
Birth_Order_Number_Cat+Maternal_Cat + Number_ANC_Visit_Cat +
Exposure_to_Media + s(DIST_CODE, bs = "mrf", xt = xt, k=3)
)
# Fit the model using the list of equations
model <- gjrm(formula = formula_list,
              data = clean_data,
              copula = "N",
              margins = c("probit", "probit"),
              model = "B")  # B for bivariate model
summary(model)
gtsummary::tbl_regression(model, intercept = T)



model1 <- gjrm(formula = formula_list1,
               data = clean_data,
               copula = "N",
               margins = c("probit", "probit"),
               model = "B")  # B for bivariate model
summary(model1)
gtsummary::tbl_regression(model1, intercept = T)
gaussian<-AIC(model1)
model2 <- gjrm(formula = formula_list1,
```

```
               data = clean_data,
               copula = "G0",
               margins = c("probit", "probit"),
               model = "B")  # B for bivariate model
summary(model2)
gumbel<-AIC(model2)
model3 <- gjrm(formula = formula_list1,
               data = clean_data,
               copula = "C0",
               margins = c("probit", "probit"),
               model = "B")  # B for bivariate model
summary(model3)
clayton<-AIC(model3)
# Plot the model with modified scale and SE options
plot(model, eq = 2, seWithMean = TRUE,
     scale = 0, pages = 1, jit = TRUE)
plot(model1, eq = 1, seWithMean = TRUE,
     scale = 0, pages = 1, jit = TRUE)
plot(model1, eq = 2, seWithMean = TRUE,
     scale = 0, pages = 1, jit = TRUE)
custom_colors <- colorRampPalette(c("blue", "white", "red"))(100)
plot(model1, eq = 1, seWithMean = TRUE, scale = 0, pages = 1, jit =
TRUE,
     col.regions = custom_colors)
plot(model1, eq = 1)
conv.check(model1)
model5<-SemiParBIV(formula_list1, data= clean_data)
conv.check(model5)
set.seed(1)
sb<-summary(model5)
sb
plot(model5, eq=1)
library(grDevices)  # For color palettes
# Custom color ramp: Red for negative, Yellow for zero, and a gradient
towards White for positive
custom_colors <- colorRampPalette(c("red", "yellow", "white"))(100)
# Plot with custom color scale
plot(model1, eq = 1, seWithMean = TRUE, scale = 0, pages = 1, jit =
TRUE,
     col.regions = custom_colors)  # Apply custom colors
```

```
plot(model1, eq = 1, seWithMean = TRUE, scale = 0, pages = 1, jit = TRUE,
     col.regions = custom_colors, zlim = c(-0.3, 0.1))  # Set color limits to match legend
data2 <- read.csv("mean_estimates1.csv")
merged_dataa <- merge(districts, data2, by.x = c("DIST_CODE"), by.y = c("DIST_CODE"), all.x=TRUE)
clean_merged<-na.omit(merged_dataa)
View(clean_merged)
tm_shape(clean_merged) +
  tm_polygons("zerod",
              title = "estimates",
              palette = "plasma",
              style = "quantile") +
  tm_layout(
    legend.outside = TRUE,
    frame = FALSE)
tm_shape(clean_merged) +
  tm_polygons("under",
              title = "estimates",
              palette = "plasma",
              style = "quantile") +
  tm_layout(
    legend.outside = TRUE,
    frame = FALSE)
# Extract the coefficients
coeff <- coef(model1)
# Extract the standard errors (from the diagonal of the variance-covariance matrix)
stderr <- sqrt(diag(vcov.gjrm(model1)))
# Calculate the odds ratios
odds_ratios <- exp(coeff)
# Combine the results into a data frame without the extra comma
results <- data.frame(
  Coefficients = coeff,
  Odds_Ratio = odds_ratios
)
# View the results
print(results)
#spatial modelling distibution
```

```
#extracting data to form a dataset with three variables: DIST_CODE,
Zero_Dose and Under_Immunised
mydataa <- c("DIST_CODE","Zero_Dose", "Under_Immunized")
mydata1 <- subset(clean_data, select = mydataa)
clean_df<-na.omit(mydata1)
View(clean_df)
cross_tab <- table(mydata1$DIST_CODE, clean_data$Zero_Dose)
#Calculating proportions for Zero Dose for each district
proportions <- prop.table(cross_tab, margin = 1)
print(proportions)
zero_dose_proportions <- proportions[, "1"]
# Print the zero dose proportions
print(zero_dose_proportions)
write.csv(zero_dose_proportions, file = "zero_dose_proportions.csv",
row.names = TRUE)
#Calculating proportions for under_Immunised
underImm <- table(mydata1$DIST_CODE, clean_data$Under_Immunized)
Under_Immunized_proportions <- prop.table(underImm, margin = 1)
print(Under_Immunized_proportions)
write.csv(Under_Immunized_proportions,             file              =
"Under_Immunized_proportions1.csv", row.names = TRUE)
Under_Immunized_proportions <- proportions[, "1"]
# Print the under immunised proportions
print(Under_Immunized_proportions)
#importing dataset with three variables
data2 <- read.csv("District Proportions1.csv")
View(data2)
#importing spatial data for Malawian district
districts = st_read(dsn = "Districts.shp", stringsAsFactors = FALSE)
#Merging spatial data with the ditrict proportions for zero dose and
under immunised
merged_data1 <- merge(districts, data2, by.x = c("DIST_CODE"), by.y =
c("DIST_CODE"), all.x=TRUE)
clean_merged<-na.omit(merged_data1)
View(clean_merged)
#Calculating neighbor list
nb <- poly2nb(clean_merged, queen=TRUE)
nb
#Calculating spatial weights list
lw <- nb2listw(nb, style="W", zero.policy = TRUE)
lw
```

```
clean_merged <- st_transform(clean_merged, crs = 4326)
zero_dose <- clean_merged$Zero_Dose
under_immunised <- clean_merged$Under_Immunised
length(zero_dose)
nrow(districts)
#Map displaying Proportions of the zero dose in Malawi
tm_shape(clean_merged) +
  tm_polygons("Zero_Dose",
              title = "Proportions of Zero Dose Children",
              palette = "plasma",
              style = "quantile") +
  tm_layout(
            legend.outside = TRUE,
            frame = FALSE)
tm_shape(clean_merged) +
  tm_polygons("Zero_Dose",
              title = "Proportions of Zero Dose Children",
              palette = "plasma",
              breaks = c(0, 0.01, 0.05, 1),  # Define the breaks for
categories
              labels = c("<1%", "1%-5%", ">5%")) +  # Custom labels
for the breaks
  tm_layout(
    legend.outside = TRUE,
    frame = FALSE,
tm_shape(clean_merged) +
  tm_polygons("Under_Immunised",
              title = "Proportions of under immunised Children",
              palette = "plasma",
              breaks = c(0, 0.01, 0.05, 1),  # Define the breaks for
categories
              labels = c("<1%", "1%-5%", ">5%")) +  # Custom labels
for the breaks
  tm_layout(
    legend.outside = TRUE,
   frame = FALSE)
#Proportions of the Under-Immunised in Malawi
tm_shape(clean_merged) +
  tm_polygons("Under_Immunised",
              title = "Proportions of Under Immunised Children",
              palette = "plasma",
```

```
            style = "quantile")
  tm_layout(
    legend.outside = TRUE,
    frame = FALSE)
# Perform Moran's I test for zero dose children
moran_test <- moran.test(zero_dose, lw)
print(moran_test)
#plotting  Moran's I for zero dose children
moran_plot <- moran.plot(zero_dose, lw, main = "Moran Scatterplot: Zero
Dose")
# Perform Moran's I test for under immunised children
moran_test1 <- moran.test(under_immunised, lw)
print(moran_test1)
#plotting moran for under-immunised
moran_plot1  <-  moran.plot(under_immunised,  lw,  main  =  "Moran
Scatterplot: Under Immunised")
#conducting local moran I for zero dose
local_moran <- localmoran(zero_dose, lw)
summary(local_moran)
#local_moran<-na.omit(local_moran)
#summary(local_moran)


clean_merged$local_moran <- local_moran[,1]

# Extract the local Moran's I values

# Plotting Local Moran's I
tm_shape(clean_merged) +
  tm_fill("local_moran", style = "quantile", title = "Local Moran's
I") +
  tm_borders() +
  tm_layout(
    legend.outside = TRUE,           # Place legend outside the map
    legend.outside.position = "right", # Position the legend to the
right
    legend.width = 0.2,              # Adjust the legend width
    legend.text.size = 1,            # Increase legend text size
    legend.title.size = 1.2,         # Increase legend title size
    title.size = 1.5,                # Increase map title size
    inner.margins = c(0.1, 0.1, 0.1, 0.1),  # Adjust inner margins
```

```
    fontfamily = "sans",            # Change font family to a commonly
available one
    frame = FALSE                    # Remove the frame around the map
  )

#conducting local moran I for Under immunised
local_moran1 <- localmoran(under_immunised, lw)
summary(local_moran1)
#local_moran1<-na.omit(local_moran1)
#summary(local_moran1)
clean_merged$local_moran1 <- local_moran1[,1]

# Plotting Local Moran's I for under Immunised
tm_shape(clean_merged) +
  tm_fill("local_moran1", style = "quantile", title = "Local Moran's
I") +
  tm_borders() +
  tm_layout(
    legend.outside = TRUE,           # Place legend outside the map
    legend.outside.position = "right", # Position the legend to the
right
    legend.width = 0.2,              # Adjust the legend width
    legend.text.size = 1,            # Increase legend text size
    legend.title.size = 1.2,         # Increase legend title size
    title.size = 1.5,                # Increase map title size
    inner.margins = c(0.1, 0.1, 0.1, 0.1),  # Adjust inner margins
    fontfamily = "sans",            # Change font family to a commonly
available one
    frame = FALSE                    # Remove the frame around the map
  )



#Calculate Getis-Ord Gi* statistics for Zero Dose
gi_star <- localG(zero_dose, lw)
summary(gi_star)

# Add Gi* statistics to your data
clean_merged$gi_star <- gi_star

View(clean_merged)
#plotting Gi* for Zero dose children
```

```
tm_shape(clean_merged) +
  tm_fill("gi_star",
          style = "quantile",
          title = "Getis-Ord Gi*") +
  tm_borders() +
  tm_layout(
    main.title = "Getis-Ord Gi* for Zero Dose Percentage",
    main.title.size = 1.2,
    legend.outside = TRUE,
    frame=FALSE
  )


gi_star1 <-localG(under_immunised, lw)
summary(gi_star1)
clean_merged$gi_star1 <- gi_star1
# Add Gi* statistics to your data
clean_merged$gi_star <- gi_star
#plotting Gi* for under Immunised
tm_shape(clean_merged) +
  tm_fill("gi_star1",
          style = "quantile",
          title = "Getis-Ord Gi*") +
  tm_borders() +
  tm_layout(
    main.title = "Getis-Ord Gi* for Under Immunised Percentage",
    main.title.size = 1.2,
    legend.outside = TRUE,
    frame=FALSE
 )
```

APPENDIX 2: Table 5: Proportions of the zero dose and under immunised

District	% of zero dose	% of under immunised	District	% of zero dose	% of under immunised
Chitipa	3.17%	11.11%	Dedza	1.49%	6.93%
Karonga	2.35%	4.12%	Ntcheu	0.41%	4.94%
Nkhatabay	2.20%	5.49%	Lilongwe city	1.90%	5.71%
Rumphi	0.55%	5.49%	Mangochi	2.12%	8.48%
Mzimba	0.98%	4.88%	Machinga	5.79%	8.68%
Likoma	0.00%	0.78%	Zomba rural	0.56%	0.56%
Mzuzu city	0.00%	5.00%	Chradzulu	2.34%	5.26%
Kasungu	3.20%	5.48%	Blantyre rural	2.88%	5.77%
Nkhota kota	1.34%	9.38%	Mwanza	1.18%	4.73%
Ntchisi	2.58%	11.34%	Thyolo	7.60%	9.94%
Dowa	2.97%	5.45%	Mulange	1.05%	5.76%
Salima	0.43%	4.27%	Phalombe	1.38%	5.99%
Lilongwe rural	0.66%	4.61%	Chikwawa	3.41%	7.80%
Mchinji	1.67%	7.08%	Nsanje	8.08%	21.21%
Balaka	0.00%	3.48%	Zomba city	0.00%	0.00%
Neno	0.96%	3.37%	Blantyre city	13.27%	20.41%

APPENDIX 3: Map of Malawi