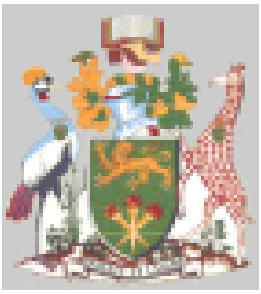


UNIVERSITY OF NAIROBI

STATISTICAL METHODS FOR CORRELATED DATA: APPLICATION TO SEVERE MALARIA CASE MANAGEMENT EVALUATION IN KENYA

BY

BEATRICE K. MACHINI (BScN, MSc Medical Statistics)

W83/52076/2017

A Thesis Submitted in Fulfillment of the Requirements for the Award of the Degree of Doctor of Philosophy (PhD) in Medical Statistics by the University of Nairobi

2022

DECLARATION

Candidate Declaration

"I declare that this thesis is my original work and has not been submitted elsewhere for examination, award of a degree, or publication. The work done by others or myself previously have been properly acknowledged in the text and referenced in accordance with the University of Nairobi requirements".

Sign:  Date: July 21, 2022

Beatrice K. Machini, BScN, MSc. Medical Statistics (W83/52076/2017),

University of Nairobi, Department of Public and Global Health,

P. O. Box 19982-00202, Nairobi; bkemunto2002@gmail.com

Supervisors Declaration

"This thesis is submitted for examination with our approval as research supervisors".

Prof. Thomas Achia. Sign:  Date: July 21, 2022

Strathmore University, Institute of Mathematical Sciences,

P.O Box 59857-00200,

Nairobi, Kenya.

Email: tachia@strathmore.edu

Dr. Hillary Kipruto. Sign:  Date: July 21, 2022

University of Nairobi, Department of Public and Global Health,

P.O Box 30197-00100,

Nairobi, Kenya.

Email: kipruto@unobi.ac.ke / kiprutoh@gmail.com

Prof. Dejan Zurovac. Sign:  Date: 22 July 2022

University of Oxford's Major Overseas Programme in Kenya,

Department of Public Health and Research, KEMRI Wellcome Trust,

P.O Box 43640-00100,

Nairobi, Kenya.

Email: dzurovac@kemri-wellcome.org/dzurovac8@gmail.com

Dr. Jacqueline Chesang. Sign:  Date: 22 July 2022

University of Nairobi, Department of Public and Global Health,

P.O Box 30197-00100,

Nairobi, Kenya.

Email: chesangj@uonbi.ac.ke

DEDICATION

“To my husband Joash and children Samuel, Ray and Stella”.

ACKNOWLEDGEMENTS

I thank God Almighty for His countless blessings, knowledge and opportunity that enabled me to complete this study successfully.

Greater appreciation is extended to my supervisors, Prof. Thomas Achia, Dr. Hillary Kipruto, Prof. Dejan Zurovac and Dr. Jacqueline Chesang for their professional technical support before and during this study. It is truly my honour to have been their student.

Special thanks also go to Prof. M'Imunya Machoki, the Principal, University of Nairobi, for his support and encouragement when I was starting this PhD study. I would also like to convey my heartfelt thanks to Mrs. Ann Wang'ombe for her mentorship in building my career in Medical Statistics.

Many thanks to Prof. Joyce Olenja, the Chair of the Department of Public and Global Health (PGH), University of Nairobi, and Prof. Julius Oyugi, the Director of the Institute of Tropical & Infectious Diseases (UNITID), whose invaluable leadership and management provided me an opportunity to learn and grow. I would also like to convey my warmest thanks to Dr Waqo Ejersa for his support during his time at the Division of National Malaria Program (DNMP).

It is not possible to individually thank everyone but my sincere gratitude goes to my friends, Dr. Beatrice Amboko, Dr. Peter Macharia and Paul Mwaniki for their support during this study. I also acknowledge the support from various institutions: UNITID, PGH, KEMRI and DNMP.

Finally, my heartfelt thanks go to my family, Joash, Sammy, Ray and Stella for their prayers, patience, love and encouragement.

RESEARCH OUTPUTS

Published work

Machini, B., Zurovac, D., Amboko, B., Malla, L., Snow, R.W., Kipruto, H. and Achia, T.N., 2020. Predictors of health workers' knowledge about artesunate-based severe malaria treatment recommendations in government and faith-based hospitals in Kenya. *Malaria journal*, *19*(1), pp.1-10.

Machini, B., Achia, T.N., Chesang, J., Amboko, B., Mwaniki, P. and Kipruto, H., 2022. Cross-sectional study to predict subnational levels of health workers' knowledge about severe malaria treatment in Kenya. *BMJ open*, *12*(1), p.e058511.

Machini, B., Achia, T.N., Kipruto, H., Amboko, B., Chesang, J., 2022. Factors associated with hospital length of stay in patients admitted with suspected malaria in Kenya: secondary analysis of a cross-sectional survey. *BMJ Open*, *12*(6), p.e059263.

ABSTRACT

Correlated data arise when clusters of observations are similar to each other. This is common in public health research where measurements on the same subject are repeatedly tracked over time and space. Health facility surveys are routinely implemented to monitor severe malaria case management. The dataset from this kind of surveys have hierarchical structure and requires statistical methods that can properly account for correlated data. This study developed a statistical model for analyzing correlated data by evaluating severe malaria case management in Kenya. This was secondary analysis using data from repeated cross-sectional inpatient malaria case management surveys undertaken at the government (GOK) and the Faith Based Organization (FBO) health facilities in Kenya from 2016 to 2019. The number of health facilities, health workers and suspected malaria admissions ranged from; 86 to 94, 330 to 367 and 2243 to 2485 respectively, across the survey period. Three methods of data collection were applied; retrospective data was extracted from the patients' files, health workers (clinicians and nurses) from the paediatric and medical ward were interviewed and, health facilities were assessed for readiness. Firstly, to evaluate the impact of correlation on outcomes, multilevel mixed effect logistic regression modelling was performed to identify the predictors of the health workers' knowledge about artesunate-based severe malaria treatment recommendations in Kenya. During modeling, the random effects and intracluster correlation were examined. Secondly, a Bayesian hierarchical spatial model beyond predictor analysis was applied to predict subnational estimates of knowledge levels. Three hierarchical models were fitted with ordinal logistic regression. The best fitting models were overlaid on a map showing all counties in Kenya. Lastly, to model factors related to hospital length of stay for severe malaria patients, competing risk approach was applied based on usual clinical setting in Kenya. Two models were fitted and their parameter estimates examined; conventional Cox regression model to obtain a cause-specific hazard (CSH) ratio and Fine and Gray competing risks method to obtain a subdistribution hazard (SDH) ratio. Evaluating the impact of correlation while adjusting for health facility and county structures, the parameter estimates were slightly varied and some of the variables that were significant in unadjusted analyses lost their significance. With respect to the treatment policy knowledge, clinicians compared to nurses were more likely to have high knowledge, both at the GOK (adjusted Odds Ratio [aOR] =1.86; 95% CI: 1.18-2.91) and FBO

health facilities (aOR=2.27; 95% CI=1.41-3.65). Health workers' knowledge about recommended artesunate dosing was significantly associated with displayed artesunate administration posters (aOR=2.17; 95% CI=1.24-3.79) at the GOK health facilities. The knowledge of artesunate dosing intervals was significantly associated with the availability of artesunate (aOR=2.18; 95% CI=1.20-3.94) at the FBO health facilities. Health workers in the paediatric ward had high knowledge about artesunate preparation compared to those in medical ward (aOR=1.99; 95% CI=1.33-2.99) at the GOK health facilities. The knowledge of preferred route of artesunate administration was significantly higher in high malaria risk areas compared to low areas. Conditional on the fixed effects covariates, the health worker knowledge on severe malaria treatment policy and artesunate preparation were slightly correlated within the same county. The random effects composed about 4% to 11% of the total residual variance. Artesunate dose and dosing interval were slightly correlated within the same health facilty. The random effects composed about 7% and 26% of the total residual variance. While, artesunate route of administration was slightly correlated within the same county in the GOK sector, and within the same health facility in the FBO sector. The random effects composed about 26% and 39% of the total residual variance in the GOK and FBO sector respectively. Adjusting for county structures in Bayesian hierarchical spatial approach, the best model fitted with spatially structured random effects and the spatial variations of the knowledge level across the 47 counties exhibited neighborhood influence. The likelihood of having high knowledge on severe malaria treatment policy was lower among nurses relative to clinicians (aOR=0.48, 95% CI: 0.25 to 0.87), health workers older than 30 years were 61% less likely to have high dosing knowledge compared to younger health workers (aOR=0.39, 95% CI: 0.22 to 0.67) while those exposed to artesunate poster had 2.4-fold increased odds of higher dosing knowledge compared to non-exposed health workers (aOR=2.38, 95% CI: 1.22 to 4.74). Based on the spatial maps, the health workers in Kisii county had high knowledge levels (>10%) on severe malaria treatment policy. In addition, Muranga, Embu, Uasin Gishu, Kiambu, and Kisumu counties had high knowledge levels (>10%) about artesunate dose, and slightly more than a third of the counties had high knowledge levels (>10%) on artesunate preparation. Modelling factors associated with the length of stay (LOS) among the severe malaria patients, the median LOS was 4 days. The factor estimates and the confidence interval spans between the SDH and CSH models were slightly varied. Among the

factors assessed for influencing LOS, respiratory rate (Subdistribution-Hazard ratio [SDHR]: 0.873; 95% CI: 0.789–0.967), oxygen saturation (SDHR: 0.859; 95% CI: 0.754–0.978), Hemoglobin(Hb)/Hematocrit (HCT) (SDHR: 0.769; 95% CI: 0.709–0.833), glucose/Random Blood Sugar(RBS) (SDHR: 0.766; 95% CI: 0.704–0.833) and documentation of at least one clinical feature of severe malaria (SDHR: 0.696; 95% CI: 0.626–0.774) were significantly associated with shortened LOS. Conversely, patients with confirmed severe malaria (SDHR: 1.214; 95% CI: 1.082–1.362) and those treated with injectable artesunate (SDHR: 1.339; 95% CI: 1.184–1.515) were significantly associated with prolonged LOS. The malaria program can utilize multilevel mixed effect logistic regression modelling to account for the hierarchical structure of the survey data. Further, Bayesian hierarchical spatial model can be used to account for the substantial heterogeneity among the health workers at various levels. In the presence of a competing risk and correlation, SDH model is the best model. The program should target on interventions likely to improve health workers' knowledge about severe malaria case management; artesunate availability, access to guidelines and exposure of artesunate poster in the wards. Based on the spatial maps, focused multidisciplinary interventions implemented can bridge the knowledge gaps identified at the subnational levels. Measurement of temperature, respiratory rate, oxygen saturation and laboratory tests (Hb/HCT, glucose/RBS) were significantly associated with shortened LOS. Treating severe malaria patients with artesunate boosted survival and increased LOS. The statistical models developed can be applied to analyse similar correlated data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACTs	Artemisinin-based Combination Therapy
AG	Andersen and Gill
AIC	Akaike Information Criterion
AIDS	Acquired Immunodeficiency Syndrome
AL	Artemether-Lumefantrine
AMFm	Affordable Medicine Facility for malaria
ANC	Antenatal Care
aOR	Adjusted Odds Ratio
AS-AQ	Artesunate-Amodiaquine
AS-MQ	Artesunate-Mefloquine
AS-SP	Artesunate-Sulfadoxine-Pyrimethamine
BIC	Bayesian Information Criterion
CAR	Conditional autoregressive
CHMT	County Health Management Team
CHVs	Community Health Volunteers
CI	Confidence Interval
CI	Credible Intervals

CIF	Cumulative Incidence Function
CIN	Clinical Information Network
CMEs	Continuous Medical Education
CRM	Continuous Ratio Model
CSH	Cause Specific Hazard
CSHR	Cause Specific Hazard Ratio
CWC	Child Welfare Clinics
DHA-PPQ	Dihydroartemisinin-Piperaquine
DIC	Deviance Information Criterion
DNMP	Division of National Malaria Program
EPR	Epidemic Preparedness and Response
EQA	External Quality Assessment (KNH
ETAT	Emergency Triage Assessment and Treatment
FBO	Faith Based Organizations
GEE	Generalized Estimating Equations
GLM	Generalized Linear Models
GLMMs	Generalized Linear Mixed Models
GLS	Generalized Least Squares

GOK	Government of Kenya
GTS	Global Technology Strategy
Hb	Haemoglobin
HCT	Haematocrit
HF	Health Facilities
HIS	Health Information and Systems
HIV	Human Immunodeficiency Virus
HRP2	Histidine Rich Protein 2
IDSR	Integrated Disease Surveillance and Response
IM	Intramuscular
IMCI	Integrated Management of Childhood Illness
IMR	Infant Mortality Rate
IPTp	Intermittent Preventive Treatment in pregnancy
IRS	Indoor Residual Spraying
ITN	Insecticide-Treated Mosquito Nets
IV	Intravenous
IVM	Integrated Vector Management
KDHS	Kenya Demographic Health Survey

KEMSA	Kenya Medical Supplies Authority
KHIS	Kenya Health Information Systems
KHSSP	Health Sector Strategic and Investment Plan
KM	Kaplan-Meier
KMIS	Kenya Malaria Indicator Survey
KMS	Kenya Malaria Strategy
LLINs	Long Lasting Insecticide Treated Nets
LOS	Length of Stay
LSM	Larval Source Management
M&E	Monitoring and Evaluation
MCA	Multiple Correspondence Analysis
MCA	Multiple Correspondence Analysis
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
MIP	Malaria in Pregnancy
MLE	Maximum Likelihood Estimation
MLM	Multilevel Modeling

MMR	Maternal Mortality Rate
MoH	Ministry of Health
OR	Odds Ratio
PGH	Public and Global Health
pLDH	Parasite Lactic Acid Dehydrogenase
POM	Proportional Odds Model
PPOM-R	Partial Proportional Odds Model with Restrictions
PPOM-UR	Partial Proportional Odds Model-Without Restrictions
PWE	Piecewise Exponential
PWP	Prentice Williams and Peterson
RDT	Rapid Diagnostic Tests
SAC	Spatial autocorrelation
SAR	Simultaneous Autoregressive Models
SCHMT	Sub-county County Health Management Team
SDH	Subdistribution Hazard
SDHR	Sub-distribution Hazard Ratio
SEVM	Spatial Eigenvector Mapping
SM	Stereotype Model

SOPs	Standard Operating Procedures
SP	Sulfadoxine-pyrimethamine
SRE	Spatially-structured Random Effects
UHC	Universal Health Coverage
UNITID	University of Nairobi Institute of Tropical & Infectious Diseases
URE	Unstructured Random Effects
WAIC	Watannabe-Akaike Information Criterion
WHO	World Health Organization
WLW	Wie Lin and Weissfeld

CHAPTER ONE

INTRODUCTION

This introductory chapter presents the study background, the problem statement, research questions, objectives and justification of the study ending with the structure of the thesis.

1.1 Background

Correlated data arise when observations are made on individuals who share certain underlying characteristics over time and space (O'Brien and Fitzmaurice, 2005). Correlation describes the relationship between data sets, in practice, correlated data can occur in many situations. In longitudinal study, the subjects are measured over time based on the outcome of interest. Measurements collected at different time points on the same subject are not independent from each other. Similarly, in studies with naturally occurring groups, responses collected from members of the same group tend to be more similar than observations from different groups. The data collected from patients or health workers from the same hospital are likely to be correlated.

In public health research, it's usual to see data that are correlated, such as clustered, hierarchical, non-independent, and geographical. It is vital to collect data on how the health facilities handle severe malaria cases in order to evaluate patient treatment. The recurrent cross-sectional health facility surveys are used on a regular basis to check the readiness and adherence of health systems to treatment guidelines. In order to correctly account for the correlation between responses from different subjects in a hierarchical dataset, appropriate approaches must be used.

Modeling population parameters is critical for establishing statistical inferences and making decisions. In public health research, regression analysis is the most often used statistical analytic tool. All data points are assumed to be statistically independent in regression analysis. Clustered sampling is frequently used to introduce correlations between observations inside a single cluster in numerous observational investigations. As a result, the correlation among the subjects is ignored in the regression analysis (Hu *et al*., 2021 Sebastiani *et al*., 2016).

Individuals who have similar underlying features throughout time or space or both might be observed to contain correlated data. Data in longitudinal studies, in which patients are followed over time to see if the desired objective has been achieved, are not kept separate. As an illustration, patients' data collected from various hospitals are grouped together by county. Studies conducted within a hospital setting are more likely to be of a similar nature. It is also more common for measurements taken at different moments in time from different groups to have similar responses to the same subject. The data structure of this type has correlations inside the hospital, which must be taken into consideration while estimating parameters.

It has been difficult for scientists to analyze and interpret correlated data since most statistical tests presume that observations are independent of each other and overlook correlation in data. Statistical power is reduced and type II error is increased when correlation is ignored because it leads to an overestimation of variability within subjects or clusters. The underestimating of *P*-values in comparisons across subjects or clusters results in substantial effects or type I errors (Sainani, 2010). Response measurement correlations between and within cluster subjects must be taken into account when doing statistical analysis (Cameron and Miller, 2015).

In the past, correlated survival and geographical data models have been classified as clustered. However, the statistical methods used for analyses ignored the correlations between data over time or space. Furthermore, in some studies evaluating length of stay (LOS) in malaria patients, competing events were not taken into account; mortality outcomes were often censored instead of being considered as a competing risk. This study examined various statistical models from a methodological standpoint that can be used to analyse correlated data, inorder to develop an appropriate statistical method for analyzing severe malaria case management survey data.

1.2 Statement of the Problem

A multilevel or hierarchical structure in the malaria survey data increases the likelihood of correlation. Inappropriate statistical inferences can be made by incorrectly analyzing associated data. The choice of the model, estimator, and summaries determines how the dependence is accounted for. Comparative survival data analysis overlooking competing risks results in a skewed estimate of the incidence rate of the outcome across time.

The surveys have been conducted by the malaria program to gauge the level of care given to patients who have been admitted with the disease. Health facilities, which are divided into epidemiological regions or counties, are used to group the participants in the surveys. Patients, health facilities, and health care workers can all have a role in determining the severity of malaria. This informs patient-level and health-care worker-level analyses, both of which are nested within the health-care facility or, county levels. Again, the data obtained from health facilities provides national estimates that do not account for subnational variation or analyze the impact of malaria interventions as they change over time.

Clustered correlation may be evident in individuals in situations when survival data have conflicting risks. According to Austin & Fine (2017), 77.4% of the high-impact journal papers examined, were potentially exposed to competing hazards that were not taken into consideration in their statistical analysis. Surviving subjects are typically assumed to be censored out of survival data analysis by conventional methods.

This study used health facility survey data collected to monitor inpatient malaria case-management. The design of the surveys involves multi-stage sampling where, the subjects are clustered within health facilities that are nested within the county or epidemiological zones. The subjects within the same hospital may have responses or outcomes that are similar. For example, two health workers selected from the same ward may respond more similarly than two health workers randomly selected from different wards or two randomly selected patients from the same hospital may have outcomes that are more similar than the outcomes of two randomly selected subjects from different hospitals. Similar data induce intra-cluster correlation between observations that must be accounted for in assessing the relationship between risk factors and health outcomes.

Correlated data is not independent, however, most statistical tests assume that observations are independent of each other and ignore correlation in data and this leads to biased parameter estimates and invalid statistical inferences. Ignoring the within cluster correlations leads to overestimation of variability while, ignoring between cluster correlations leads to underestimation of variability. In this study, we focus on statistical methods for analysing correlated data to develop models for evaluating severe malaria case management in Kenya.

1.3 Research questions

1. Which statistical methods are suitable for correlated data adjusting for structures?
2. What is the impact of correlation on dichotomous, polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI) and prediction?
3. What are factors associated with the duration of hospitalization of severe malaria?

1.4 Objectives of the Study

1.4.1 Main Objective

To develop statistical methods for correlated data by evaluating severe malaria case management data in Kenya, 2016 to 2019.

1.4.2 Specific Objectives

a. To investigate the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (standard rrrors and confidence intervals), and prediction, adjusting for health facility and county structures in Kenya.

b. To model factors associated with the length of stay among severe malaria patients using a competing risk survival analysis approach, adjusting for health facility structures in Kenya.

c. To apply Bayesian hierarchical approach to analyse polytomous data, adjusting for county structures in Kenya.

1.5 Justification of the study

This study will inform the Ministry of Health in planning hospital services, interventions to enhance the quality for severe malaria case management and improve healthcare service delivery. The study explored statistical models for analysing the correlated malaria survey data by assessing the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI), and prediction. In addition, modelling using competing risk survival analysis approach was implemented. The statistical models examined will be used as a source of information to other statisticians and researchers with similar research problems.

Through this work, predictors of the inpatient health workers' knowledge about severe malaria treatment policy were identified. This will inform the malaria program on key interventions to enhance malaria case management policy. In addition, based on the prediction of knowledge at subnational levels and the spatial maps, the program can employ focused multidisciplinary interventions. Lastly, the factors affecting hospital LOS for patients admitted with suspected malaria can be targeted to enhance the quality of care.

1.6 Scope of the study

This study utilized retrospective inpatient malaria survey data collected from 2016 to 2019 to evaluate malaria case management in Kenya. The study focused on the statistical methodologies for modelling correlated data while adjusting for health facility and county structures.

1.7 Structure of the thesis

Chapter 1 presents an overview of correlated data and the consequences of ignoring correlated data. In addition, statistical methods for analysing correlated data have been discussed. The

chapter has also described the problem statement, research questions, objectives and justification of the study. It ends with the structure of the thesis.

Chapter 2 presents an overview of malaria burden, epidemiology, and strategic interventions. The second section has elaborated on various statistical models of correlated data. It provides an overview of correlated data and the impact of ignoring correlation in data. The theoretical approaches to statistical analysis of the correlated data have been discussed and supported by relevant studies. These include: marginal, conditional, and survival models. It further discusses the approaches to spatial correlated data. The previous statistical methodologies to evaluate severe malaria case management are reviewed. The chapter further provides in-depth methodological review of statistical models used in these analyses by study objectives. These models included; multilevel mixed-effects ordinal logistic regression, Bayesian hierarchical spatial modelling, and competing risk in correlated survival data. Lastly, the chapter highlights the gaps identified in the literature.

Chapter 3 presents the study area, design, populations, and sample size including the sampling procedure and data collection procedures. It further describes data management process for secondary analysis, analytical approaches and study variables. The chapter further describes the exploratory data analyses performed, statistical analysis implemented per study objective and concludes with ethical consideration.

Chapter 4 presents the results for the objective: to investigate the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI), and prediction, adjusting for health facility and county structures in Kenya. This was demonstrated using multilevel mixed effect ordinal and binary logistic regession modelling to

determine predictors of the inpatient health workers' knowledge about artesunate-based severe malaria treatment recommendations in government and faith-based organisation hospitals in Kenya. It begins with background information, followed by the methods section that includes a brief description of the data sources. Further, it presents the results, discussion of the results and a conclusion. The work presented in this chapter has been published by *Malaria Journal* (Machini *et al*., 2020).

Chapter 5 presents the results for the objective: to apply Bayesian hierarchical approach to analyse polytomous data, adjusting for county structures in Kenya. A Bayesian hierarchical ecological spatial model beyond predictor analysis was performed to test for the best fitting spatial effects model to predict subnational levels of health workers' knowledge of severe malaria treatment policy, artesunate dosing, and preparation in Kenya. It begins with an introduction; discusses the methodology used, national standards, the implementation context, and data sources; and describes the outcomes and factors examined.

The chapter further presents the summary and exploratory analysis, Bayesian methods for ordinal logistic models, and Bayesian statistical inference. It also presents the study results, provides three comparative hierarchical models that were fitted, as well as the spatial random effects of the posterior means of the probability of health workers having high knowledge of severe malaria treatment policy, artesunate dosing, and preparation, respectively, overlaid on a map showing all the counties in Kenya. The chapter concludes with a discussion and a conclusion on the Bayesian hierarchical ecological spatial modelling. The work presented in this chapter has been published by *BMJ Open* (Machini B, *et al*., 2022).

Chapter 6 presents the results from the objective: to model factors associated with the LOS among severe malaria patients using a competing risk survival analysis approach, adjusting for health facility structures in Kenya. Factors associated with hospital length of stay for patients admitted with suspected malaria, were modeled using a competing risk approach in Kenya. The chapter begins with an introduction, discusses methods to include the description of data, presents a standard case for uncomplicated and severe malaria, and describes the outcomes and factors examined. The chapter provides statistical analysis, involving competing risk analysis, based on both a subdistribution model and a cause-specific model. This chapter presents the descriptive results, followed by a multivariate analysis for the cause-specific hazard and subdistribution hazard for patients admitted with suspected malaria, with length of stay adjusted for other factors at the various stages of hospitalisation. The chapter concludes with a discussion on the competing risk analysis and conclusions. The works presented in this chapter has been published by *BMJ Open* (Machini B, *et al*., 2022).

Finally, Chapter 7 concludes with the overall discussion, conclusion and recommendations of the study by objectives. It further provides the strengths and limitations and implications for future research.

CHAPTER TWO

LITERATURE REVIEW

This chapter provides an overview of malaria burden, epidemiology and strategic interventions in Section 2.1. Integrative review of statistical models for correlated data is presented in Section 2.2, followed by the previous statistical methodologies to evaluate severe malaria in Section 2.3. Section 2.4 presents a broad methodological review of the statistical models by study objectives and ends by identifying the gaps in the literature in Section 2.5.

2.1 Overview of malaria

2.1.1 Malaria burden and epidemiology

Globally, malaria is a major issue especially in underdeveloped countries. Malaria was responsible for an estimated 229 million cases and 409,000 fatalities in 2019, with the WHO African Region responsible for 94% of all cases (WHO, 2019a). Since three-quarters of the population is susceptible to malaria, the disease is a public health and socioeconomic issue in Kenya. Largely, the country's malaria prevalence was 6% (KMIS, 2020).

There are four main locations of malaria transmission in Kenya: the highlands, lakes, coasts, seasonal transmission, and areas with minimal risk of infection (MoH, 2019b). Low-risk zones have lower transmission intensity than the endemic zone, which has the maximum transmission intensity. However, throughout time, malaria endemic areas have decreased, and areas with low transmission rates have increased (Figure 2.1). Most counties have a prevalence rate of less than 1%, as shown by the malaria transmission risk map (Figure 2.2).

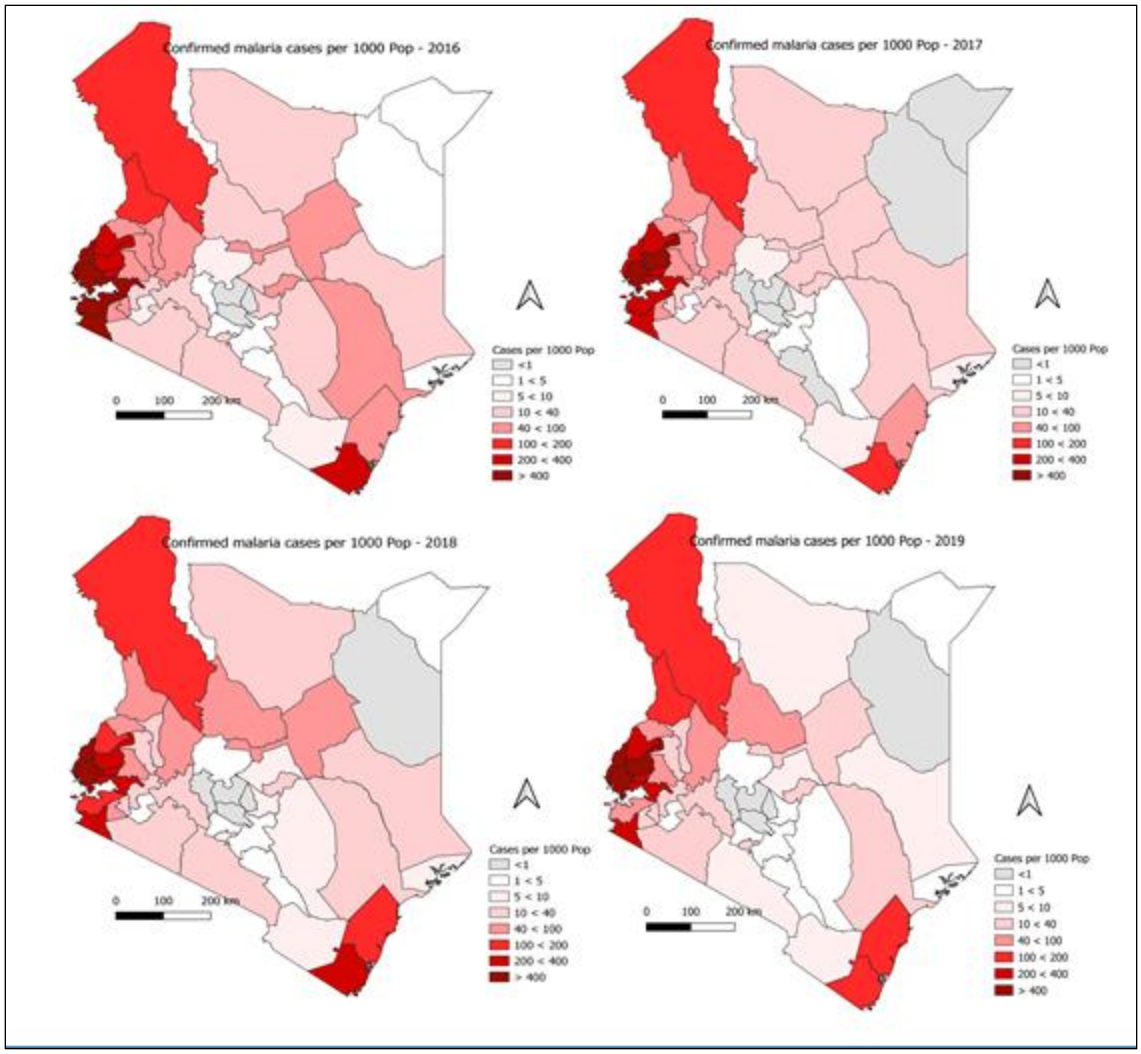


Figure 2. 1 Kenya malaria incidence maps, 2016-2019

Source: MoH, 2019

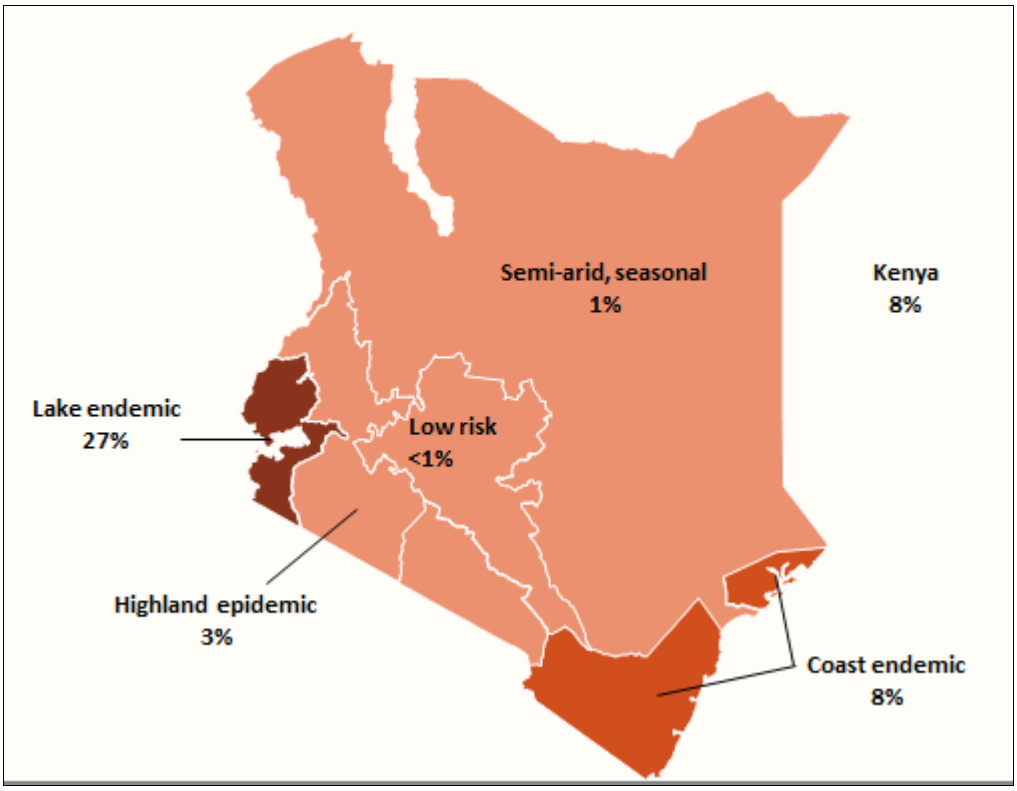


Figure 2. 2 Malaria prevalence in Kenya by zones, 2015
Source: Kenya Malaria Indicator Survey, 2015

2.1.2 Malaria strategic interventions

Malaria case management is conducted in all of the epidemiological zones, regardless of their geographic location (MoH, 2015b). Malaria prevention, diagnosis, and treatment are all included in the first pillar of the Global Technical Strategy (GTS) for 2016–2030 (WHO, 2015a). By 2023, Kenya has set a goal to reduce malaria incidence and deaths by at least 75% compared to 2016 levels (MoH, 2019b). One of the strategic objectives is to treat all suspected malaria cases in Kenya in accordance with the country's malaria treatment guidelines by 2023, and one way to do this is to improve the capacity of the country's integrated malaria case management and the quality assurance of malaria testing (MoH, 2019b). Health care providers need to be trained in

malaria treatment and diagnosis, and malaria medication must be available at reasonable prices as part of effective case management techniques (MoH, 2015b; MoH, 2021).

National malaria treatment recommendations specify that, “all suspected malaria patients should be tested using rapid diagnostic tests (RDTs) or malaria microscopy prior to treatment,” and only those individuals that test positive should be treated with the recommended antimalarial. To diagnose malaria parasitically, microscopy is considered “the gold standard,” due to its ability to identify parasite species as well as density (Moody, 2002; Feleke, Tarko and Hadush, 2017; WHO, 2009). RDTs can also detect even low levels of parasites (MoH, 2020b; Ugah *et al*., 2017). Empirical treatment may still be used in cases where a malaria diagnostic test cannot be carried out (MoH, 2015b; WHO, 2015b). However, depending on how it manifests clinically, malaria can be classed as uncomplicated or severe and treatment depends on the categorization as recommended (MoH, 2020a; WHO, 2012; Dondorp *et al*., 2010; Zurovac *et al*., 2014).

The severe malaria treatment policy in Kenya has recommended artemether lumefantrine (AL) for uncomplicated malaria and parenteral artesunate for severe malaria (Wilairatana, Tangpukdee, and Krudsood, 2013; MoH, 2015b; WHO, 2015b). Children weighing less than 20 kg should be given 3.0 mg/kg per dose of artesunate, while children weighing more than 20 kg and adults should receive 2.4 mg/kg of artesunate. Artesunate powder contains sodium artesunate that is a mixture of artesunic acid and sodium bicarbonate. The solution is diluted with normal saline or 5 percent dextrose solution for both intravenous (IV) or intramuscular (IM) injection administration. A range of programmatic measures have been implemented by the malaria program and it’s supporting partners to promote the policy of treatment. These include: the development of treatment guidelines, procurement and distribution of artesunate, development

and dissemination of job aids, supportive supervision and training health workers on malaria cases management (WHO, 2009; MoH, 2016; Zurovac *et al*., 2018; Worges, 2018; Eliade *et al*., 2019; MoH, 2019b; MoH, 2020a).

Protection against mosquito bites, malaria chemoprophylaxis, and vaccination have been used to prevent malaria. In Kenya, the use of long-lasting insecticide-treated nets, indoor residual spraying, and larval source management have been exploited (WHO, 2007; Rek *et al*., 2020; WHO, 2017a). In addition, Intermittent Preventive Treatment in pregnancy (IPTp) throughout the antenatal period is practiced to reduce the negative health impacts of MIP (MoH, 2019b; WHO, 2019b). The malaria vaccine is one of the most recent advancements in the fight against malaria in young African children (MoH, 2020a; Mahase, 2021; Gumulira, 2021) targeting the Plasmodium falciparum, which causes more than 90% of severe malaria worldwide (Snow, 2015 and Marteau *et al*., 2021; Zekar and Sharman, 2020).

Malaria elimination is becoming more common around the world (WHO, 2016). The Kenya Malaria Strategy adopted a goal of eliminating malaria in areas that had previously reported low or no malaria transmission, in accordance with the national health sector strategy (MoH, 2019b; WHO, 2015a). Kenya prioritized four counties (Laikipia, Nyeri, Kirinyaga, and Nyandarua) in the low-risk zone for malaria eradication efforts (MoH, 2019b).

Despite WHO providing information on global surveillance standards to strengthen surveillance systems, many countries with a high malaria burden have feeble surveillance systems (WHO, 2018; WHO, 2021). One objective in the KMS 2019-2023 has is to strengthen malaria surveillance and to use the information to improve programmatic decisions. The program

relies on the routine national Kenya Health Information Systems (KHIS) and surveys to monitor and evaluate malaria interventions (MoH, 2019a; MoH, 2019b).

2.2 Integrative review of statistical models for correlated data

2.2.1 Introduction to correlated data

Correlated data with a multilevel structure are regularly encountered in public health research. In multilevel structures, population elements are grouped into aggregates, and often, data is generated from both the individual and the aggregated groups. An example would be information from subjects from the same area in a given county. In this study, knowledge levels from health workers from the same hospital may be similar than knowledge levels of other health workers from different hospitals. The patients or health workers can be level-one observation units, the health facility can be level-two observation units and county can be level-three observation units. In this case we refer to levels of measurement and observational units at each level. These types of hierarchies are likely to generate correlations in data. Correlation describes the relationship between the individuals in datasets. Ignoring correlation leads to inaccurate standard errors and erroneous statistical inferences (Sainani, 2010). In categorical data, the correlation coefficient is a poor measure of association because it is guarded by the mean parameters (Liang and Zeger, 1993). There is a need to account for the dependence in correlated data, and this is determined by choice of the model, estimator, and summaries (Cameron and Miller, 2015).

2.2.2 Approaches to statistical analysis of correlated data

The general linear models that assume independent and uncorrelated Gaussian residuals are often implemented with an objective of describing the relationship of a response with explanatory variables. The generalized linear models extend these approaches to include binary responses,

count data, and timing of event data. Approaches used to analyse correlated datasets have often ignored clusters in time and space, with clusters being reduced to independent observations, regression analysis implemented using fixed effects analysis of variance approaches, and ignoring the clusters (Galbraith, Daniel and Vissel, 2010).

In the presence of correlated outcomes, marginal models and conditional models can be implemented. The marginal expectation is modelled as a function of independent variables and the within-cluster dependence are modelled separately (Lee and Nelder, 2004). Marginal models are estimated using generalized estimating equations (GEEs), or marginalised multilevel regression or hierarchical linear models. GEE evaluates longitudinal and other correlated response data, principally if the outcomes are binary (Luo *et al.*, 2021; Liang and Zeger, 1986).

GEE methods do not explicitly model variations between clusters. Instead, it focuses on the similarity of the residuals in the corresponding cluster. However, in analysing clustered data using GEE, the standard errors tend to be erroneous and the regression coefficients for predictor variables lower than the subject-specific models. Thus, the GEE approach cannot handle multilevel clustering inherent in survey data since explicit terms for the between cluster variation are not present. GEE only allows fixed effects, and treats variability within groups as a nuisance term by incorporating it into the intercept (Hanely *et al.*, 2003). This approach has been used to evaluate the impact of President's Malaria Initiative (PMI) in reducing malaria burden in sub-Saharan Africa between 2004 and 2014 analysed data using generalized estimating equations. The findings showed that PMI contributed significantly to increasing the coverage of malaria control interventions and reducing under-five mortality in SSA (Ye and Duah, 2019).

The multilevel models address the problem of correlated error by having both fixed and random effects during modelling. Fixed-effect models include subject-specific regression coefficients and the error term assumes normality among observations from the same subject (Moen et al., 2016), while the mixed-effect model includes a random effect as an individual subject (O'Brien and Fitzmaurice, 2005). Multilevel models partition variance into between-group and within-group effects, resulting in valid inferences (LaHuis et al., 2014). Conditional models have coefficients for predictors that are random variables and that are unique to each subject. Marginal models are population-average models whereas conditional models are subject-specific (Muff, Held and Keller, 2016). Multilevel models have been used to examine the effects of individual and community-level bed net usage on malaria prevalence among under-fives in the Democratic Republic of Congo (Levitz et al 2018)

2.2.3 Statistical models for correlated survival analysis

Survival data is applied when assessing the expected time to some event of interest (Kartsonaki, 2016). Correlated survival data occur as the result of repetitive events that a person experiences, or when observations are grouped together. The cluster effect leads to dependencies among outcomes within each cluster, and ignoring the cluster effect narrows the confidence interval (CI) of the estimated rate, which can lead to misleading conclusions. Therefore, it is necessary to adjust the correlation within the individual (Amorim and Cai, 2015).

Austin, (2017) describes approaches of evaluating multilevel survival data: the frailty, piecewise exponential (PWE), and discrete time models with mixed-effects. Frailty models are extended to incorporate cluster-specific random effects to account for within-cluster homogeneity in outcomes (Gorfine and Hsu, 2011). The shared frailty model has added appeal

that frailty can be used to model intragroup correlation (Dixon, Darlington and Desmond, 2011). The Fine-Gray model can be used to introduce a shared frailty structure to examine heterogeneity (Fine and Gray, 1999; Katsahian, 2011). This was extended to the clustered data setting so that the CIF can be estimated by adjusting for prognostic aspects while accommodating correlation within clusters (Zhou *et al*., 2012). SDH regression model with multivariate frailty has been used to analyse clustered data based on hierarchical likelihood estimation approach developed to fit models and draw inferences (Ha *et al.*, 2016). In PWE approach, the discrete time in survival models measure discrete values and can simply integrate the multilevel structure of the data (Austin, 2017).

Various studies have been done using this approach. One of these studies was done in Ghana to analyze relapsed cases using the frailty regression method and considered risk heterogeneity at the individual level (Cairns *et al*., 2015). In another study, South-East Asia Quinine Artesunate Malaria Trial, examined associations with cause-specific hazard (CSH) analysis using Cox proportional hazards regression and Subdistribution-Hazard (SDH) ratios, using Fine and Gray model with cumulative incidence accounting for competing risks (Keene *et al.*, 2018). In addition, some statistical models have been suggested to account for intra-subject correlation as a result of repeated events in survival analysis based on the risk set such as the independent increment, marginal and conditional models (Amorim and Cai, 2015; Villegas, Julià and Ocaña, 2013).

2.2.3.1 Independent increment model

The independent increment model assumes that the correlation between event times for a person can be explained by past events or unaffected by earlier events that occurred to the same subject

so baseline hazards for all events are common (Anderson and Gill, 1982). This implies that the time increments between events are conditionally uncorrelated, given the covariates. It is a suitable model when correlations among events for each individual are induced by measured covariates (Amorim and Cai, 2015).

These models can be fitted as a Cox model with the addition of a robust SE estimator, and hazard ratios are interpreted as the effect of the covariate on the recurrence rate over the follow-up period (Guo, Gill and Allore, 2008). The assumption of mutual independence of the events within a subject is equal to the assumption of independent increments in the counting process inside each individual. To illustrate this without considering the rank of recurrence, let us assume a recurrent event for the i^{th} subject follow a proportional hazard model, where the hazard function (Amorim and Cai, 2015) is:

$$h_i\,(t) = h_0\,(t)\exp\big(\beta'\,Z_i\,(t)\big). \tag{2.1}$$

where the risk is set as,

$$Y_{ij}\,(t) = I\big(X_{ij-1} < t < X_{ij}\big), \tag{2.2}$$

where the risk set time *t* is,

$$\sum_{ij} Y_{ij}\ (t). \tag{2.3}$$

2.2.3.2 Conditional model

The conditional model assumes that a subject is not at risk for a successive event till a prior event occurs (Prentice, Williams and Peterson, 1981). Let us assume that a subject is not at risk for the j^{th} event until the subject experiences event j-1. This model allows the baseline hazard to vary from recurrence to recurrence, the hazard function for the j^{th} event for the i^{th} subject (Villegas, Julià and Ocaña, 2013):

$$h_{ij}(t) = h_{0j}(t)\ \exp\Big(\beta'_j Z_i(t)\Big). \tag{2.4}$$

where,

$$Y_{ij}(t) = I\big(X_{ij-1} < t < X_{ij}\big), \tag{2.5}$$

where the risk set at time t is different for each j,

$$\sum_{i,} Y_{ij}\ (t). \tag{2.6}$$

According to the definition of the risk set, the number of subjects is dramatically decreased as j increases. Stable coefficient estimates cannot be obtained for higher ranks of j. The hazard function at time t for the j^{th} recurrence is conditional to the entire previous failures. This model allows different baseline hazards; therefore, estimations for the current event may be affected by earlier events (Villegas, Julià and Ocaña, 2013).

2.2.3.3 Marginal model

The marginal model considers all recurrent events of the same subject as a single counting process and does not require time-varying covariates to reflect the past history of the process

(Wei, Lin and Weissfeld, 1989). These models can also be fit using stratified models with robust SEs (Amorim and Cai, 2015).

The risk set indicator is:

$$Y_{ij}(t) = I\big(X_{ij} \geq t\big), \tag{2.7}$$

where the risk set at time *t* for the j^{th} recurrence is,

$$\sum_{i,} Y_{ij}(t), \tag{2.8}$$

where the hazard function of the marginal model for the j^{th} event for the i^{th} subject is,

$$h_{ij}(t) = h_{0j}(t)\exp\left(\beta'_j Z_i(t)\right). \tag{2.9}$$

Observation from subjects may be correlated, however, the β estimation are reliable in the case of two events (Lin, 1994).

2.2.4 Approaches to spatial correlated data

Spatial correlation is exhibited by the existence of spatial dependence among observations and failure to account for the underlying correlations among data tends to narrow the confidence intervals. Mixed-effects models spatial correlations by using random effects for geostatistical data and conditional autoregressive (CAR) structure for areal data (Clayton and Kaldor, 1987; Cressie, 1993; Breslow and Clayton, 1993; Besag, York and Mollie, 1991; Waller *et al*., 1997).

Spatial autocorrelation (SAC) happens when the obsevations sampled at neighbouring settings exhibit similar values and this complicates the analysis of spatial data. The SAC can be checked using Moran's I plots (Legendre and Legendre, 1998), or Geary's C correlograms and

semi-variograms. The presence of spatial autocorrelation poses a challenge drawing inferences (Lennon, 2000; Dormann, 2007; Anselin, 2002).

These approaches have been used when analysing spatial data. In Southern Africa, a study was done using geospatial tools such as geographic information system and spatial statistic methods to detect spatial patterns of malaria to identify malaria hot spots for better planning and management of cases at the country level (Gwitira *et al.*, 2018).

For correlated data, Generalized Linear Models (GLMs) are not sufficient (Zeger, Liang and Albert, 1988). There are other various statistical approaches such as: autocovariate regression, spatial eigenvector mapping (SEVM), generalized least squares (GLS), conditional autoregressive models (CAR), simultaneous autoregressive models (SAR), generalized linear mixed models (GLMM), and generalized estimation equations (GEE). These are discussed in details in the next sections.

2.2.4.1 Autocovariate regression

The autocovariate regression captures spatial autocorrelation arising from endogenous progressions in additional covariates included into a generalized linear model (GLM). This model estimates how much the response variable at any one site will reflect the response values at surrounding sites as an extension of the GLM model done by adding a distance-weighted function of neighbouring response values to the model's explanatory variables (Dormann *et al*., 2007).

2.2.4.2 Generalized least squares

Generalized least squares (GLS) methods fit a variance-covariance matrix constructed on the non-independence of spatial observations using parametric functions (Dormann *et al*., 2007).

2.2.4.3 Autoregressive models

Simultaneous autoregressive models (SAR) and conditional autoregressive models (CAR), model the errors produced and include weight matrices that state the strength of interaction between neighbouring localities. Both models incorporate spatial autocorrelation using neighbourhood matrices, stipulating the association between the response values or residuals at each locality and those at neighbouring localities (Dormann *et al*., 2007).

2.2.4.4 Spatial generalized linear mixed models (GLMM)

Spatial generalized linear mixed models (GLMM) have linear predictors with random effects and within-group errors that may be spatially autocorrelated (Müller, Scealy and Welsh, 2013).

2.2.4.5 Conditional model

The conditional model has a term in the model for each group resulting to group effect that can either be modelled as fixed or random effect. The mixed-effects model shows the unobserved elements responsible for the similarity between certain observations. In a longitudinal survey, one can use a random effect, specific to each subject, expressing how much a subject's trajectory is translated as compared to what is expected according to its characteristics (Tutz and Oelker, 2017; Agresti, 2007; Pendergast, 1996; Allison, 1999).

2.2.4.6 Generalized estimating equations

Generalized Estimating Equations (GEE) divide the data into smaller clusters prior to modelling the variance-covariance relationship. For responses that are repeatedly measured in time or space, the GEE method uses a parameterized correlation matrix to consider the intra-cluster correlation of the sample units, but the inter-cluster correlation is considered zero (Bender, Augustin and Blettner, 2005). For example:

Let $Y_{ij}, j = 1, \dots, n_{j,}\ \ i = 1, \dots, K$ denote the *j*th measurement on the *i*th subject. There are n_j measurements on subject *i* and $\sum_{i=1}^{k} n_i$ total measurements.

The vector of measurements on the *i*th subject is $Y = \left[Y_{I1,\dots,}Y_{ini}\right]'$, corresponding vector of means $\mu_i = \left[\mu_{i1,\dots,}\mu_{ini}\right]'$ and *V* is the covariance matrix of *Y*.

The vector of independent variables for the *j*th measurement on the *i*th subject is $X_{ij} = \left[x_{ij1,\dots,}x_{ijp}\right]'$.

The GEE for estimating the $p \times 1$ vector of regression parameters β is an extension of the independence estimating equation to correlated data (Liang and Zeger, 1986):

$$S(\beta) = \sum_{i=1}^{K} \frac{\partial \mu'_i}{\partial \beta} V_i^{-1}\left(Y_i - \mu_i(\beta)\right) = 0 \qquad (2.10)$$

Since

$$g\left(\mu_{ij}\right) = x'_{ij}\beta, \qquad (2.11)$$

where *g* is the link function, the $p \times n_i$ matrix of partial derivatives of the mean in relation to the regression parameters for the *i*th subject:

$$\frac{\partial \mu'_i}{\partial \beta} = \begin{bmatrix} \frac{x_i 11}{g'(\mu_{i1})} & \cdots & \frac{x_{in_i} 1}{g'(\mu_{in_i})} \\ \vdots & & \vdots \\ \frac{x_i 1p}{g'(\mu_{i1})} & \cdots & \frac{x_{in_i} p}{g'(\mu_{in_i})} \end{bmatrix} \quad (2.12)$$

2.2.4.7 The Bayesian approach

Bayesian approach allows prior beliefs of information to be included in calculating expected values based either on a CAR or an auto-logistic implementation (Dormann *et al*., 2007). Bayesian hierarchical regressions lay prior distributions on exposure-specific regression coefficients to stabilize assessment and incorporate available prior knowledge (Ferrari and Dunson, 2021). Bayesian networks employ random effects in modelling correlation in sample elements without increasing the Type I error (Kruschke and Vanpaemel, 2015). In utilizing this approach, a study done in China identified the presence of spatial effects in influencing the effects of built environment on car ownership and use by combining multilevel Bayesian model and CAR model to control for spatial autocorrelation (Wang *et al*., 2018).

2.2.4.8 Spatial eigenvector mapping

Spatial SEVM is grounded on the spatial preparation of data points that can be decoded into independent variables that capture spatial effects at different spatial resolutions. At analysis stage, eigenvectors that best decrease spatial autocorrelation in the residuals are selected overtly as spatial predictors. An eigenvector denotes a specific spatial patterning that result in varied spatial autocorrelation over space (Dormann *et al*., 2007). This approach has been used among other studies to investigate spatial factors determining malaria occurrences in Korea. Multilevel model was used to simultaneously analyze the variables in different spatial scales, and

eigenvector spatial filtering to explain the spatial autocorrelation in the malaria occurrence data (Kim and Kim, 2019).

2.3 Previous statistical methodologies to assess severe malaria case management

In this section, previous studies are reviewed to explore the various statistical methodologies used to assess severe malaria case management. Most of the studies reviewed were conducted in Africa. The studies provided the clinical setting in which the measures were applied, the statistical models computed, the results and conclusions drawn.

A prospective cohort study explaining malaria diagnosis and treatment patterns in adult inpatients admitted to Uganda's medical and gynecological wards used multivariate logistic regression to identify risk factors for missing day 1 antimalarial drug administration. One in four inpatients delayed the start of hospital malaria treatment by at least one calendar day. The study concluded that the hospital should encourage prompt availability of malaria test-results to promote the timely initiation and completion of anti-malarial treatment to improve the quality of care for hospitalized malaria patients (Kiguba, Karamagi and Bird, 2021).

A cross-sectional study was done in Northern Nigeria to evaluate hospital and health worker readiness for policy implementation, health workers' knowledge about case-management recommendations, and the quality of inpatient management for patients admitted with suspected malaria following artesunate treatment. The study used descriptive statistics to conclude that translation of new treatment policy for severe malaria into inpatient practice was compromised by lack of malaria diagnostics, stock-outs of artesunate, and suboptimal health workers' practices (Ojo *et al*., 2020).

A cross-sectional descriptive study conducted in Sudan to assess hospital readiness, healthcare provider knowledge, and care for patients with severe malaria showed overall adherence to 2.2% of guidelines at the hospital level. It was concluded that the management of severe malaria at hospital level was suboptimal and most patients were not treated according to the national guidelines (Elnour *et al*., 2019).

A study done to evaluate prescription compliance according to the WHO recommendation in eight public health facilities in Ghana and Uganda, used log-binomial regression model to identify predictors for compliance. Standard errors were adjusted for clustering of patients in health facilities. The study found that injectable artesunate was the most commonly prescribed medicine in the management of severe malaria. However, adherence to the WHO recommendation of at least three doses of injectable anti-malarials in 24 hours followed by a full course of ACT was low, at less than 30% (Ampadu, Asante and Bosomprah, 2019).

A cross-sectional health facility survey done to monitor levels and trends in health system readiness to implement new treatment policies; health workers' coverage with interventions and their treatment knowledge; and malaria case-management practices for patients admitted to paediatric and medical wards utilised cluster-adjusted comparisons and reported improvement of general quality of malaria case-management, albeit with a few programmatic gaps to optimise policy translation (Zurovac *et al.,* 2018).

A similar study assessed the level of knowledge of healthcare practitioners on the preparation of injectable artesunate for the treatment of severe malaria in public health facilities in Tanzania and performed descriptive statistics to summarize the demographic characteristics while, comparison of categorical data was done by Chi-squared test. The study concluded that

the level of knowledge among healthcare providers on preparation of injectable artesunate in public health facilities was low (Mikomangwa *et al.,* 2019).

A study conducted in Angola to assess the readiness of healthcare facilities for the diagnosis and treatment of malaria and to assess the quality of malaria case management used a logistic regression model to examine the relationship between outcome and explanatory variables. Spatial heatmaps generated by Gaussian kernel smoothing were created for each province. The findings highlight differences and similarities between provinces and emphasized the importance of continuous training and monitoring of health care staff in malaria case management, especially in areas with low malaria infection. (Plucinski *et al.,* 2017).

A similar study evaluated the practice of diagnosing and treating malaria patients admitted to healthcare facilities in Malawi. Chi-square was used to test the differences between the categorical variables. The risk ratio was estimated using a log-binomial regression model using the generalized estimation equation approach to explain the correlation of repeated measurements within the same healthcare facility. In summary, most patients diagnosed with severe malaria received recommended intravenous treatment and suggested the need to improve healthcare worker recognition and documentation of severe signs and symptoms to enhance the criteria for severe malaria diagnosis (Shah *et al*., 2016).

A descriptive cross-sectional study of inpatient malaria case management practices in Kenya and concluded that there were improvements in inpatient malaria care, high rates of presumptive treatment for test negative children and likely over-use of injectable anti-malarial drugs were observed (Amboko *et al*., 2016).

A case study done in Hawassa to detect malaria mortality in hospital risk factors, and to model and simulate related risk factors, analysed data using classical logistic regression and Bayesian logistic regression approaches. The two approaches were compared using standard errors of model parameters and revealed that Bayesian modelling approach gave estimates with smaller standard errors than the classical approach (Gute *et al.*, 2015).

Morbidity and mortality can be reduced through appropriate diagnosis and treatment. Bayesian approach was applied to model the malaria-related hospital mortality risk. The risk of dying in hospital was lower in the dry season and for children who travelled a distance of less than 5 kilometres to the hospital, but increased for those who were referred to the hospital. The results also indicated significant differences in both structured and unstructured spatial effects, and the health facility effects revealed considerable differences by type of facility or practice (Kazembe *et al*., 2008).

Keene *et al.* (2018), modelled length-of-stay (LOS) for severe malaria cases accounting for the competing event of death. Clinical factors on admission and during hospitalisation influence LOS in severe malaria, presenting targets to improve health and service efficiency. Artesunate has the potential to increase LOS (Keene *et al*., 2018). Similar studies were conducted to determine under-researched factors related to LOS in German-imported *P. falciparum* hospitals. This retrospective observational study used multivariate Cox proportional hazards regression over time to discharge as an endpoint for inpatient adults to identify factors. In summary, early recognition of disease severity, along with targeted supportive care, can lead to the prevention of overt organ failure, lowering LOS and reducing pressure on bed capacity (Hoffmeister, 2021).

In Ghana, where follow-up was conducted, the survival curve was flattened, indicating that unexposed children were not infected with malaria. They analyzed relapsed cases using the frailty regression method and considered risk heterogeneity at the individual level. Interventions that reach primarily urban populations would limit the overall impact, as some urban populations were virtually risk-free, even in endemic areas (Cairns *et al*., 2015).

Another longitudinal study in southwestern Ethiopia used a spatially correlated CAR vulnerability model to examine the impact of hydropower plants on malaria infection. The parameters of the model were estimated using the Bayesian framework using the Markov Chain Monte Carlo (MCMC) approach. Parameter estimation using the neighborhood assumption (spatial-correlated CAR frailty model) proved to be economical. Malaria control intervention programs need to take into account the spatial variation of malaria infection in order to achieve sustainable and efficient malaria control in the study area (Wondaya *et al*., 2016).

2.4 Methodological review of statistical models by study objectives

This section presents a statistical approach to multilevel mixed-effect ordinal logistic regression modelling, Bayesian hierarchical spatial modelling and competing risk analysis in correlated survival data.

2.4.1 Multilevel mixed-effects logistic regression modelling

Multilevel mixed-effects ordinal logistic regression modelling was implemented to adjust for the impact of correlation on dichotomous and polytomous data in terms of estimation of parameters, precision measures (SE & CI) and prediction while adjusting for county structures. In this section, the basics of logistic regression modelling are discussed and extended to polychotomous logistic regression modelling. Ordinal logistic regression has been explicitly discussed in section

2.4.1.3 and finally, the main model, the multilevel mixed-effects ordinal logistic regression is discussed (section 2.4.1.5).

2.4.1.1 Logistic regression model

Logistic regression models discrete outcomes using predictor variables. It can be either binary or multinomial logistic regression (Hosmer, Lemeshow and Sturdivant, 2013). Binary logistic regression estimates the success probability when the exploratory variables are given, and if the dependent variables are ordered, the model applied is ordinal logistic regression (Sperandei, 2014). The binary logistic regression can be termed as simple or multiple depending on the predictor variables (Tabachnick and Fidell, 2014; Bewick, Cheek and Ball, 2005). The simple and multiple logistic regressions are discussed in the following discussion (Hosmer, Lemeshow and Sturdivant, 2013).

The simple logistic regression model is given as:

$$ln\left[\frac{P(Y)}{1-P(Y)}\right] = \beta_0 + \beta_1 X_1 \,, \tag{2.13}$$

and the multiple logistic regression as:

$$ln\left[\frac{P(Y)}{1-P(Y)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p, \tag{2.14}$$

where *ln* is the natural logarithm, *Y* is the dichotomous outcome, *p* is the expected probability that the outcome is present; X_1 through X_p are predictor variables; β_0 is the intercept, $\beta_1, \ldots, \beta_p$ are the regression coefficients. The outcome is the expected log of the odds that the outcome is present and is written as:

$$p = \frac{e^z}{1 + e^z}, \tag{2.15}$$

where,

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p.$$

The model assumes that the observations are independent and that the natural logarithm of the odds ratio and the measurement variables are linearly related, but does not assume that the measurement variables are normally distributed (Hosmer, Lemeshow and Sturdivant, 2013).

The logistic regression model assumes that the dependent variables are from Bernoulli or binomial distribution depending on the number of trials. The true conditional probabilities are a logistic function of the independent variables, no important variables are omitted, no extraneous variables are included and the independent variables are measured correctly. Logistic regression requires a large sample size and assumes linearity of independent variables to log odds. The independent variables are not linear combinations of each other. The model assumes perfect multicollinearity which makes estimation impossible, while strong multicollinearity makes estimates innaccurate (Yu, Jiang and Land, 2015). The MLE is used to estimate the slopes and intercept of the best-fitting equation in a multiple logistic regression (Harrell, 2015). Menard (2002) defines the appropriateness of the model involves testing whether the model assumptions hold by assessing the distribution of the residuals. The Hosmer-Lemeshow test assesses the goodness of fit of a model and allows for any number of explanatory variables, either continuous or categorical.

2.4.1.2 Polychotomous logistic regression

Polychotomous outcome can be classed as either multinomial or ordinal categories. In polychotomous logistic regression analysis, more than one logit model is fit to the data, as there are more than two outcome categories. Multinomial logistic regression is used to model nominal or categorical outcome variables, that cannot be ordered in any meaningful way and for which there are more than two categories in which the log odds of the outcomes are modelled as a linear combination of the predictor variables. When using clustered data, the multinomial logistic regression model should be used to obtain the parameter estimates, and a clustered bootstrap approach should be used to obtain correct standard errors (Bae *et al*., 2016; Dell-Kuster *et al*., 2018). However, the multinomial logistic regression model is not used when the dependent variable is ordered instead ordinal logistic regression is used.

2.4.1.3 Ordinal logistic regression model

Ordinal logistic regression is implemented when a dependent variable has more than two categories and the values of each category have a meaningful progressive order where a value is definitely 'higher' than the previous one. The ordered logit model can be used to estimate the probability that the unobserved variable Y falls within the various threshold limits. For instance, using three cut-off terms (outcome levels), let the estimated value of Z and the assumed logistic distribution of the disturbance term be defined as (Williams and Quiroz, 2020):

$$Z_i = \sum_{k=1}^{K} \beta_k X_k \tag{2.16}$$

The probability of Y can be estimated as:

$$P(Y = 1) = \frac{1}{1 + \exp(Z_i - k_1)}, \tag{2.17}$$

$$P(Y = 2) = \frac{1}{1 + \exp(Z_i - k_2)} - \frac{1}{1 + \exp(Z_i - k_1)}, \tag{2.18}$$

and

$$P(Y = 3) = \frac{1}{1 + \exp(Z_i - k_2)}. \tag{2.19}$$

2.4.1.4 Types of ordinal logistic regression models

There are several ordinal logistic regression models such as the proportional odds model (POM) [the partial proportional odds model, without restrictions (PPOM-UR) and with restrictions (PPOM-R)]; the continuous ratio model (CRM), and the stereotype model (SM). The most frequently used ordinal logistic regression model in practice is the constrained cumulative logit model called the proportional odds model (Das and Rahman, 2011).

The ordinal logistic regression model assume the response variable is ordinal, the explanatory variables are continuous or categorical, there is no multicollinearity, and the odds are proportional (an independent variable has an identical effect at each cumulative split of the ordinal dependent variable). To check whether the proportional odds (parallel lines) assumption holds, the Brant Test is used (Brant, 1990). The generalized ordered logit model can fit three special cases of the generalized model: the proportional odds/parallel-lines model, the partial proportional odds model, and the logistic regression model (Williams and Quiroz, 2020).

In complex survey sampling designs involving the use of diverse strata, clustered sampling techniques, and unequal selection probabilities, it is incorrect to conduct the proportional odds model analysis for the ordinal response variable without taking the survey sampling designs into consideration. Nevertheless, multilevel models can analyse complex sampling survey data when data structures are hierarchical (Hahs-Vaughn, 2005; Thomas and Heck, 2001; Liu and Koirala, 2013).

The basic objective of ordered logit models is calculation of accumulative probability for dependent variable being greater than the j^{th} category (Brant, 1990; Liu and Agresti, 2005). The POM assumption is that the effect of independent variable is same across all the categories of dependent variable (McCullagh, 1980). Generalized model is used to relax the proportional odds assumption (Arfan and Sherwani, 2017).

For a response variable Y with C categories and a set of predictors $X = (X_1, \dots, X_k)'$ having the effect, the parameters $\beta = (\beta_1, \dots, \beta_k)'$ and the probability of response variable being less than or equal to category j can be modelled by the logistic distribution as follows (Williams and Quiroz, 2020):

$$\gamma_j = P(Y \leq y_j | X) = \frac{\exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}{1 + \exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}. \qquad (2.20)$$

where *j*=1,2, 3, …*C*-1.

The above proportional odds model gives the cumulative probability Y_j of category *j*, and for the response variable with categories *C*, the last category of the cumulative probability is *C-1* and is always equal to one. The above model can also be written as:

$$P(Y \leq y_j|X) = \frac{1}{1 + \exp[-\alpha_j + (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}, \quad (2.21)$$

The probability that response variable lies in the category greater than *j* is

$$P(Y > y_j|X) = 1 - P(Y \leq y_j|X). \quad (2.22)$$

This means that

$$P(Y > y_j|X) = 1 - \frac{\exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}{1 + \exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}, \quad (2.23)$$

and that

$$P(Y > y_j|X) = \frac{1}{1 + \exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})]}. \quad (2.24)$$

The odds of response variable being less than or equal to category greater than *j* will be

$$\frac{P(Y \leq y_j|X)}{P(Y > y_j|X)} = \exp[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})], \quad (2.25)$$

and the logit model is the natural log of odds ratio being the linear function of *k* independent variables:

$$Log\left[\frac{P(Y \leq y_j|X)}{P(Y > y_j|X)}\right] = \alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK}). \quad (2.26)$$

where *j*=1,2, 3, …*C*-1.

The proportional odds model assumes that the explanatory variables have the same effect on the response variable across all the categories of the response variable. Under the assumption of proportional odds the β remains same and only intercepts contrasts for various categories of

response variable. When the sign of β is negative the interpretation is that one unit increases in predictor variable increases the log odds of being in the category greater than *j*. The cumulative logit model is expressed as (Arfan and Sherwani, 2017):

$$\gamma_j = \frac{\exp\left[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})\right]}{1 + \exp\left[\alpha_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK})\right]}, \tag{2.27}$$

where α_j are the intercepts and are different for each comparison ordinal categorical variable, and the relation between α_j is $\alpha_1 < \alpha_2 \ldots < \alpha_{c-1}$ which ensures that $\gamma_1, \gamma_2, \ldots, \gamma_{c-1}$. The slope coefficient $\beta_1,\ldots,\beta_k$ are the same for all the categories of dependent variable, for continuous variables, the slope coefficients change in log odds for one unit change in predictor and for nominal predictors the slope coefficient represent the effect of each category of nominal variable as compared to reference category (Arfan and Sherwani, 2017).

2.4.1.5 Multilevel mixed-effects ordinal logistic regression

Mixed-effects ordered logistic regression is ordered logistic regression encompassing both fixed and random effects. For multilevel data, where observations are nested within clusters (wards, hospitals) or repeatedly assessed over time, mixed effects regression models are often used to account for the inherent dependency of the data. Mixed-effects logistic regression models are defined for analysis of longitudinal ordinal outcomes, where observations are observed to be clustered within subjects. Random effects included in the model account for correlation in the clustered observations. Naturally, the error variance and the variance of the random effects are considered to be homogeneous. These variance terms characterize the within-subjects (error variance) and between-subjects (random-effects variance) variation in the data (Hedeker, Demirtas, and Mermelstein, 2009).

Multilevel models are for hierarchical nested data structure that allow error elements at different level of hierarchy and may be correlated over time (Arfan and Sherwani, 2017). For example, using two level health worker outcomes where health workers are nested within a health facility in a given county. A multilevel model estimates the residual at different levels, both health facility and county levels. Thus the total residual variance is divided into parts one for between health facility (level two units, health facility residual) and one for within health facility (between level one units, health workers' residuals).

Multilevel models provide proper standard errors when data points are not independent. The modelling approach is appropriate when the one is interested in relationships both within and between clustered groups. Multilevel models have ability of handling more than two levels in the response variable (Austin and Merlo, 2017; Kassahun *et al.*, 2014).

Multilevel logistic regression modelling allows accounting for subject clustering at high levels. Multilevel structure data has within-cluster subjects, like hospitals, which may have correlated responses as an effect of mutual circumstantial products such as the hospital setting, staff, and administration on the outcome. Two patients selected randomly from the same hospital may be highly related than the ones selected at random for the outcomes of two patients randomly selected from diverse hospitals (Austin and Merlo, 2017). If the dependent variable is ordinal in nature (high, medium, and low), the multilevel analysis is appropriate in estimating the variation in responses occurring within and between clusters in higher-level elements (Austin, 2017).

In multilevel modelling, between clusters represents random effects due to inclusion of the cluster level in the model. The multilevel ordinal logistic regression models hierarchical and

ordinal dependent variable that follows the logistic distribution and is nested with higher levels (Khiari and Rejeb, 2015). For example:

Let the Y_{cij} ordered categorical response of i^{th} individual in the cluster with C ordered categorical coded as C=1, 2, …,c. The cumulative probability for ordered response up to category c is $P_{ijc} = P(Y_{ij} \leq c)$. The multilevel random intercept cumulative log odds model for ordinal response is written as:

$$\log\left(\frac{P(Y_{ij} \leq c|X)}{P(Y_{ij} > c|X)}\right) = \alpha_c + u_{0j}. \quad (2.28)$$

This measures the odds of Y_{cij} being in the category less than or equal to C as compared to greater than the category C, and u_{0j} is the random effect of level two units and is assumed to follow normal distribution N(0, τ_0). The above model random intercept model when there is no explanatory variables. When the model has fixed explanatory variables, the model is written as

$$\log\left(\frac{P(Y_{ij} \leq y_c|X)}{P(Y_{ij} > y_c|X)}\right) = \alpha_c + X_{ij}\beta + u_{0j}, \quad (2.29)$$

where X_{ij} is the data matrix of fixed predictors, hence fixed effects βs are the same as for the simple proportional odds model.

2.4.1.6 Intra-cluster correlation (ICC)

The ICC specifies the proportion of unexplained variance that is at the cluster level. It accounts for the association of clustered data by comparing the variance within clusters to the variance between clusters:

$$ICC = \frac{\sigma_v^2}{(\sigma_v^2 + \sigma^2)}, \tag{2.30}$$

where σ_v^2 is the cluster or level-2 variance and σ^2 is the level-1 variance (Agresti, 2003).

2.4.2 Bayesian hierarchical spatial modelling

Bayesian hierarchical spatial modelling was applied to analyse polytomous data. This section introduces Bayesian approach and discusses Bayesian inference with Bayes' theorem (Section 2.4.2.2). The Markov chain Monte Carlo is elaborated in Section 2.4.2.6, followed by a model comparison in Bayesian computation. Finally, Bayesian hierarchical spatial modelling is explicitly tackled in Section 2.4.2.8.

2.4.2.1 Bayesian approach

Bayesian data analysis is an approach to statistical modelling where relative credibility of parameter values is reallocated using Bayes' rule (Kruschke & Liddell, 2018; Kruschke, 2010; Dohoo, Martin and Stryhn, 2012). Bayes' theorem provides a technique of reviewing probability estimates as more information is acquired. Prior probability is the probability before additional information becomes available while, the revised probability using the additional information is the posterior probability (Dienes, 2016). The initial uncertainty is expressed by a prior distribution of the quantities of interest. Current data and assumptions concerning how they were generated are summarised based on MLE. The uncertainty about the unknown parameters is quantified using probability, and the unknown parameters are regarded as random variables. The posterior distribution for the quantities of interest is achieved by linking the prior distribution and the likelihood (Corani *et al.*, 2017; Gelman *et al.*, 2017). The Bayesian approach allow for a

more flexible incorporation of other obstacles like observer bias, missing data, or when accounting for detection probabilities (Ma and Chen, 2018).

2.4.2.2 Bayesian inference with Bayes' theorem

Conditional probability is a vital element of Bayes' theorem is (Sánchez and StataCorp, 2017). The Bayes' rule:

$$P(B|A = \frac{P(A|B)P(B)}{P(A)}. \tag{2.31}$$

where P(A|B) is the conditional probability of A given B; P(B|A) is the conditional probability of B given A; P(B) is the marginal probability of B; and P(A) is the marginal probability of A. In sequential updating, Bayes' theorem can be used to revise the probability estimate given additional information *C*, where *P* (A|B) becomes the prior probability, as that is the estimate before observing *C*.

The posterior probability is:

$$P(A|C,B) = \frac{\mathrm{P(C|A,B)P(A|B)}}{\mathrm{P(C|B)}}, \tag{2.32}$$

Bayesian inference is a method of analysing statistical models, combining prior knowledge about the model parameters. It is built on the posterior distribution of the parameters, and provides summaries containing posterior means, their MCMC standard errors and credible intervals. It provides a more intuitive interpretation in terms of probabilities. In Bayesian testing there is no need for alpha adjustment as inferences can be drawn many times without risking increased likelihood of false conclusions (Li and Fearnhead, 2018; Dienes, 2011; Kruschke, 2010).

Let the posterior distribution for θ, given the data be denoted as *f*(θ|data) acquired from the prior density *f*(θ) and the likelihood *L*(θ) as follows:

$$f(\theta|data)\alpha\, f(data|\theta)f(\theta). \tag{2.33}$$

which is the same as:

Posterior distribution $\propto$ *likelihood* × *prior distribution,*

and even further as follows:

Posterior distribution = Constant × *Likelihood* × *prior distribution.*

After $f(\theta|data)$, is calculated point estimators for θ can be calculated such as the mean, median or mode of $f(\theta|data)$ 95% credible intervals for components of θ are intervals that contain 95% of the posterior density of those components (Sharifi-Malvajerdi *et al.*, 2019).

2.4.2.3 Likelihood function

In the binomial case of small areas within which events are observed, let the number of cases be denoted as *y* and follows an independent binomial distribution, conditional on the probability that an individual is a case, defined as:

$$P(y|\theta) = \text{Binomial}(n, \theta). \tag{2.34}$$

The likelihood is then given by

$$L(y|\theta, n) = \prod_{i=1}^{m} \binom{n}{y} \theta^{y}(1 - \theta)^{(n-y)}. \tag{2.35}$$

Assuming that a logistic link is appropriate for the probability and that a random effect at the individual level is to be included, v_i. Hence

$$\theta = \frac{\exp(\alpha_0 + v_i)}{1 + \exp(\alpha_0 + v_i)}, \tag{2.36}$$

would represent a basic model with intercept to capture the overall rate and prior distribution for the intercept and the random effect could be assumed to be $\alpha_0 \sim N(0, \tau_{\alpha_0})$ and $v_i \sim N(0, \tau_v)$. The hyperprior distribution for the variance parameters could be gamma, inverse gamma, or uniform (Lawson, 2018).

2.4.2.4 Prior distribution

Entirely parameters in Bayesian models are stochastic and are apportioned proper probability distribution based on a priori knowledge about the parameters (van de Schoot *et al.*, 2021). Parameters of prior distributions are a kind of hyperparameter (Banner, Irvine and Rodhouse, 2020). In a beta distribution (prior distribution), p is a parameter of the underlying Bernoulli distribution and a and β are hyperparameter.

We shall consider a Beta prior for θ, that is $\theta \sim Beta\ (\alpha, \beta)$, so that the probability density function of (θ) becomes

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{(\alpha-1)}\ (1-\theta)^{(\beta-1)}. \tag{2.37}$$

2.4.2.5 Posterior distribution

The posterior distribution is derived by multiplying the prior distribution overall parameters by the likelihood function:

$$(\theta|y) \sim \text{Beta}\ (y + \alpha, n - y + \beta). \tag{2.38}$$

For Bayesian analysis, the Gibbs sampler is used to approximate the properties of the marginal posterior distributions for each parameter (Huang and Wand, 2013). Most of posterior

distributions do not have a closed form and must be simulated using Markov chain Monte Carlo (MCMC) approaches like the Metropolis–Hastings (MH) or Gibbs method or both. The outputs are summarised using posterior mean and median, credible intervals (CI) and highest posterior density intervals (van de Schoot *et al.,* 2021).

2.4.2.6 Markov chain Monte Carlo (MCMC)

MCMC approaches are a category of algorithms for sampling from a probability distribution established on constructing a Markov chain that has the specified distribution as its stationary distribution. The state of the chain after a variety of steps is then used as a sample of the preferred distribution. The significance of the sample improves as a function of the amount of steps. The Metropolis algorithm is an MCMC process for attaining a sequence of random walks that uses an acceptance or rejection rule to converge to the target distribution.

The simplest MCMC algorithm is the Gibbs sampler. The Gibbs sampling approach constructs a Markov chain where the probability of the next sample is calculated as the conditional probability given the prior sample. MCMC algorithms are sensitive to their starting point, and often require burn-in phase to draw samples of the target distribution of interest. Inferences are drawn only when the samples obtained are stationary and the chains converged are used to summarize the posterior distribution and compute quantiles. To assess convergence, multiple simulations with starting points spread throughout the parameter space are run, and the distributions from each simulation to the mixed results are compared (Salinelli and Tomarelli, 2014).

2.4.2.7 Model comparison in Bayesian computation

There are various model comparison approaches that have been utilised to select the best-fitting model. The Akaike information criterion (AIC), deviance information criterion (DIC), Bayesian information criterion (BIC) and Watannabe–Akaike information criterion (WAIC) (Aswi *et al*., 2020; Gelman *et al.*, 2013).

2.4.2.8 Bayesian hierarchical spatial model

With correlated data, Bayesian networks employ random effects to model the correlation within sample elements without increasing the Type I error (Kruschke and Vanpaemel, 2015). This approach is used in modelling the spatial context to capture the spatial heterogeneity and autocorrelation in clustered data. The multilevel structure is accurate in accounting for the spatial heterogeneity of hierarchical data and the conditional autoregression model stipulate the spatial autocorrelation at the different levels (Ding *et al.*, 2017; Gelman *et al.*, 2003).

CAR modelling specifications date at least to Besag (1974). Spatial data are directly or indirectly referenced to a locality on the space. The CAR model could be a continuous Markov random field with a conditional probability density function classification and intended to model spatial phenomena that are highly related to a particular local context (Besag, 1974, Cressie, 1993). The CAR models allow for borrowing of strength between neighbouring counties determined by boundaries such that neighbouring counties have similar risk whereas distant counties are expected to point outvariation in risk. The thought of spatial autocorrelation in spatial data analysis is that values of variables in nearby locations are more similar or related than those far apart (Lawson and Lee, 2017; Kyung and Ghosh, 2009).

The conditional autoregressive and SAR models are aimed at modelling spatially autocorrelated data centred on neighbourhood relationships. The conditional autoregressive prior (De Oliveira, 2012) introduces the spatial structure in a hierarchical model. Combining spatially structured and unstructured random effects to the predictable logistic regression model account for over-dispersion and residual spatial structure. Bayesian multilevel logistics models the spatial heterogeneity that exists among groups, while the CAR models spatial autocorrelation (Chunfu, 2017).

2.4.3 Competing risks in correlated survival data

Survival approach evaluates data in which the time till the event is of interest. A competing event in survival analysis is defined as an outcome other than the primary outcome of interest that occurs when conducting a study over a follow-up period (Kleinbaum and Klein, 2012). Competing risk analysis was performed to model factors related with hospital LOS for severe malaria patients accounting for competing risk adjusting for health facility. This section discusses the competing risk analysis (Sections 2.4.3.1) and the accompanying regression approaches in Section 2.4.3.2.

2.4.3.1 Competing risk analysis

Competing risk methodology is a distinctive type of survival analysis that appropriately estimates the marginal probability of an event in the existence of competing events (Pintilie, 2007). Competing risks are often encountered during correlated survival data analysis. Austin and Fine (2017) found that more than three-quarters of high-impact journal articles reviewed were possibly subject to competing risks and were not accounted for in their statistical analysis. According to Austin, (2017) conventional survival models do not account for dependence arising

from clustered data. Outcomes can be correlated if the subjects are nested within related levels interfering with the independent observations assumption. When cluster characteristics like health facilities are not measured, they prompt homogeneity within clusters, affecting the outcome or covariates that have not been accounted for at subject level, as for patients or health workers with similar values within the cluster. Examining the within-cluster correlation strength determines how similar within cluster observations are likely to be. A higher within-cluster correlation gives a more evident clustering effect (Galbraith *et al.*, 2010). Ignoring competing risks in statistical analysis leads to a subjective estimate of incidence of the outcome over time. There is a need to evaluate the impact of correlation on data to ascertain a suitable statistical method of analysing correlated data.

In the competing risks context, diverse event types are regarded as jointly exclusive and use of traditional methods suggested by Kaplan and Meier (1958), for estimation of survival probabilities lead to biased estimates of event probabilities when competing risks are treated as censored observations. Kaplan Meier (KM) curves overestimate the incidence of the outcome over time, and Cox models inflate the relative differences between groups, resulting in biased hazard ratios.

The cause-specific hazard rate for the competing risks can be valued from observable data by presenting a Cox-type regression model to assess the impact of covariates on the cause-specific hazard rates. When cause-specific hazards are modelled, the estimated probability of an event of interest up to a given in time, represented by the cumulative incidence function in the competing risk setting, depends on the cause-specific hazard rates for all possible types of events (Prentice *et al*., 1978).

In 1985, Larson and Dinse offered a piecewise exponential model to assess the conditional event time distributions and later Gray (1988) introduced the subdistribution hazard as an adjusted risk set to keep individuals that failed from a competing event in the risk set for future time points. The method was advanced to a regression model by Fine and Gray (1999), to measure the effect of covariates on the subdistribution hazard. Klein and Andersen (2005) adjusted this method by using the cumulative incidence function (CIF) as measure of interest. The GEE is used to approximate the effect of covariates on the CIF and to provide robust standard errors resulting to accurate statistical estimates (Liang and Zeger, 1986; Nicolaie, Houwelingen and Putter, 2010).

2.4.3.2 Regression approaches for the competing risk-setting

There are two statistical methodologies for survival analysis in the existence of competing risks: the cause-specific hazard (CSH) model and the sub-distribution hazard (SDH) model. The CSH rate is the instantaneous rate of occurrence of the k^{th} event in subjects who are currently event free. The SDH rate is the instantaneous rate of occurrence of the given type of event in subjects who have not yet experienced an event of that type. The CSH model is suitable for examining the aetiology of a disease, including treatment effects, while the SDH model is suitable for examining a prognosis or predicting an individual's risk. The CIF allows for estimation of the incidence of the occurrence of an event while accounting for competing risk (Austin, Lee and Fine, 2016; Lau, Cole and Gange, 2009; Austin and Fine, 2017).

2.4.3.2.1 Cause-specific hazard regression

CSH regression modelled with Cox regression treat failures from the cause of interest as events and failure from other causes as censored observation. The exponentiated regression coefficient

from a CSH model indicates the extent of the relative change in the CSH function in relation to one-unit change in the covariate (Austin and Fine, 2017).

2.4.3.2.2 Subdistribution hazards regression

The SDH model allows one to approximate the effect of covariates on the CIF for the event of interest (Austin, Lee and Fine, 2016). The exponentiated regression coefficient from a Fine Gray SDH model indicates the extent of the relative change in the SDH function associated with one-unit change in the particular covariate (Austin and Fine, 2017).

In medical settings, the two hazard-based regression approaches, are used to analyze competing risk data. The CSH regressions focus on the immediate risk, whereas the SDH regressions have direct linkage to the CIF (Austin, Lee and Fine, 2016). Competing risk analysis includes use of non-parametric methods, such as cumulative incidence function (CIF) curves between groups, for assessment (Cox, 2018).

Applications that involve competing risks may have clustered correlations in individuals. When working with clustered data, Zhou *et al.* (2012) suggest using the Fine Gray model to estimate the CIF using regression model of SDH with multivariate frailty based on hierarchical likelihood estimation method developed to fit the models and draw inferences (Ha *et al*., 2016).

2.5 Gaps identified in the literature

There is scant evidence on the interplay between the health workers' knowledge and practice. Monitoring malaria quality of care studies have reported key knowledge paucities around artesunate-based treatment policies (Zurovac *et al.,* 2018), but have not explored the health worker or health facility predictors influencing the malaria case management practices for severe malaria. Health system challenges have not been tackled, as they are predictors of whether the

required commodities and services to manage severe malaria can be provided. This being a common phenomenon in the relevant studies reviewed based on descriptive analyses done, despite the studies being multilevel (Ojo *et al.*, 2020; Mikomangwa *et al.*, 2019). Studies (Zurovac *et al*., 2018; Zurovac *et al*., 2014; Amboko *et al*., 2020; Moen *et al*., 2016) considered clustering by implementing cluster adjustments and correlation matrices based on hypothetical expectations in the analyses, without considering spatial correlations between clusters (Corani *et al*., 2017; Berger, De Oliveira and Sanso, 2001; Shor *et al*., 2007).

Correlation of data overtime or space was not considered in the various levels in some of the studies (Plucinski *et al*., 2017; Shah *et al*., 2016; Amboko *et al*., 2016; Gute *et al*., 2015; Kazembe *et al*., 2008; Cairns *et al*., 2015). Most previous studies reviewed had some form of clustering, and the methodologies employed during analysis included basic descriptive analysis, logistic regression, log-binomial modelling, and Bayesian and survival approaches (Kiguba, Karamagi, and Bird, 2021; Ojo *et al*., 2020; Elnour *et al*., 2019; Ampadu, Asante and Bosomprah, 2019; Zurovac *et al*., 2018; Mikomangwa *et al*., 2019). However, unfamiliarity with statistical methods for correlated data among researchers often creates challenges in implementation and interpretation. Consequently, the correlations tend to be removed from the datasets or simply ignored. Ignoring correlations will lead to either overestimation or underestimation of the variability resulting to invalid inferences. There is need to assess the impact of correlation on data, to determine the most suitable statistical method of analysing correlated data.

A few studies have been done on LOS in malaria patients but have not accounted for competing events in correlated data; often, the analysis censored death outcomes instead of having them as a competing risk (Austin and Fine, 2017). Overlooking competing risks in

analysis result to subjective estimates of incidence of the outcome over time. Identifying the factors predicting LOS is fundamental in quality of care analyses. However, previous studies have dwelt on assessing readiness to implement the treatment policy following a large clinical trial (Al Farsi *et al*., 2019; Phillips *et al*., 2009; Dondorp *et al*., 2010). A similar study done prior to policy implementation had a defined primary outcome with major implications (Keene *et al*., 2018); however, there are few post-policy follow-up studies on malaria case management based on usual clinical settings within the documented literature. This study examined the impact of correlation and developed statistical methods for analyzing severe malaria case management in Kenya.

CHAPTER THREE

STUDY METHODOLOGY

The study utilized secondary health facility survey data collected to monitor the quality of care for the inpatients in Kenya from 2016 to 2019. The study methodology so described relate to how the data was generated. Therefore, this chapter presents detailed information about study methods, data management process for secondary analysis and the statistical analysis by study objective. The findings of this study have since been published as presented in Chapters four, five, and six.

3.1 Methods for primary study

3.1.1 Study area

The primary study was conducted at the government (GOK) County Referral and major Faith Based Organization (FBO) health facilities in Kenya.

3.1.2 Study design

This was a repeated cross sectional study design to monitor the inpatient malaria quality of care surveys from 2016 to 2019 in Kenya.

3.1.3 Study populations

The inpatient health workers on duty during survey days. Inclusion criteria for health workers' interviews are:

- Health workers on day shift duty during survey days
- Paediatric and medical ward clinicians
- Peadiatric and medical ward nurses

- Health workers providing informed consent

Exclusion criteria for health workers are:

- Health workers on night shift or off duty during survey days
- Student nurses

Inclusion criteria for hospitals are:

- Major County Faith Based Organizations
- Government County referral hospitals

Exclusion criteria for hospitals are:

- Hospitals with admitted patients under antimalarial drug trials

Inclusion criteria for retrospective review of patient files are:

- Patient files meeting study definition of suspected malaria
- Patients discharged from paediatric ward
- Patients discharged from medical ward

Exclusion criteria for retrospective review of patient files:

- Patient files not meeting study definition of suspected malaria
- Patients discharged from other than paediatric and medical wards

3.1.4 Sample size determination and sampling procedure

3.1.4.1 Sample size determination

The formula used for the sample size calculation is as follows:

$$n=\frac{\text{deff x } [Z_{1-\alpha}\sqrt{2P\ (1-P)} + Z_{1-\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}]^2}{(P_2-P_1)^2}. \quad (3.1)$$

where

$Z_{1-}\alpha$ = 1.96 (5% significance) is standard value for type I error

$Z_{1-}\beta$ = 0.84 (80% power) is standard value for type II error

P_1 = the value of key outcome at time 1(45%)

P_2 = the value of key outcome at time 2 (50%)

$P = (P_1 + P_2)/2$

deff = design effect (1.8)

The required sample was 2814 suspected malaria admissions. With an assumption of recruiting an average of 4 health workers per health facility, a minimum of 360 health workers was required per survey from a total of 90 health facilities.

3.1.4.2 Sampling Procedure

The GOK county referral hospitals and the FBOs equivalent to referral hospitals health facilities were selected purposively. As sampling for proportionality is not the main concern in health facility selection, purposive sampling allowed for quick access to the targeted population. 90 hospitals (47 GOK and 43 FBO) participated in the study. There were 26 high-risk regions near Lake Victoria and along the Indian Ocean coast among the 90 surveyed hospitals, while the remaining 64 were form low-risk areas (Macharia *et al*., 2018, KMIS, 2015). A random sample of clinicians and nurses working in various paediatric and medical wards were surveyed in each hospital as part of the study and interviews were conducted using the fishbowl sampling technique. The medical records of patients admitted to each of the examined institutions were retrieved and used to compile data retrospectively. At GOK and FBO hospitals, 30 consecutive

patients (15 from paediatric and 15 from medical wards) were selected for data extraction prior to screening inpatient and laboratory registries.

3.1.5 Data collection tools, training of personnel

Data was collected using standardized questionnaires; 1) health worker interview form, 2) hospital assessment form, and 3) patient level data extraction form. The research assistants were trained a week preceding data collection (Zurovac *et al*., 2018).

3.1.6 Data collection procedures

Three methods of data collection were applied. The research assisstants reviewed the patients file and extracted data retrospective entering the data into the exit forms. The health workers from the paediatric and medical ward (clinicians and nurses) were randomly selected and interviewed after obtaining the informed consent based the inclusion criteria. The interviews were used to gather information on the demographics of health workers, their exposure to in-service training, guidelines, and supportive supervision, as well as information on how they handle severe malaria patients. Multiple-choice questions were used to test the knowledge component. Finally, the hospital was assessed for readiness (Zurovac *et al*., 2018). Medicines and job aids availability were examined in the pharmacy and admission wards. After the interviews and knowledge assessments were completed, all participating health workers were notified of the right responses, provided national malaria case management guidelines, and, if missing, distributed artesunate administration posters to be exposed on the wall.

3.1.7 Quality assurance and control procedures

Quality assurance was applied during and after the study period. Initially the data collection forms were piloted, pretested and refined. During training, the research assistants went through

concordance testing up to more than 90%. During the actual fieldwork, the research assistants reviewed their daily data collection forms together with their team supervisors and during data entry, the data collection forms were double entered and checked for consintency. Lastly, the forms were kept securely after fieldwork.

3.2 Data management for secondary analysis

This section describes data for secondary analysis, the analytical approaches, the study variables and the initial exploratory data analyses performed. It further describes the statistical analysis per study objective.

3.2.1 Description of data for secondary analysis

The secondary analysis utilized health facility survey data collected to monitor the quality of care for the inpatients in Kenya from 2016 to 2019 (Table 3.1). The data sets were merged based on the study objective requirements. To investigate the impact of correlated data while adjusting for health facility and county structures, the health facility and health worker datasets for 2016 and 2017 were merged according to health facility ownership (FBO and GOK) and independent analysis was performed. A total of 94 and 86 health facilities from GOK and FBO respectively, 367 and 330 health workers from GOK and FBO respectively were included in the study. To evaluate time to discharge for patients suspected with malaria in the presence of a competing event, admissions for 2018 from the hospitals were analysed. A total of 2396 suspected severe malaria patients' admissions from 90 health facilities were included in the study. Finally, prediction of county level estimates on health workers' knowledge levels about artesunate using Bayesian approach, health facility and health worker datasets for 2019 was merged and analysis performed. A total of 349 health workers were included in the study but four had missing

information, hence 345 health workers were included in the final dataset for analysis and 89 health facilities.

Table 3.1 Study population

Year	**2016**	**2017**	**2018**	**2019**
	N	**N**	**N**	**N**
Health facilities	94	86	90	89
Health workers	367	330	336*	349
Suspected severe malaria patients' admissions	2386*	2243*	2396	2485*

*not applicable in this study

3.2.2 Analytical approaches

Initially, to investigate the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI) and prediction multilevel modelling approach that adjusted for the health facilities and county structures were implemented. Then, Bayesian hierarchical spatial modelling was fitted to analyse the polytomous data adjusting for county structures. Lastly, the factors associated with LOS among severe malaria patients was implemented using competing risk approach adjusting for health facility structures based on usual clinical setting in Kenya.

In this context, substantive applications of frequentist and Bayesian approaches to multilevel and competing risk analysis were explored while assessing the impact of correlation adjusting for health facility or county structures.

3.2.3 Study variables

This section highlights the study variables that were categorized as health worker outcomes, health facility and health worker level factors, length of stay outcome variables and patient level factors examined.

3.2.3.1 Preliminary list of health worker outcomes

a. Correct knowledge of recommended severe malaria treatment

    a1) For paediatrics and non-pregnant population

    a2) For expectant mothers in their first trimester

    a3) For expectant mothers in their second and third trimester

    a4) Composite treatment policy knowledge for all categories of patients

b. Correct knowledge of recommended artesunate dose

    b1) For paeditrics weighing less than 20kg

    b2) For patients weighing 20kg and more

    b3) Composite dosing knowledge for both weight categories of patients

c. Correct knowledge of artesunate dosing interval in hours

    c1) After the first dose

    c2) After the second dose

    c3) After the third dose

    c3) Composite knowledge of artesunate dosing interval

d. Correct knowledge of artesunate reconstitution and dilution solutions

d1) Knowledge of bicarbonate for reconstitution

d2) Knowledge of normal saline or 5% dextrose for dilution of reconstituted artesunate

d3) Composite knowledge of artesunate preparation

e. Correct knowledge of recommended route of artesunate administration

3.2.3.2 Preliminary list of the health worker and health facility factors examined

A) Hospital level

- Hospital ownership (government vs faith based)
- Ward allocation (paediatric vs medical)
- Malaria endemicity (high vs low)
- Exposure to artesunate poster (yes vs no)
- Access to artesunate dosing wheel (accessible vs not accessible)
- Availability of artesunate (available vs not available)

B) Health worker level

- Gender (male vs female)
- Health workers' age (grouped)
- Health worker cadre was grouped as nurse vs clinician. The clinicians included medical officer, clinical officer, medical officer intern, clinical officer intern and consultants.
- Years of inpatient experience (grouped)
- Case management training (trained vs not trained)

- Malaria treatment guidelines (accessible vs not accessible)
- Paediatric protocol (accessible vs not accessible)
- Supportive supervision (yes vs no)

3.2.3.3 Preliminary list of length of stay outcome for severe malaria patients

a) Length of stay defined as time to discharge from hospital (days) was the event of interest

b) Time to death was considered a competing event

3.2.3.4 Preliminary list of the patient factors to be examined

a) General information (age, sex, weight and ward).

b) Patient factors assessed on admission (pulse, respiratory rate, temperature, fever complaint and blood pressure).

c) Documented clinical features of severe malaria features (altered consciousness, convulsions, prostration, severe anaemia, respiratory distress, jaundice, shock, abnormal bleeding, renal failure, haemoglobinuria, hypoglycaemia and pulmonary oedema).

d) Monitoring of inpatients during hospitalization (respiratory rate, temperature, blood pressure, oxygen saturation and pulse rate).

e) Laboratory investigation done (malaria test on admission, malaria test post-admission, malaria test result on admission, Hb/HCT done and Glucose/RBS test done).

f) Diagnosis (Health workers' malaria diagnosis on admission and confirmed severe malaria).

g) Treatment during the hospitalization (artesunate).

3.2.4 Exploratory data analyses

Exploratory data analyses were done to detect mistakes in the data, check for assumptions, and determine relationships between explanatory and outcome variables and to help select appropriate models. According to the assessment, the missing information from both variables was scanty and classified as missing completely at random (MCAR). Hence, complete case analysis approach restricting analysis to subjects with complete data for all variables was considered during analysis. Descriptive analysis was conducted to categorize the correct knowledge on management of severe malaria. Composite indicators on outcome variables were created using Multiple Correspondence Analysis (MCA) established from the health worker outcome variables (Section 3.2.3.1). The reliability index was measured using Cronbach's alpha (Ayele and Mwambi, 2014).

3.3 Statistical analysis by study objective

3.3.1 To investigate the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI), and prediction, adjusting for health facility and county structures

In order to assess this objective, the predictors of the inpatient health workers' knowledge about artesunate-based severe malaria treatment recommendations in hospitals were examined. The composite indicator from the MCA for the health workers' knowledge on severe malaria treatment policy, artesunate dosing and interval, preparation and preferred route of administration were the dependent variables. Independent variables were classified as health worker level (individual) or health facility level (contextual) factors influencing the knowledge of artesunate treatment policy as described in Section 3.2.3.2. The factors examined included

those believed to be related to the knowledge outcomes since reflecting programmatic interventions (training, supervision, guidelines, commodity availability, and job-aids), as well as those likely to influence or confound the association between interventional factors and outcomes (malaria risk, hospital ownership, and demographics).

To determine predictors of health workers' knowledge, hospital and health worker level were examined for each of the composite indicator of knowledge outcomes applying multilevel mixed effects ordinal logistic regression modelling to address the correlated nature of data (Figure 3.1). The proportion corresponding to 95% confidence interval for each of the composite indicator around the knowledge outcomes (Section 3.2.3.1) were estimated at the health worker level after adjusting for the clustering at the hospital and county level.

Univariable analyses were performed to identify the predictor variables to include in the multivariable analysis (Table 4.4-4.8). The estimates from univariable analyses generated unadjusted odds ratio, *P*-values, and 95% CI. Brant test was used to test for assumption of proportionality for ordinal logistic regression models. Testing of hypothesis and approximation of confidence interval was done at 0.05 alpha levels. Factors with a *P*-value <0.15 from the univariate analysis and interaction terms with *P*-value <0.05 were entered into multivariate mixed effects logistic regression models to adjust for confounding. Factors not meeting the entrance criteria for the multivariate analysis were added to the models, one factor at a time to gauge any change of the odds ratio (OR) and incase of any significant change, the variable would be retained in the multivariate model. All predictors' analyses estimating 95% confidence intervals and *P*-values were adjusted for clustering at county and hospital level. During analysis some of the predictor variables; the years of experience, artesunate dosing wheel, a one-day

orientation training on artesunate and on-job artesunate training were dropped due to multicollinearity.

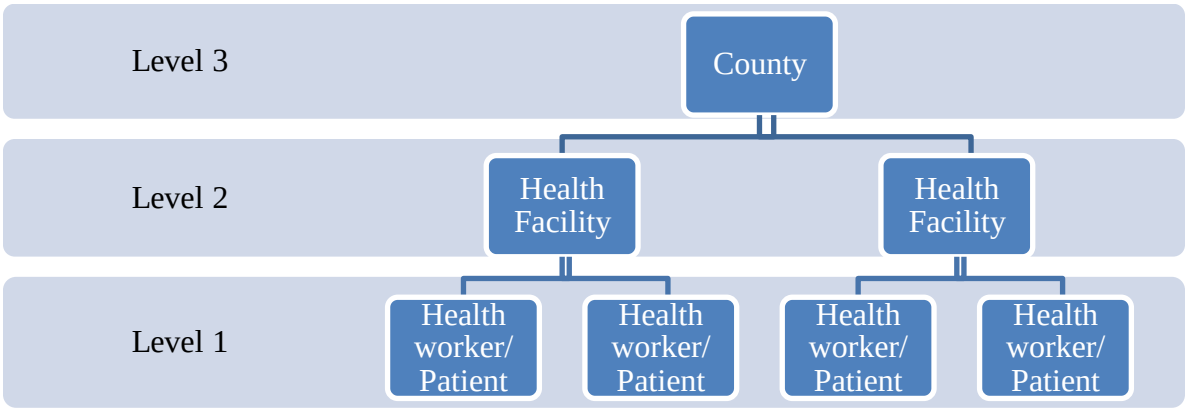


Figure 3. 1 Illustration of multilevel modeling
Source: Author

The impact of correlation on the dichotomous and polytomous response was performed during the predictor analysis. Initially, the data were analysed conventionally while ignoring correlation in multivariate logistic regression; subsequently, the data were analysed adjusting for clustering at county and health facility levels in both FBO and GOK sector. Then, the variance within the statistical estimates; the standard errors and the confidence intervals were examined. The standard errors of the coefficient assessed the variability between the estimates while the CI assessed the practical significance of study results and provided the expected range for the true odds ratio for the population to fall within. During analyses, the random effects and intracluster correlation were examined conditional on the fixed-effects covariates.

3.3.2. To model factors related with hospital length of stay for severe malaria patients accounting for competing risk adjusting for health facility and county structures

Competing risk approach was implemented. The primary outcome variable was time to discharge and death was considered as a competing risk, while, the other outcomes were censored (Section 3.2.3.3). The time to event was calculated in days beginning the time the patient with suspected malaria was admitted till the time the patient was discharged or died. Therefore, secondary analysis was implemented at the patient level. The patient and health facility level data were linked to the datasets.

CIF, SDH and CSH models were implemented to assess the impact of covariates on the cumulative probability of discharged in the presence of a competing event during hospitalization. The factors related with LOS for the inpatients were assessed using the CSHR while, SDHR assessed the relationship with cumulative incidence in the presence of competing risk in correlated data. Factors that were significant (P-value<0.05) during univariable analyses were included into a multivariate model. Schoenfeld residual test was used to test for the proportional hazard's assumption conditions.

Survival curves were estimated and cumulative incidence function curves to compare the risks illustrated by artesunate treatment variable. Initially, the data was analysed using standard methods while ignoring the competing risk and Kaplan-Meier curves were used, subsequently, the data was analysed using modes that account for competing risk. CIF curves were used to compare risk models.

3.3.3 To apply Bayesian hierarchical approach to analyze dichotomous or polytomous and survival data

Health workers' knowledge about severe malaria treatment policy, artesunate dose and preparation were considered in Bayesian modelling. The outcomes were built using MCA approach based on the composite indicator of health workers' knowledge outcomes (Section 3.2.3.1) resulting to three levels (high, medium or low) and the explanatory variables for analysis were categorized as health worker or health facility (section 3.2.3.2). During univariable analysis, estimated odds ratio (OR) and 80% Credible Intervals (CI) identified significant explanatory variables that were incorporated in multilevel analysis. Subsequently, three hierarchical models were fitted using ordinal logistic regression analysis adjusting for clustering at the county level. The models included the ordinal logistic regression with: spatially structured random effects, unstructured spatially random effects, and convolution models. The models were compared by the deviance information criterion (DIC) and robust simulations were done using three chains of Markov Chain Monte Carlo (MCMC) algorithms. The posterior means, odds ratio, quantiles, median, standard deviation and their 95% CI were used to evaluate the significance of the factors. The best fitting model of health workers' knowledge outcome were mapped at subnational levels.

3.4 Ethical considerations

The studies were approved by Kenyatta National Hospital/the University of Nairobi Ethics and Research Committee (KNH/UON ERC), KNH-UON P643/10/2015. The DNMP granted permission to nest the study (Appendix 1). This study was therefore nested within the severe

malaria quality of care for the inpatient survey and was approved by KNH/UON ERC P233/04/2018 (Appendix 2).

Health workers on duty during survey days were the only human subjects involved in the survey. Prior to the interviews, the health workers consented and the interviews conducted in private places. During extraction of data from routine records the information recorded from the patient files was maintained confidentially and anonymous to the greatest extent possible. The data collected were used for the purpose of the study only.

CHAPTER FOUR

OBJECTIVE ONE: ASSESSING THE IMPACT OF CORRELATION ON DICHOTOMOUS AND POLYTOMOUS OUTCOMES IN TERMS OF ESTIMATION OF PARAMETERS, PRECISION MEASURES, AND PREDICTION, ADJUSTING FOR HEALTH FACILITY AND COUNTY STRUCTURES IN KENYA.

In order to study this objective, the predictors of the inpatient health workers' knowledge about artesunate-based severe malaria treatment recommendations in hospitals were examined using multilevel mixed effect logistic regression modelling approach that account for clustering of health workers within the health facilities and counties in Kenya, 2016-2017. This chapter provides background information (Section 4.1), methods (Section 4.2), knowledge outcome and definitions are explained in Section 4.3. The results are discussed in Section 4.4 while, discussion on predictor analysis are presented in Section 4.5. The chapter ends with conclusion in Section 4.6. This scientific work has been published by *Malaria Journal*.

4.1 Background

Despite a falling prevalence of *Plasmodium falciparum* infection (Snow *et al*., 2015; Macharia *et al*., 2018), severe malaria is a common cause of admission in hospitals in Kenya (Aketch *et al*., 2019). Alongside the World Health Organization (WHO) recommendation in 2012 (WHO, 2012), Kenya was among the first African countries to adopt the artesunate treatment policy for severe malaria (MoH, 2012a). The Malaria Programme revised national malaria guidelines to reflect the new policy, procured and facilitated distribution of injectable artesunate to health facilities, developed training curriculums and job aids around new case-management standards, and implemented in-service training programs to facilitate readiness of clinicians and nurses to

deliver new treatment standards (MoH, 2012b; MoH, 2015a). Of estimated 400 public hospitals in the country, about three-quarters are government-owned and the remaining are faith-based organization (FBO) hospitals (Ouma *et al.*, 2015). National artesunate implementation equally targeted government and FBO health workers, but subsidized artesunate was available only to the government hospitals procuring medicines through the Kenya Medical Supply Agency (KEMSA).

Health workers' knowledge about new treatment policy and recommended use of the new medicines is one of the basic pre-requisites determining the readiness of the health system to implement any drug policy (Lomas *et al.*, 1989; Berhe *et al.*, 2018; Yoo *et al.*, 2019). Several studies have suggested major knowledge deficiencies about artesunate-based treatment recommendations (Zurovac *et al.*, 2018; Mikomangwa *et al.*, 2019), but no study has examined predictors of the health workers' knowledge. This study examined the predictors of the inpatient health workers' knowledge about artesunate-based severe malaria treatment recommendations in government and FBO hospitals in Kenya.

4.2 Methods

4.2.1 Data sources

The methods used in all hospital surveys were the same as previously published (Zurovac *et al.*, 2018) and details of data used in this study are provided in Section 3.2.1. Data collection procedures are elaborated in section 3.1.6.

4.2.2 Knowledge outcomes and definitions

Artesunate-based treatment recommendations for severe malaria were the basis for the selection of five knowledge outcomes. The outcomes reflected the correctness of health workers'

knowledge about 1) severe malaria treatment policy, 2) artesunate dose, 3) dosing intervals, 4) artesunate preparation, and 5) preferred route of artesunate administration. The preferred route of administration had two levels of categorization, but all other outcomes were characterized on a three-point scale. Table 4.1 provides the definitions of each outcome's knowledge categories.

Table 4. 1 Categories of the knowledge outcomes and study definitions

Knowledge outcomes	**National recommendations**	**Knowledge categories**	**Category definitions**
Treatment policy for severe malaria	Artesunate for the following 3 severe malaria populations: 1) children & non-pregnant adults; 2) pregnant women in 1st trimester; 3) pregnant women in 2nd & 3rd trimester	High	Artesunate response for all 3 severe malaria populations
		Medium	Artesunate response for 2 severe malaria populations
		Low	Artesunate response for one or none of the populations
Artesunate dose	2 weight categories: 1) 3 mg/kg for child <20kg, 2) 2.4 mg/kg for patient >20kg	High	Correct response for 2 weight categories
		Medium	Correct response for one weight category
		Low	No correct response for any of the weight categories
Artesunate dosing interval	3 dosing intervals: 1) 12 hours between 1st & 2nd dose 2) 12 hours between 2nd & 3rd dose 3) 24 hours between 3rd & 4th dose	High	Correct response for all 3 dosing intervals
		Medium	Correct response for 2 dosing intervals
		Low	Correct response for one or none of the dosing intervals
Artesunate preparation	Solutions for 2 artesunate preparation steps: 1) bicarbonate for reconstitution	High	Correct response for 2 preparation steps
		Medium	Correct response for one

Knowledge outcomes	National recommendations	Knowledge categories	Category definitions
	2) saline or 5% dextrose for dilution		preparation step
		Low	No correct response for any of the preparation steps
Preferred route of artesunate administration	Intravenous slow bolus	High	IV slow bolus response
		Low	Any other response

4.2.3 Statistical analysis

For each of the five outcomes, descriptive analyses were conducted using frequencies and percentages for each of the predictor variables. The knowledge of health workers about the treatment of severe malaria was investigated using multilevel ordinal and binary logistic regression models for four ordinal and one binary outcome (Table 4.1). Gender, age, pre-service training, ward allocation, exposure to malaria guidelines, in-service malaria case management training, and supportive supervision were factors studied at the health worker level. At the hospital level, the factors included availability of artesunate, the presence of artesunate-administration posters, and endemicity. A multilevel modeling strategy was used in order to account for the clustering of health workers within hospitals. Unadjusted regression models for each outcome were used to select candidate predictor variables, and multicollinearity was assessed between those with *P*-values<0.15. Multivariate regression models were used to adjust our estimations for the specified variables. Health facilities and county structures were introduced as random effects in all regression models while the survey round was a fixed effect. The Brant test was used to assess the proportional odds assumption in the final multivariable ordinal regression models (Brant, 1990). As the surveys in government and FBO hospitals were

conducted in different years all analyses were stratified by hospital ownership. *P*-values and 95% confidence intervals (CI) accompany the odds ratios derived from the multivariable models. The 0.05 significance threshold was used for hypothesis testing. Stata 14 was used for all analyses (StataCorp, College Station, TX, USA).

4.3 Results

4.3.1 Characteristics of study health workers

Table 4.2 presents characteristics of 367 and 330 health workers respectively interviewed at the government and FBO hospitals. In both sectors, most health workers were female, younger than 35 years, having less than five years of inpatient experience and working in low malaria risk areas. Nurses and clinicians as well as paediatric and medical ward health workers were similarly represented within and between hospital sectors. Compared to the FBO sector, government health workers were however more commonly female (61.9% *vs.* 51.2%), older than 35 years (37.6% *vs.* 17.3%) and with more than five years of experience (43.9% *vs.* 24.5%). Regarding the exposure to the relevant interventions, over three-quarters of health workers in both sectors worked at hospitals with artesunate in stock, however, only about a third had access to malaria guidelines and less than a quarter received in-service malaria case-management training and supportive supervision in the past three months. While only minor differences were observed between the sectors with respect to the training and supervision, government health workers less commonly had access to malaria guidelines (32.2% *vs.* 39.8%), but more frequently worked in wards with displayed artesunate administration poster (61.0% *vs.* 47.0%).

Table 4. 2 Characteristics of study health workers

Health worker characteristics	Government hospitals N=367		Faith based hospitals N=330	
	n	%	n	%
Gender				
Male	140	38.1	161	48.8
Female	227	61.9	169	51.2
Age[a]				
35-70 years	138	37.6	57	17.4
21-35 years	229	62.4	271	82.6
Pre-service training				
Clinician	175	47.7	156	47.3
Nurse	192	52.3	174	52.7
Inpatient experience[b]				
<5 years	203	55.8	249	75.5
>5 years	161	44.2	81	24.5
Ward allocation				
Paediatric	185	50.4	168	50.9
Medical	182	49.6	162	49.1
Malaria endemicity				
High	102	27.8	88	26.7
Low	265	72.2	242	73.3
Exposure to artesunate interventions				
Case management training	87	23.7	66	20.0
Malaria guideline	118	32.2	131	39.7
Supportive supervision	39	10.6	29	8.8
Artesunate administration poster	224	61.0	157	47.6
Artesunate in stock	276	75.2	257	77.9

[a] Denominator excludes 2 health workers with missing information in faith-based hospitals

[b] Denominator excludes 3 health workers with missing information in government hospitals

4.3.2 Knowledge of artesunate-based severe malaria treatment recommendations

Based on hospital ownership and knowledge categories, Table 4.3 shows how well health workers understand artesunate-based severe malaria treatment recommendation. A third of government and FBO health workers had high knowledge of artesunate treatment policies, a third knew all dosing intervals, and about half knew preparation solutions (49.9% vs. 55.8%). About half to two-thirds of health workers knew the artesunate dose for both weight categories (50.8% vs. 66.7%), and over three-quarters knew the preferred method of administration (78.7% vs 82.4%).

Table 4. 3 Health workers' knowledge about artesunate-based severe malaria treatment recommendations

Health workers' knowledge:	**Government hospitals**		**Faith based hospitals**	
	N=367		**N=330**	
	n	**%**	**n**	**%**
Treatment policy for severe malaria[a]				
High	113	30.8	108	32.9
Medium	131	35.7	101	30.8
Low	123	33.5	119	36.3
Artesunate dose[b]				
High	186	50.8	220	66.7
Medium	100	27.3	67	20.3
Low	80	21.9	43	13.0
Artesunate dosing intervals				
High	123	33.5	110	33.3
Medium	136	37.1	137	41.5

Health workers' knowledge:	Government hospitals		Faith based hospitals	
	N=367		N=330	
	n	%	n	%
Low	108	29.4	83	25.2
Artesunate preparation				
High	183	49.9	184	55.8
Medium	134	36.5	114	34.5
Low	50	13.6	32	9.7
Preferred route of administration				
High	289	78.7	272	82.4
Low	78	21.3	58	17.6

[a] Denominator excludes 2 health workers with missing information in faith-based hospitals
[b] Denominator excludes 1 health worker with missing information in government hospitals

4.3.3 Results of univariate logistic regression analysis

Tables 4.4-4.8 show the results of univariate logistic regression analysis investigating the relationship between 11 factors and five knowledge outcomes on artesunate-based severe malaria treatment recommendation for each of the two hospital ownership sectors.

Table 4. 4 Univariable ordinal logistic regression analysis of predictors of knowledge on severe malaria treatment policy, by hospital ownership

	Government hospitals						FBO hospitals					
	N=367	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	*P*-value	N=328	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	*P*-value
Age[a]												
35-70 years	138	50(36.2)	51(37.0)	37(26.8)	1.0(ref)		57	23(40.4)	15(26.3)	19(33.3)	1.0(ref)	
21-35 years	229	73(31.9)	80(34.9)	76(33.2)	1.29(0.87-1.92)	0.205	269	96(35.7)	85(31.6)	88(32.7)	1.10(0.62-1.93)	0.750
Gender												
Female	227	82(36.1)	82(36.1)	63(27.8)	1.0(ref)		168	69(41.1)	50(29.8)	49(29.2)	1.0(ref)	
Male	140	41(29.3)	49(35.0)	50(35.7)	1.45(0.97-2.16)	0.072	160	50(31.3)	51(31.9)	59(36.9)	1.58(1.03-2.43)	0.034
Cadre												
Nurse	192	78(40.6)	65(33.9)	49(25.5)	1.0(ref)		174	80(46.0)	51(29.3)	43(24.7)	1.0(ref)	
Clinician	175	45(25.7)	66(37.7)	64(36.6)	1.91(1.29-2.82)	0.001	154	39(25.3)	50(32.5)	65(42.2)	2.51(1.64-3.85)	<0.001
Ward												
Medical	182	58(31.9)	72(39.6)	52(28.6)	1.0(ref)		162	60(37.0)	55(34.0)	47(29.0)	1.0(ref)	
Paediatric	185	65(35.1)	59(31.9)	61(33.0)	1.02(0.70-1.49)	0.932	166	59(35.5)	46(27.7)	61(36.7)	1.24(0.82-1.87)	0.308
Endemicity												
Low	265	90(34.0)	91(34.3)	84(31.7)	1.0(ref)		240	83(34.6)	73(30.4)	84(35.0)	1.0(ref)	
High	102	33(32.4)	40(39.2)	29(28.4)	0.95(0.59-1.55)	0.848	88	36(40.9)	28(31.8)	24(27.3)	0.70(0.37-1.33)	0.274
CM Guidelines												
No	249	85(34.1)	92(36.9)	72(28.9)	1.0(ref)		198	92(46.5)	54(27.3)	52(26.3)	1.0(ref)	
Yes	118	38(32.2)	39(33.1)	41(34.7)	1.17(0.78-1.78)	0.447	129	27(20.9)	46(35.7)	56(43.4	2.89(1.81-4.61)	0.000
CM training												
No	280	108(38.6)	93(33.2)	79(28.2)	1.0(ref)		263	105(39.9)	77(29.3)	81(30.8)	1.0(ref)	
Yes	87	15(17.2)	38(43.7)	34(39.1)	2.09(1.33-3.30)	0.002	65	14(21.5)	24(36.9)	27(41.5)5)	1.83(1.07-3.14)	0.027
Supervision												
No	328	117(35.7)	115(35.1)	96(29.3)	1.0(ref)		299	109(36.5)	92(30.8)	98(32.8)	1.0(ref)	
Yes	39	6(15.4)	16(41.0)	17(43.6)	2.21(1.18-4.16)	0.014	29	10(34.5)	9(31.0)	10(34.5)	1.46(0.65-3.26)	0.359
AS poster												
No	143	55(38.5)	51(35.7)	37(25.9)	1.0(ref)		171	65(38.0)	60(35.1)	46(26.9)	1.0(ref)	
Yes	224	68(30.4)	80(35.7)	76(33.9)	1.44(0.95-2.18)	0.084	157	54(34.4)	41(26.1)	62(39.5)	1.44(0.91-2.28)	0.122
AS in stock												
No	91	30(33.0)	39(42.9)	22(24,2)	1.0(ref)		73	37(50.7)	18(24.7)	18(24.7)	1.0(ref)	
Yes	276	93(33.7)	92(33.3)	91(33.0)	1.27(0.78-2.07)	0.337	255	82(32.2)	83(32.5)	90(35.3)	2.01(1.08-3.73)	0.028
Survey												
Baseline	185	72(38.9)	70(37.8)	43(23.2)	1.0(ref)		163	68(41.7)	51(31.3)	44(27)	1.0(ref)	
Follow up	182	51(28)	61(33.5)	70(38.5)	1.89(1.28-2.78)	0.001	165	51(30.9)	50(30.3)	64(38.8)	1.74(1.15-2.63)	0.009

[a] Denominator excludes 2 health workers with missing information in faith-based hospitals

Table 4. 5 Univariable ordinal logistic regression analysis of predictors of artesunate dose knowledge, by hospital ownership

	Government hospitals						FBO hospitals					
	N=366	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	p-value	N=330	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	p-value
Age[a]												
35-70 years	137	38(27.7)	38(27.7)	61(44.5)	1.0(ref)		57	9(15.8)	12(21.1)	36(63.2)	1.0(ref)	
21-35 years	229	42(18.3)	62(27.1)	125(54.6)	1.77(1.14-2.73)	0.010	271	34(12.5)	55(20.3)	18267.2)	1.41(0.73-2.72)	0.309
Gender												
Female	226	58(25.7)	54(23.9)	114(50.4)	1.0(ref)		169	24(14.2)	38(22.5)	107(63.3)	1.0(ref)	
Male	140	22(15.7)	46(32.9)	72(51.4)	1.37(0.88-2.11)	0.162	161	19(11.8)	29(18.0)	113(70.2)	1.23(0.74-2.05)	0.421
Cadre												
Nurse	191	48(25.1)	53(27.7)	90(47.1)	1.0(ref)		174	33(19.0)	36(20.7)	105(60.3)	1.0(ref)	
Clinician	175	32(18.3)	47(26.9)	96(54.9)	1.45(0.96-2.18)	0.078	156	10(6.4)	31(19.9)	115(73.7)	2.21(1.33-3.67)	0.002
Ward												
Medical	181	43(23.8)	63(34.8)	75(41.4)	1.0(ref)		162	21(13.0)	33(20.4)	108(66.7)	1.0(ref)	
Paediatric	185	37(20.0)	37(20.0)	111(60.0)	1.94(1.29-2.94)	0.002	168	22(13.1)	34(20.2)	112(66.7)	0.94(0.58-1.53)	0.815
Endemicity												
Low	264	61(23.1)	72(27.3)	131(49.6)	1.0(ref)		242	36(14.9)	44(18.2)	162(66.9)	1.0(ref)	
High	102	19(18.6)	28(27.5)	55(53.9)	1.25(0.63-2.48)	0.521	88	7(8.0)	23(26.1)	58(65.9)	1.08(0.44-2.68)	0.868
CM Guidelines[b]												
No	249	63(25.3)	66(26.5)	120(48.2)	1.0(ref)		198	32(16.2)	43(21.7)	123(62.1)	1.0(ref)	
Yes	117	17(14.5)	34(29.1)	66(56.4)	1.76(1.11-2.80)	0.016	131	11(8.4)	24(18.3)	96(73.3)	1.50(0.86-2.61)	0.151
CM training												
No	279	61(21.9)	84(30.1)	134(48.0)	1.0(ref)		264	38(14.4)	53(20.1)	173(65.5)	1.0(ref)	
Yes	87	19(21.8)	16(18.4)	52(59.8)	1.41(0.83-2.39)	0.198	66	5(7.6)	14(21.2)	47(71.2)	1.17(0.61-2.26)	0.640
Supervision												
No	327	76(23.2)	86(26.3)	165(50.5)	1.0(ref)		301	43(14.3)	63(20.9)	195(64.8)	1.0(ref)	
Yes	39	4(10.3)	14(35.9)	21(53.8)	1.13(0.56-2.26)	0.739	29	0(0.0)	4(13.8)	25(86.2)	4.15(1.24-13-92)	0.021
AS poster												
No	143	43(30.1)	48(33.6)	52(36.4)	1.0(ref)		173	27(15.6)	45(26.0)	101(58.4)	1.0(ref)	
Yes	223	37(16.6)	52(23.3)	134(60.1)	2.92(1.75-4.86)	<0.001	157	16(10.2)	22(14.0)	119(75.8)	2.57(1.40-4.73)	0.002
AS in stock												
No	91	30(33.0)	20(22.0)	41(45.1)	1.0(ref)		73	15(20.5)	16(21.9)	42(57.5)	1.0(ref)	
Yes	275	50(18.2)	80(29.1)	145(52.7)	2.13(1.15-3.92)	0.015	257	28(10.9)	51(19.8)	178(69.3)	1.71(0.80-3.68)	0.167
Survey												
Baseline	185	50(27.0)	61(33.0)	74(40.0)	1.0(ref)		164	24(14.6)	42(25.6)	98(59.8)	1.0(ref)	
Follow up	181	30(16.6)	39(21.5)	112(61.9)	2.41(1.59-3.67)	0.000	166	19(11.4)	25(15.1)	122(73.5)	1.91(1.17-3.12)	0.010

Table 4. 6 Univariable ordinal logistic regression analysis of predictors of artesunate dosing interval knowledge, by hospital ownership

	GOK hospitals						FBO hospitals					
	N	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	p-value	N	Low n (%)	Medium n (%)	High n (%)	OR (95% CI)	p-value
Age												
35-70 years	138	42(30.4)	58(42.0)	38(27.5)	1.0(ref)		57	14(24.6)	26(45.6)	17(29.8)	1.0(ref)	
21-35 years	229	66(28.8)	78(34.1)	85(37.1)	1.48(0.97-2.25)	0.067	271	69(25.5)	109(40.2)	93(34.3)	1.05(0.61-1.83)	0.854
Sex												
Female	227	70(30.8)	83(36.6)	74(32.6)	1.0(ref)		169	44(26.0)	76(45.0)	49(29.0)	1.0(ref)	
Male	140	38(27.1)	53(37.9)	49(35.0)	1.31(0.86-1.99)	0.211	161	39(24.2)	61(37.9)	61(37.9)	1.35(0.88-2.06)	0.172
Cadre												
Nurse	192	59(30.7)	81(42.2)	52(27.1)	1.0(ref)		174	53(30.5)	73(42.0)	48(27.6)	1.0(ref)	
Clinician	175	49(28.0)	55(31.4)	71(40.6)	1.50(1.01-2.23)	0.046	156	30(19.2)	64(41.0)	62(39.7)	1.90(1.25-2.90)	0.003
Ward												
Medical	182	58(31.9)	70(38.5)	54(29.7)	1.0(ref)		162	41(25.3)	66(40.7)	55(34.0)	1.0(ref)	
Paediatric	185	50(27.0)	66(35.7)	69(37.3)	1.44(0.97-2.13)	0.074	168	42(25.0)	71(42.3)	55(32.7)	0.96(0.64-1.45)	0.863
Endemicity												
Low	265	87(32.8)	91(34.3)	87(32.8)	1.0(ref)		242	65(26.9)	97(40.1)	80(33.1)	1.0(ref)	
High	102	21(20.6)	45(44.1)	36(35.3)	1.49(0.76-2.94)	0.250	88	18(20.5)	40(45.5)	30(34.1)	1.20(0.65-2.21)	0.566
CM Guidelines												
No	249	75(30.1)	101(40.6)	73(29.3)	1.0(ref)		198	60(30.3)	80(40.4)	58(29.3)	1.0(ref)	
Yes	118	33(28.0)	35(29.7)	50(42.4)	1.79(1.14-2.82)	0.012	131	23(17.6)	56(42.7)	52(39.7)	1.85(1.17-2.91)	0.008
CM training												
No	280	82(29.3)	102(36.4)	96(34.3)	1.0(ref)		264	73(27.7)	106(40.2)	85(32.2)	1.0(ref)	
Yes	87	26(29.9)	34(39.1)	27(31.0)	0.96(0.59-1.56)	0.864	66	10(15.2)	31(47.0)	25(37.9)	1.49(0.88-2.52)	0.142
Supervision												
No	328	100(30.5)	128(39.0)	100(30.5)	1.0(ref)		301	77(25.6)	127(42.2)	97(32.2)	1.0(ref)	
Yes	39	8(20.5)	8(20.5)	23(59.0)	2.55(1.23-5.29)	0.012	29	6(20.7)	10(34.5)	13((44.8)	1.82(0.82-4.03)	0.142
AS poster												
No	143	52(36.4)	56(39.2)	35(24.5)	1.0(ref)		173	48(27.7)	78(45.1)	47(27.2)	1.0(ref)	
Yes	224	56(25.0)	80(35.7)	88(39.3)	1.94(1.19-3.17)	0.008	157	35(22.3)	59(37.6)	63(40.1)	1.59(1.01-2.51)	0.047
AS in stock												
No	91	37(40.7)	33(36.3)	21(23.1)	1.0(ref)		73	29(39.7)	26(35.6)	18(24.7)	1.0(ref)	
Yes	276	71(25.7)	103(37.3)	102(37.0)	2.17(1.22-3.86)	0.008	257	54(21.0)	111(43.2)	92(35.8)	2.02(1.12-3.65)	0.020
Survey												
Baseline	185	61(33.0)	78(42.2)	46(24.9)	1.0(ref)		164	46(28.0)	70(42.7)	48(29.3)	1.0(ref)	
Follow up	182	47(25.8)	58(31.9)	77(42.3)	1.85(0.86-2.76)	0.002	166	37(22.3)	67(40.4)	62(37.3)	1.47(0.97-2.21)	0.067

Table 4. 7 Univariable ordinal logistic regression analysis of predictors of artesunate preparation knowledge, by hospital ownership

	GOK hospitals						FBO hospitals					
	N	**Low n (%)**	**Medium n (%)**	**High n (%)**	**OR (95% CI)**	**p-value**	**N**	**Low n (%)**	**Medium n (%)**	**High n (%)**	**OR (95% CI)**	**p-value**
Age												
35-70 years	138	21(15.2)	44(31.9)	73(52.9)	1.0(ref)		57	5(8.8)	14(24.6)	38(66.7)	1.0(ref)	
21-35 years	229	29(12.7)	90(39.3)	110(48.0)	0.89(0.59-1.34)	0.570	271	26(9.6)	100(36.9)	145(53.5)	0.60(0.32-1.11)	0.103
Sex												
Female	227	34(15.0)	83(36.6)	110(48.5)	1.0(ref)		169	13(7.7)	56(33.1)	100(59.2)	1.0(ref)	
Male	140	16(11.4)	51(36.4)	73(52.1)	1.20(0.80-1.82)	0.377	161	19(11.8)	58(36.0)	84(52.2)	0.70(0.45-1.10)	0.120
Cadre												
Nurse	192	26(13.5)	62(32.3)	104(54.2)	1.0(ref)		174	19(10.9)	53(30.5)	102(58.6)	1.0(ref)	
Clinician	175	24(13.7)	72(41.1)	79(45.1)	0.74(0.50-1.10)	0.142	156	13(8.3)	61(39.1)	82(52.6)	0.87(0.56-1.34)	0.520
Ward												
Medical	182	31(17.0)	76(41.8)	75(41.2)	1.0(ref)		162	17(10.5)	56((34.6)	89(54.9)	1.0(ref)	
Paediatric	185	19(10.3)	58(31.4)	108(58.4)	2.00(1.34-2.99)	0.001	168	15(8.9)	58(34.5)	95(56.5)	1.10(0.71-1.69)	0.671
Endemicity												
Low	265	37(14.0)	95(35.8)	133(50.2)	1.0(ref)		242	24(9.9)	79(32.6)	139(57.4)	1.0(ref)	
High	102	13(12.7)	39(38.2)	50(49.0)	1.00(0.61-1.64)	0.985	88	8(9.1)	35(39.8)	45(51.1)	0.82(0.46-1.45)	0.493
CM Guidelines												
No	249	35(14.1)	97(39.0)	117(47	1.0(ref)		198	20(10.1)	73(36.9)	105(53.0)	1.0(ref)	
Yes	118	15(12.7)	37(31.40	66(5.9)	1.37(0.88-2.11)	0.157	131	11(8.4)	41(31.3)	79(60.3)	1.33(0.84-2.12)	0.229
CM training												
No	280	43(15.4)	105(37.5)	132(47.1)	1.0(ref)		264	28(10.6)	95(36.0)	141(53.4)	1.0(ref)	
Yes	87	7(8.0)	29(33.3)	51(58.6)	1.66(1.02-2.70)	0.041	66	4(6.1)	19(28.8)	43(65.2)	1.80(1.00-3.24)	0.049
Supervision												
No	328	45(13.7)	119(36.3)	164(50.0)	1.0(ref)		301	31(10.3)	103(34.2)	167(55.5)	1.0(ref)	
Yes	39	5(12.8)	15(38.5)	19(48.7)	0.95(0.50-1.83)	0.888	29	1(3.4)	11(37.9)	17(58.6)	1.33(0.60-2.96)	0.481
AS poster												
No	143	31(21.7)	46(32.2)	66(46.2)	1.0(ref)		173	19(11.0)	72(41.6)	82(47.4)	1.0(ref)	
Yes	224	19(8.5)	88(39.3)	117(52.2)	1.56(1.01-2.39)	0.043	157	13(8.3)	42(26.8)	102(65.0)	1.97(1.24-3.13)	0.004
AS in stock												
No	91	17(18.7)	35(38.5)	39(42.9)	1.0(ref)		73	12(16.4)	25(34.2)	36(49.3)	1.0(ref)	
Yes	276	33(12.0)	99(35.9)	144(52.2)	1.65(0.98-2.76)	0.058	257	20(7.8)	89(34.6)	148(57.6)	1.61(0.90-2.88)	0.110
Survey												
Baseline	185	29(15.7)	73(39.5)	83(44.9)	1.0(ref)		164	19(11.6)	65(39.6)	80(48.8)	1.0(ref)	
Follow up	182	21(11.5)	61(33.5)	100(54.9)	1.49(1.00-2.21)	0.049	166	13(7.8)	49(29.5)	104(62.7)	1.78(1.15-2.76)	0.009

Table 4. 8 Univariable binary logistic regression analysis of predictors of the knowledge about preferred route of artesunate, by hospital ownership

	GOK hospitals					FBO hospitals				
	N	Low n (%)	High n (%)	OR (95% CI)	p-value	N	Low n (%)	High n (%)	OR (95% CI)	p-value
Age										
35-70 years	138	31(22.5)	107(77.5)	1.0(ref)		57	14(24.6)	43(75.4)	1.0(ref)	
21-35 years	229	47(20.5)	182(79.5)	1.11(0.62-1.99)	0.735	271	44(16.2)	227(83.8)	1.55(0.67-3.61)	0.309
Sex										
Female	227	52(22.9)	175(77.1)	1.0(ref)		169	33(19.5)	136(80.5)	1.0(ref)	
Male	140	26(18.6)	114(81.4)	1.30(0.71-2.38)	0.398	161	25(15.5)	136(84.5)	1.11(0.55-2.22)	0.767
Cadre										
Nurse	192	39(20.3)	153(79.7)	1.0(ref)		174	29(16.7)	145(83.3)	1.0(ref)	
Clinician	175	39(22.3)	136(77.7)	0.89(0.51-1.55)	0.675	156	29(18.6)	127(81.4)	0.91(0.47-1.77)	0.779
Ward										
Medical	182	44(24.2)	138(75.5)	1.0(ref)		162	26(16.0)	136(84.0)	1.0(ref)	
Paediatric	185	34(18.4)	151(81.6)	1.56(0.89-2.73)	0.120	168	32(19.0)	136(81.0)	0.68(0.34-1.33)	0.259
Endemicity										
Low	265	66(24.9)	199(75.1)	1.0(ref)		242	53(21.9)	189(78.1)	1.0(ref)	
High	102	12(11.8)	90(88.2)	2.97(1.03-8.56)	0.044	88	5(5.7)	83(94.3)	6.86(1.28-36.79)	0.025
CM Guidelines										
No	249	58(23.3)	191(76.7)	1.0(ref)		198	36(18.2)	162(81.8)	1.0(ref)	
Yes	118	20(16.9)	98(83.1)	1.68(0.88-3.22)	0.117	131	22(16.8)	109(83.2)	0.77(0.35-1.69)	0.512
CM training										
No	280	67(23.9)	213(76.1)	1.0(ref)		264	47(17.8)	217(82.2)	1.0(ref)	
Yes	87	11(12.6)	76(87.4)	1.99(0.91-4.36)	0.086	66	11(16.7)	55(83.3)	1.18(0.48-2.91)	0.723
Supervision										
No	328	73(22.3)	255(77.7)	1.0(ref)		301	54(17.9)	247(82.1)	1.0(ref)	
Yes	39	5(12.8)	34(87.2)	1.57(0.52-4.72)	0.421	29	4(13.8)	25(86.2)	0.99(0.24-4.09)	0.986
AS poster										
No	143	42(29.4)	101(70.6)	1.0(ref)		173	35(20.2)	138(79.8)	1.0(ref)	
Yes	224	36(16.1	188(83.9	2.44(1.21-4.90)	0.012	157	23(14.6)	134(85.4)	1.78(0.78-4.07)	0.173
AS in stock										
No	91	24(26.4)	67(73.6)	1.0(ref)		73	25(34.2)	48(65.8)	1.0(ref)	
Yes	276	54(19.6)	220(80.4)	1.74(0.80-3.81)	0.164	257	33(12.8)	224(87.2)	5.35(1.61-17.82)	0.006
Survey										
Baseline	185	45(24.3)	140(75.7)	1.0(ref)		164	32(19.5)	132((80.5)	1.0(ref)	
Follow up	182	33(18.1)	149(81.9)	1.57(0.89-2.74)	0.118	166	26(15.7)	140(84.3)	1.46(0.75-2.86)	0.259

4.3.4 The impact of correlation on the outcome variables adjusting for structures

The impact of correlation on the dichotomous and polytomous outcome was performed using multilevel mixed effect ordinal and binary logistic regression modelling (Tables 4.9-4.13). Adjusting for health facilities and counties structures, the coefficients and their standard errors, confidence intervals (CI) for the odds ratio were slightly wider. In addition, the models that ignored the influence a cluster can exert on the outcome, potentially yielded false conclusions that accommodated more variables being significant, and upon accounting for the clustering at health facilities and counties levels of the hierarchy, concluded that the associations of some of the variables were no longer significant.

Conditional on the fixed-effects covariates (Table 4.14), **s**evere malaria treatment policy was only slightly correlated within the same county. Random effects composed about 5% and 11% of the total residual variance in the GOK and FBO sector respectively. Artesunate dose was slightly correlated within the same health facility. Random effects composed about 19% and 26% of the total residual variance in the GOK and FBO sector respectively. Artesunate dosing interval was slightly correlated within the same health facility. Random effects composed about 18% and 7% of the total residual variance in the GOK and FBO sector respectively. Artesunate preparation was only slightly correlated within the same county. Random effects composed about 4% and 7% of the total residual variance in the GOK and FBO sector respectively.

4.3.5 Predictors of health workers' knowledge about treatment recommendations

At least one knowledge outcome in the government sector passed the selection threshold of $P<0.15$ for multivariable analyses, but only the ward allocation of health workers in the FBO

sector failed to meet these criteria. Multivariable results for government and FBO hospitals are provided in Tables 4.9-4.13, for each of the knowledge outputs studied.

With respect to the treatment policy knowledge, clinicians compared to nurses were more likely to have high knowledge, both at the government (adjusted odds ratio [aOR] =1.86; 95% CI: 1.18-2.91) and FBO hospitals (aOR=2.27; 95% CI=1.41-3.65). In the government hospitals, health worker's treatment policy knowledge was also statistically significantly associated with training exposure (aOR=2.31; 95% CI=1.44-3.72) and follow-up surveys (aOR=1.83; 95% CI=1.22-2.74) while at the FBO hospitals the artesunate availability (aOR=2.01; 95% CI=1.05-3.85) and access to guidelines (aOR=2.41; 95% CI=1.48-3.93) were significant predictors (Table 4.9).

Health workers' knowledge about recommended artesunate dosing was statistically significantly associated with displayed artesunate administration posters (aOR=2.17; 95% CI=1.24-3.79), among paediatric compared to medical ward health workers (aOR=1.99; 95% CI= 1.30-3.04) and during the follow-up compared to the baseline survey (aOR=2.01; 95% CI= 1.28-3.16) within government hospitals. At the FBO hospitals, only health workers' cadre was significant where clinicians were more likely to have correct dosing knowledge than nurses (aOR=2.24; 95% CI=1.33-3.77) (Table 4.10).

Regarding the knowledge of artesunate dosing intervals (Table 4.11), the availability of artesunate (aOR=2.18; 95% CI=1.20-3.94) and health workers' cadre (aOR: 1.76; 95% CI=1.15-2.69) were statistically significantly associated at the FBO hospitals while at the government hospitals the only statistically significant predictor was the follow up compared to baseline survey (aOR=1.55; 95% CI=1.02-2.37).

The knowledge of preferred artesunate administration route via intravenous slow bolus was statistically significantly higher in high compared to low malaria risk areas, both among government (aOR=2.97; 95% CI=1.04-8.46) and FBO health workers (aOR=5.79; 95% CI=1.17-28.67). Furthermore, the same knowledge outcome was associated with displayed artesunate posters (aOR=2.19; 95% CI=1.05-4.57) in the government hospitals and the artesunate availability (aOR=4.73; 95% CI=1.50-14.89) in FBO hospitals (Table 4.13). Finally, only one significant predictor, paediatric compared to medical ward allocation (aOR=1.99; 95% CI=1.33-2.99) at the government hospitals was associated with the knowledge about artesunate preparation (Table 4.12).

Table 4. 9 Predictors of health workers knowledge about severe malaria treatment policy

Health workers' knowledge on treatment policy in the GOK and FBO sectors										
	Unadjusted for clustering					Adjusted for clustering*				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
GOK sector										
Sex										
Female			1.0(ref)					1.0(ref)		
Male	0.16	0.22	1.17	0.75-1.81	0.486	0.16	0.23	1.17	0.74-1.85	0.497
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.57	0.22	1.77	1.14-2.75	**0.011**	0.62	0.23	1.86	1.18-2.91	**0.007**
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.81	0.23	2.25	1.43-3.53	**0.000**	0.84	0.24	2.31	1.44-3.72	**0.001**
Supervision										
No			1.0(ref)					1.0(ref)		
Yes	0.5	0.33	1.65	0.87-3.14	0.127	0.5	0.34	1.65	0.84-3.22	0.065
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.23	0.2	1.26	0.84-	0.267	0.19	0.23	1.21	0.78-	0.398

Health workers' knowledge on treatment policy in the GOK and FBO sectors										
Parameter estimates	Unadjusted for clustering					Adjusted for clustering*				
	Coefficient	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
				1.87					1.90	
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.56	0.2	1.75	1.18-2.60	**0.005**	0.6	0.21	1.83	1.22-2.74	**0.004**
FBO sector										
Sex										
Female			1.0(ref)					1.0(ref)		
Male	-0.02	0.23	0.98	0.62-1.55	0.944	0.02	0.25	1.02	0.63-1.65	0.941
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.78	0.23	2.18	1.38-3.45	**0.001**	0.82	0.24	2.27	1.41-3.65	**0.001**
CM Guidelines										
No			1.0(ref)					1.0(ref)		
Yes	0.82	0.23	2.26	1.45-3.53	**<0.001**	0.88	0.25	2.41	1.48-3.93	**<0.001**
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.33	0.28	1.38	0.81-2.39	0.238	0.29	0.3	1.33	0.74-2.39	0.335
AS poster										
No			1.0(ref)							
Yes	0.2	0.25	1.22	0.75-2.00	0.426	0.07	0.31	1.07	0.59-1.95	0.829
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.71	0.26	2.03	1.21-3.40	**0.007**	0.7	0.33	2.01	1.05-3.85	**0.036**
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.37	0.25	1.45	0.88-2.37	0.141	0.48	0.28	1.61	0.94-2.77	0.085

CM Case Management; *AS* Artesunate; * adjusted for health facility and county structures

Table 4. 10 Predictors of health workers knowledge about artesunate dose

Health workers' knowledge on artesunate dose in the GOK and FBO sectors										
	Unadjusted for clustering					Adjusted for clustering*				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
GOK sector										
Age										
35-70 years			1.0(ref)					1.0(ref)		
21-35 years	0.32	0.26	1.38	0.82-2.30	0.217	0.53	0.29	1.7	0.96-3.01	0.071
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.21	0.25	1.24	0.75-2.04	0.4	0.13	0.28	1.14	0.66-1.95	0.645
Ward										
Medical			1.0(ref)					1.0(ref)		
Paediatric	0.62	0.21	1.85	1.24-2.78	0.003	0.69	0.22	1.99	1.30-3.04	0.002
CM Guidelines										
No			1.0(ref)					1.0(ref)		
Yes	0.24	0.22	1.27	0.82-1.98	0.281	0.4	0.25	1.49	0.91-2.42	0.11
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.72	0.22	2.06	1.34-3.15	0.001	0.77	0.28	2.17	1.24-3.79	0.007
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.18	0.25	1.2	0.74-1.94	0.46	0.32	0.33	1.38	0.72-2.65	0.335
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.68	0.21	1.97	1.30-2.99	0.001	0.7	0.23	2.01	1.28-3.16	0.002
FBO sector										
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.7	0.24	2.02	1.26-3.23	0.003	0.81	0.27	2.24	1.33-3.77	0.002
Supervision										

Health workers' knowledge on artesunate dose in the GOK and FBO sectors										
	Unadjusted for clustering					Adjusted for clustering*				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
No			1.0(ref)					1.0(ref)		
Yes	1.03	0.56	2.8	0.93-8.40)	0.066	1.21	0.65	3.37	0.94-12.0)	0.061
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.59	0.28	1.8	1.04-3.10	0.035	0.7	0.39	2.02	0.93-4.37	0.074
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.25	0.27	1.28	0.75-2.19	0.359	0.27	0.32	1.31	0.70-2.45	0.392

CM Case Management; *AS* Artesunate; * adjusted for health facility and county structures

Table 4. 11 Predictors of health workers knowledge about artesunate dosing interval

Health workers' knowledge on artesunate dosing interval in the GOK and FBO sector										
	Unadjusted for clustering					Adjusted for clustering				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
GOK sector										
Age										
35-70 years			1.0(ref)					1.0(ref)		
21-35 years	0.01	0.24	1.01	0.63-1.63	0.959	0.18	0.27	1.19	0.71-2.02	0.509
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.28	0.24	1.32	0.82-2.12	0.252	0.21	0.26	1.23	0.75-2.05	0.412
Ward										
Medical			1.0(ref)					1.0(ref)		
Paediatric	0.35	0.2	1.42	0.96-2.08	0.076	0.4	0.21	1.49	1.00-2.23	0.052
CM Guidelines										
No			1.0(ref)					1.0(ref)		
Yes	0.24	0.22	1.27	0.83-1.94	0.269	0.43	0.24	1.54	0.97-2.45	0.069
Supervision										
No			1.0(ref)					1.0(ref)		

Health workers' knowledge on artesunate dosing interval in the GOK and FBO sector										
	Unadjusted for clustering					Adjusted for clustering				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
Yes	0.84	0.36	2.32	1.14-4.69	0.02	0.74	0.4	2.09	0.96-4.55	0.062
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.4	0.21	1.49	0.98-2.25	0.061	0.36	0.27	1.43	0.85-2.42	0.179
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.39	0.24	1.47	0.92-2.35	0.107	0.42	0.32	1.52	0.82-2.82	0.183
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.4	0.2	1.5	1.00-2.24	0.047	0.44	0.22	1.55	1.02-2.37	0.042
FBO sector										
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	0.53	0.21	1.69	1.12-2.56	0.013	0.56	0.22	1.76	1.15-2.69	0.009
CM Guidelines										
No			1.0(ref)					1.0(ref)		
Yes	0.45	0.23	1.57	1.01-2.45	0.045	0.46	0.24	1.59	0.99-2.55	0.054
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.23	0.27	1.26	0.74-2.14	0.398	0.21	0.28	1.24	0.71-2.16	0.457
Supervision										
No			1.0(ref)					1.0(ref)		
Yes	0.29	0.38	1.33	0.63-2.83	0.451	0.35	0.41	1.42	0.64-3.16	0.391
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.45	0.25	1.57	0.97-2.54	0.067	0.4	0.28	1.49	0.86-2.60	0.154
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.82	0.26	2.27	1.37-	0.002	0.78	0.3	2.18	1.20-	0.01

Health workers' knowledge on artesunate dosing interval in the GOK and FBO sector										
Parameter estimates	Unadjusted for clustering					Adjusted for clustering				
	Coefficient	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
				3.76					3.94	
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.04	0.24	1.04	0.64-1.67	0.879	0.09	0.26	1.09	0.66-1.82	0.733

CM Case Management; *AS* Artesunate; * adjusted for health facility and county structures

Table 4. 12 Predictors of health workers knowledge about artesunate preparation

Health workers' knowledge on artesunate preparation in the GOK and FBO sector										
Parameter estimates	Unadjusted for clustering					Adjusted for clustering				
	Coefficient	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
GOK sector										
Cadre										
Nurse			1.0(ref)					1.0(ref)		
Clinician	-0.22	0.2	0.8	0.54-1.20	0.287	-0.23	0.21	0.8	0.53-1.20	0.272
Ward										
Medical			1.0(ref)					1.0(ref)		
Paediatric	0.67	0.2	1.95	1.31-2.91	0.001	0.69	0.21	1.99	1.33-2.99	0.001
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.45	0.24	1.57	0.97-2.54	0.064	0.46	0.25	1.58	0.96-2.60	0.071
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.36	0.22	1.43	0.93-2.18	0.101	0.34	0.24	1.4	0.88-2.22	0.154
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.23	0.24	1.26	0.78-2.03	0.338	0.3	0.27	1.35	0.79-2.30	0.278
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.27	0.21	1.3	0.86-1.98	0.202	0.26	0.21	1.3	0.85-1.98	0.221

Health workers' knowledge on artesunate preparation in the GOK and FBO sector										
Parameter estimates	Unadjusted for clustering					Adjusted for clustering				
	Coefficient	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
FBO sector										
Age										
35-70 years			1.0(ref)					1.0(ref)		
21-35 years	-0.35	0.32	0.71	0.38-1.31	0.271	-0.37	0.33	0.7	0.36-1.31	0.26
Sex										
Female			1.0(ref)					1.0(ref)		
Male	-0.35	0.22	0.71	0.45-1.10	0.121	-0.40	0.24	0.67	0.42-1.06	0.09
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.49	0.29	1.64	0.92-2.90	0.092	0.6	0.31	1.82	0.99-3.35	0.056
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.54	0.26	1.72	1.04-2.86	0.036	0.51	0.3	1.67	0.93-2.99	0.087
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	0.47	0.26	1.6	0.95-2.68	0.076	0.48	0.31	1.61	0.88-2.96	0.121
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up			1.3	0.79-2.16	0.301	0.32	0.27	1.38	0.81-2.36	0.24

CM Case Management; *AS* Artesunate; * adjusted for health facility and county structures

Table 4. 13 Predictors of health workers knowledge about artesunate route of administration

Health workers' knowledge on artesunate route of administration in the GOK and FBO sector										
	Unadjusted for clustering					Adjusted for clustering				
Parameter estimates	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**	**Coefficient**	**Std. Error**	**OR**	**95% CI**	**p-value**
GOK sector										
Ward										
Medical			1.0(ref)					1.0(ref)		
Paediatric	0.38	0.27	1.46	0.87-2.46	0.151	0.45	0.29	1.56	0.89-2.76	0.123
Endemicity										
Low			1.0(ref)					1.0(ref)		
High	0.89	0.34	2.44	1.24-4.81	0.01	1.09	0.53	2.97	1.04-8.46	0.042
CM Guidelines										
No			1.0(ref)					1.0(ref)		
Yes	0.14	0.3	1.15	0.63-2.08	0.647	0.25	0.34	1.29	0.66-2.52	0.461
CM training										
No			1.0(ref)					1.0(ref)		
Yes	0.76	0.36	2.13	1.04-4.34	0.038	0.67	0.41	1.95	0.86-4.39	0.108
AS poster										
No			1.0(ref)					1.0(ref)		
Yes	0.71	0.28	2.04	1.19-3.50	0.01	0.78	0.38	2.19	1.05-4.57	0.037
Survey										
Baseline			1.0(ref)					1.0(ref)		
Follow up	0.18	0.27	1.19	0.70-2.04	0.515	0.18	0.31	1.2	0.66-2.20	0.549
FBO sector										
Endemicity										
Low			1.0(ref)					1.0(ref)		
High	1.42	0.49	4.13	1.58-10.84	0.004	1.76	0.82	5.79	1.17-28.67	0.031
AS in stock										
No			1.0(ref)					1.0(ref)		
Yes	1.16	0.31	3.19	1.72-5.92	0.000	1.55	0.59	4.73	1.50-14.89	0.008

CM Case Management; *AS* Artesunate; * adjusted for health facility and county structures

Table 4.14 Random effects and Intracluster correlation

	Health workers' knowledge				
	Treatment policy	**Artesunate dose**	**Artesunate dosing interval**	**Artesunate preparation**	**Artesunate route of administration**
GOK sector					
Random effects					
County [var (CI)]	0.19(0.04-0.90)	9.31e-33	2.08e-29	0.14(0.02-1.04)	1.13(0.44-2.92)
HF [var (CI)]	2.11e-30	0.78(0.37-1.67)	0.70(0.32-1.56)	3.56e-30	6.71e-33
Intracluster correlation					
ICC County	0.05	2.286e-33	5.216e-30	0.04	0.26
ICC HF	0.05	0.19	0.18	0.04	0.26
Observations	367	367	367	367	367
FBO sector					
Random effects					
County [var (CI)]	0.40(0.14-1.16)	7.64e-33	0.23(0.06-0.98)	0.25(0.05-1.18)	2.45e-32
HF [var (CI)]	2.10e-33	1.16(0.54-2.48)	7.71e-32	6.88e-35	2.06(0.75-5.69)
Intracluster correlation					
ICC County	0.11	1.718e-33	0.07	0.07	4.571e-33
ICC HF	0.11	0.26	0.07	0.07	0.39
Observations	327	330	329	328	330

HF-Health Facility; ICC-Intracluster correlation

4.4 Discussion

When modelling using fixed effects in analyses that are not cluster-adjusted, it is difficult to isolate the effect of covariates at the group level (Lesaffre *et.al.,* 2011). Using the ordinary

standard error ignoring clustering will lead to confidence intervals which are too narrow and *P*-values which are too small (Mansournia et al., 2021). Ignoring independence when analyzing cluster correlated data result into biased estimates of standard errors leading to invalid test statistics and CI, hence, misleading inferences (Sainani, 2010; Cameron & Miller 2015; Greenland *et al*., 2016). The model had a random intercept for health facility and county (health facility nested in counties) indicating the amount of variability between health facility, and between counties, in the outcome. The random effects in the model indicated the amount of variability within and between health facilities and counties in the outcome. The intracluster correlation (ICC) at the county level was close to zero or negligible; while at the health facility level the ICC had slight variability with positive CIs across all outcomes. The ICC considered in robust standard error estimations led to accurate inferences (Cameron & Miller 2015; Greenland *et al*., 2016). Both levels had mixed slight variability across the outcomes and could consider running two-level model (Sommet & Morselli, 2017).

Inpatient health workers' knowledge of artesunate-based treatment guidelines for severe malaria at both government and non-governmental organizations (FBO) hospitals in Kenya was sub-optimal five years after the change in national policy. For artesunate policy implementers in Kenya and other African nations, this study indicates a number of important patterns. Health care providers in the public sector have reported better understanding of the new severe malaria treatment policy after undergoing in-service training, but this has not been found to affect the more nitty-gritty standard aspects of artesunate treatment such as dosage, dosing intervals, how the drug should be administered or the preferred route of administration. When it came to the frequency with which FBO hospital health workers were exposed to malaria prevention guidelines, a similar tendency emerged. In Kenya (Toda *et al*., 2018) and elsewhere in Africa,

the lack of training and exposure to malaria guideline in association with health workers' knowledge has been documented (Shavo *et al*., 2014; Jinadu *et al*., 2018; Kurtz, 2016). Observed limited beneficial effects may be due to the short time allocated to severe malaria (only 2-3 hours) within the 3-day malaria case management training curriculum, the inability of guidelines to transfer subtler knowledge information, or the suboptimal quality of implementation when interventions like in-service training are delivered programmatically on a large, national scale (Kurtz, 2016; Eboreime *et al*., 2019).

The correct dosing knowledge was aided by exposed posters in the ward that stated the suggested dosage, preparation, and administration of the drug. This was in contrast to training and access to treatment guidelines. The findings are consistent with a previous study in Kenya that found that healthcare professionals' understanding of surveillance was boosted by poster reminders (Toda *et al*., 2018). Even if job aids have been shown to have a positive impact on the government sector, commodity availability appears to be the most important contextual factor in FBO hospitals when it comes to treatment policy and knowledge about artesunate use. Lack of subsidized artesunate may be a factor in the lack of knowledge in FBO hospitals, where absence of artesunate is unlikely due to previous stock-outs, but more likely the result of failed implementation due to the high cost of artesunate.

In comparison to nurses, clinicians have greater degrees of expertise. The pre-service training effect was expectedly marked in terms of drug policy knowledge and dosing, given the non-prescribing role of nurses in the inpatient context. In the hospital inpatient context, nurses who regularly execute this duty do not have more advanced expertise of artesunate preparation. Similarly, in Tanzania, doctors' expertise of artesunate preparation was higher than that of nurses (Mikomangwa *et al*., 2019). Artesunate preparation and dosing knowledge is higher among

paediatric ward health care personnel in government institutions, which may reflect a larger emphasis on paediatric malaria care in the past (Murray *et al*., 2012; Desai *et al*., 2014). Adult patients in the hospital wards should be the focus of future treatments.

Correct treatment policy and suggested dose have been linked to follow-up surveys at the government hospitals, even though only 5% of health personnel are surveyed on a regular basis. Surveys at the study hospitals may not only be a measurement and monitoring activity, but may also be an intervention in and of themselves, based on the right responses to knowledge assessments, the dissemination of national malaria guidelines and artesunate posters.

Finally, a few caveats must be made clear. First and foremost, our findings do not represent actual clinical practices, despite the fact that understanding of the new treatment strategy is an important prerequisite for its implementation. As a second point, these findings only relate to large government and FBO hospitals, but they do not reflect the knowledge and determinants of smaller inpatient facilities. Finally, it's possible that some of the significant associations were missed due to insufficient power, whereas additional comparisons could have shown asociations that were purely coincidental.

4.5 Conclusions

Multilevel modelling allowed data to be analyzed at one level while accounting for variance at other levels resulting in more accurate estimates. In order to improve government health workers' awareness of artesunate-based severe malaria treatment recommmendations, programmatic interventions such as posters in the wards, targeted health workers in the medical wards, and knowledge assessments are likely to be effective. FBO hospitals should prioritise the provision of artesunate and the training of nurses.

CHAPTER FIVE

OBJECTIVE TWO: APPLYING BAYESIAN HIERARCHICAL APPROACH TO ANALYSE POLYTOMOUS DATA, ADJUSTING FOR COUNTY STRUCTURES IN KENYA.

In this objective, application of Bayesian hierarchical ecological spatial modelling beyond predictor analysis was developed to test for the best fitting model to predict subnational artesunate knowledge levels across 47 counties in Kenya, 2019. This chapter provides specific background information (Section 5.1), methods (Section 5.2) statistical analysis (Section 5.3) and results in Section 5.4. It ends with a discussion in Section 5.5. This scientific work has been published by *BMJ Open.*

5.1 Background

Malaria is a major public health problem. In 2012, the WHO recommended the use of parenteral artesunate for the treatment of severe malaria (WHO, 2015b). This treatment policy was adopted and implemented across malaria-endemic countries in Africa (MoH, 2015b; WHO, 2015b). Health workers' knowledge of evidence-based treatment recommendations is one of the basic requirements for a healthcare system's readiness to implement any new drug policy. In Kenya, concerted efforts have been made to support the WHO policy and monitor its implementation using various health facility surveys that report national levels, trends, and predictors of artesunate knowledge deficiencies among hospital health workers (Zurovac *et al*., 2018). Similar cross-sectional studies have reported inadequate health workers' knowledge of artesunate-based treatment recommendations (Ojo *et al.*, 2020; Mikomangwa *et al.*, 2019). The data from these studies were multi-level and spatially correlated in nature. Traditionally, such data have been

analysed by applying cluster adjustments and correlation matrices based on theoretical assumptions (Zurovac *et al*., 2018; Zurovac *et al*., 2014; Amboko *et al*., 2020; Moen *et al*., 2016), without considering spatial correlations between clusters (Corani *et al*., 2017; Berger, De Oliveira and Sanso, 2001; Shor *et al*., 2007).

Bayesian hierarchical spatial modelling accounts for correlation by introducing effects at different levels of a hierarchy to estimate random effects together with other model parameters accounting for variability within and between sites (Box and Tiao, 2011; Kruschke and Liddell, 2018; Kruschke, 2010; Kruschke and Vanpaemel, 2015; Sánchez, 2017; Perezgonzalez, 2016). The random effects incorporated into fixed-effects models capture the heterogeneity across clusters in the regression coefficients, accounting for the dependence of observations from the same cluster (Sharifi-Malvajerdi *et al*., 2019; Li and Fearnhead, 2018; Dienes, 2011; Austin and Merlo, 2017), leading to accurate conclusions (Bae *et al*., 2016; Dickinson and Basu, 2005).

The Bayesian multilevel models account for the spatial heterogeneity existing among groups, and the conditional autoregressive (CAR) models spatial autocorrelation based on neighbourhood relationships (Lawson and Lee, 2017; Obaromi, 2019; Aswi *et al*., 2020; Gelman *et al*., 2013; Wang *et al.*, 2018; Liu and Zhu, 2016). In this study, neighbourhood was defined using queen adjacency, where a county was considered a neighbour if it shared either a vertex or a node. In this study, a Bayesian hierarchical ecological spatial model beyond predictor analysis was applied to test for the best fitting model to predict subnational artesunate knowledge levels across 47 counties in Kenya.

5.2 Methods

5.2.1 Data sources

This was secondary analysis of a cross-sectional cluster sample survey health facility data conducted in Kenyan hospitals (Figure 5.1). The methodology has been described in detail (Section 3.2.1).

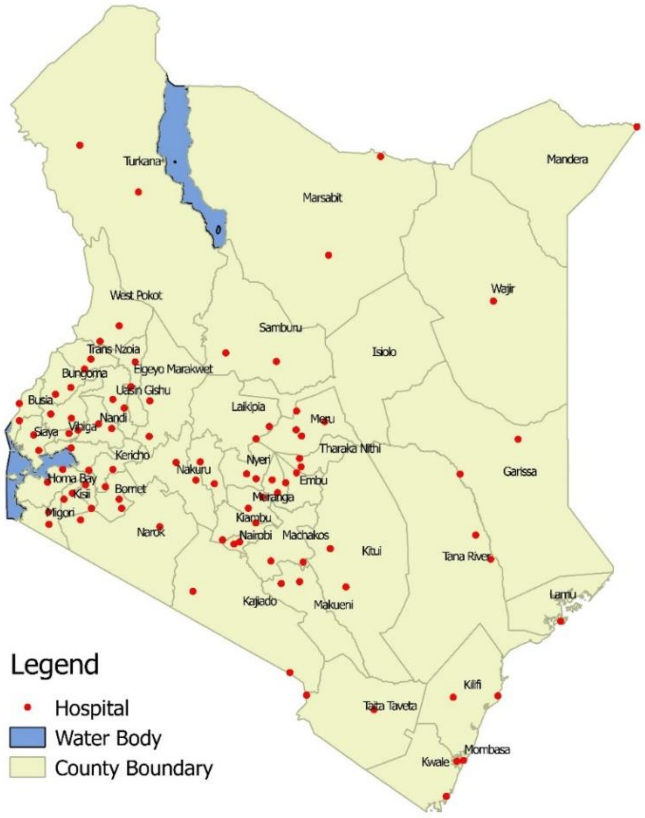


Figure 5. 1 Map of Kenya showing the survey hospitals

5.2.2 Outcomes, definitions, and factors examined

The study considered three response variables reflecting the correctness of the health workers' knowledge about recommended antimalarial treatment for severe malaria, artesunate dose, and preparation. These variables were constructed using a multiple correspondence analysis (MCA)

approach based on variables measured during the survey. MCA is a data analysis technique for nominal categorical data and is used to detect and represent underlying structures in a complex dataset (Ayele and Mwambi, 2014). Prior to computing the MCA, the health worker outcome variables were recoded into dichotomous variables, allowing the variables to take a value of zero or one. The resulting polytomous knowledge response was ordered on a three-point scale: high, medium, or low (Table 4.1).

The variables selected for analysis were based on previous studies (Zurovac *et al*., 2018) and classified as individual or contextual predictors. The health worker level attributes were considered as individual predictors and they included: gender (male *vs.* female), cadre (clinician *vs.* nurse), age (21-30 *vs.* 31-60), years of experience (<10 years vs. >10 years), admission ward allocation (medical *vs.* paediatric), artesunate training (yes *vs.* no), access to malaria guidelines (yes *vs.* no), and access to paediatric protocols (yes *vs.* no). The contextual variables presented heath facility level characteristics, including availability of artesunate (yes *vs.* no), display of artesunate administration posters (yes *vs.* no), availability of artesunate dosing job aids (yes *vs.* no), and malaria endemicity classification (high *vs.* low).

5.3 Statistical analysis

5.3.1 Summary and exploratory analysis

The study utilized descriptive statistics to sum up the demographics of health workers and facilities. Using the Bayesian technique, univariate analysis, estimated odds ratio (OR), and credible intervals (CI) were determined. The multilevel modeling showed significant predictors (80% CI) linked with health workers' knowledge of severe malaria treatment policy, dosage, and preparation of artesunate. An ordinal logistic regression analysis adjusted for clustering at the

county level was then carried out to quantify the predictive variable's (95% CI) effects by fitting three hierarchical models.

5.3.2 Bayesian method for ordinal logistic regression model

The generic form of the binomial models served as the foundation for the development of the ordinal logistic model. When dealing with observations that fit into mutually exclusive categorical groups, a strong class of models called latent variable models can be used to represent the model's logic (Agresti, 2003; Tutz, 2011). Inferences can be drawn more accurately because of the framework's adaptability. As shown below, we started with ordinal logistic regression analysis and then moved on to Bayesian hierarchical spatial modeling.

Let Y_{ij} be a trichotomous outcome variable taking values 1, 2 or 3 if the *j*-th health worker in the *i*-th county $i = 1, \dots, 47$ had low, medium or high artesunate knowledge, respectively.

This variable is a categorized version of a continuous latent (utility) variable defined by

$$Z_{ij} = \eta + \varepsilon_{ij}, \tag{5.1}$$

where η is a predictor depending on covariates and parameters and ε_{ij} is the error term. The two variables Y_{ij} and Z_{ij} are linked by $Y_{ij} = k$ if and only if

$$\theta_{k-1} < Z \leq \theta_k, k = 1,2,3, \tag{5.2}$$

with thresholds $-\infty < \theta_0 < \theta_1 < \theta_2 < \theta_3 < \infty$. In a multinomial logit model setting, the error variables in (1) are independent across the categories and assumed to be standard extreme value distributed with function F. Hence, it follows that Y_{ij} obeys a cumulative logit model. The predictor is then defined as:

$$P(Y_{ij} \leq k|\eta) = F(\theta_k - \eta). \quad (5.3)$$

If F in equation (3) is chosen to be the logistic distribution function, the influence of covariates is modelled using the multinomial logit model given as:

$$P(Y_{ij} = k|Y_{ij} \geq k, \eta) = \frac{\exp(\eta)}{1 + \exp(\eta)} = \theta_k - \eta. \quad (5.4)$$

In this study, the following 3 versions of this cumulative link model for ordinal-scaled observation were implemented as:

$$\text{Model 1: } \log\left(\frac{\gamma_{ijk}}{1-\gamma_{ijk}}\right) = \theta_k - \left(x_{ij}^T\beta + u_i\right),\ i = 1, \dots, 47, j = 1, \dots, n_i, k = 1,2,3, \quad (5.5)$$

$$\text{Model 2: } \log\left(\frac{\gamma_{ijk}}{1-\gamma_{ijk}}\right) = \theta_k - \left(x_{ij}^T\beta + v_i\right),\ i = 1, \dots, 47, j = 1, \dots, n_i, k = 1,2,3, \quad (5.6)$$

$$\text{Model 3: } \log\left(\frac{\gamma_{ijk}}{1-\gamma_{ijk}}\right) = \theta_k - \left(x_{ij}^T\beta + u_i + v_i\right), i = 1, \dots, 47, = 1, \dots, n_i, k\ 1,2,3, \quad (5.7)$$

where,

$$\gamma_{ijk} = P(Y_{ij} \leq k) = \pi_{ij1} + \pi_{ij2} + \cdots + \pi_{ijk}\ with \sum_{k=1}^{3} \pi_{ijk} = 1 \quad (5.8)$$

are cumulative probabilities, η is the linear predictor and x_{ij}^T is a *p*-vector of regression variables for the parameters, β without a leading column for an intercept and F is the inverse link function; $\theta_k, k = 0,1,2,3$ are thresholds for cumulative ordinal logit model, u_i is a spatial structured

component random effect for the $i-$th county with a conditional autoregressive (CAR) distribution $u_i|\boldsymbol{u}_{-i}\sim N\left(\bar{u}_{\delta_i},\frac{\sigma_u^2}{n_{\delta_i}}\right)$, where $\bar{u}_{\delta_i}=n_{\delta_i}{}^{-1}\sum_{j\in\delta_i}u_{j,}$ δ_i and n_{δ_i} represent the set of neighbors and the number of neighbors for the $i-$th county respectively; and v_i is an unstructured spatial random effect for the $i-$th county defined as $v_i\sim N(0,\sigma_v^2)$. The first model and second models (Model 1, Model 2) were ordinal logistic regressions with spatially structured and unstructured random effects respectively, the third model (Model 3) was a convolution model fit by combining both structured and unstructured spatial random effects. In implementing Bayesian analysis, a set of posterior means of the relative risks was then used to create maps to visualize the high to low health workers' knowledge levels by borrowing information from all health workers.

5.3.3 Bayesian Statistical inference

During the model assessment, significant individual and contextual predictors were included in the model simultaneously. The predictive performance of the three hierarchical models was compared using the deviance information criterion (DIC), and a smaller DIC was regarded as a better model. Sensitivity analysis was performed by assuming three chains of Markov Chain Monte Carlo (MCMC) algorithms (Salinelli and Tomarelli, 2014), specifying the same model and prior information from different starting values and comparing the variance within each chain with the variance between chains. Large MCMC samples were used to establish better estimates. In executing this analysis, 10,000 iterations with a burn-in of 500 and thinning of one were run to reduce autocorrelation and avoid bias in the standard error estimate of the posterior mean. Model convergence was assessed using trace plots, histograms, and autocorrelation graphs, monitored by R-hat convergence diagnostic, which is the ratio of the spread of all the

values combined with the mean spread of each chain. The posterior means/odds ratio, quantiles, median, standard deviation, and the corresponding 95% credible interval (CI) were used to assess the significance of all parameters (Edward, Lindman and Savage, 1963). The spatial random effects from the best-fitting model (structured, unstructured, or convolution) of health workers with high knowledge of treatment policy, artesunate dosing, and preparation were overlaid on a map showing all counties in Kenya. Initial analysis was conducted using StataCorp.14 (Stata Statistical Software: Release 14. College Station, TX, StataCorp LP). The Bayesian models were fitted using the R2OpenBUGS statistical package.

5.4 Results

5.4.1 Health worker characteristics

The majority of the 345 health workers interviewed were female (59.7%), aged 21-30 years (62%), had less than 10 years of inpatient experience (82.6%), and 72.5 percent worked in low malaria risk areas. A quarter (24.6 percent) had access to dosage aids, 36.8 percent had been capacity build on the use of artesunate, and 40.9 percent had access to malaria treatment guidelines. The majority of health workers (90.7 percent) worked in hospitals that had artesunate in stock and exposed artesunate administration posters (82.9 percent). In the sample (Table 5.1), the health workers' ward allocation, cadre, and paediatric protocol exposures had a similar distribution.

Table 5. 1 Distribution of the health workers' characteristics

	N=345	
	n	**Percent (%)**
Predictor variables		
Gender		
Male	139	40.3
Female	206	59.7
Health worker cadre		
Clinician	159	46.1
Nurse	186	53.9
Age		
21-30	214	62.0
31-60	131	38.0
Years of experience		
>10years	60	17.4
<10years	285	82.6
Ward allocation		
Medical	170	49.3
Paediatric	175	50.7
Exposure to artesunate interventions		
Trained on artesunate	127	36.8
Malaria treatment guidelines	141	40.9
Paediatric protocol	186	53.9
Artesunate poster	286	82.9
Artesunate dosing wheel	85	24.6
Availability of artesunate	313	90.7
Endemicity		
Low	250	72.5
High	95	27.5

5.4.2 Health workers level of knowledge on severe malaria treatment policy, artesunate dose, and preparation

Using Cronbach's alpha coefficient, dependability of the MCA-derived indices were evaluated, and a value of > 0.7 indicated substantial intra-correlation among the variables (Ayele, Zewotir, and Mwambi, 2014). They were 0.7674 for knowledge of severe malaria treatment policy, 0.8901, and 0.7810 for knowledge of artesunate dose and preparation. The resulting polytomous knowledge response was ordered on a three-point scale: high, medium, or low.

More than a third of health workers had a high level of knowledge about artesunate treatment policy for severe malaria (32.8 percent), whereas 73.9 percent and 70.9 percent of health workers had a high level of knowledge about the recommended dose and preparation of artesunate, respectively (Table 5.2). An ordinal logistic regression analysis was used to examine the relationship between 12 factors and three knowledge outcomes. Of the 12 factors examined, two, ten, and one factor(s) met the inclusion criteria for multivariable analysis (80% CI) with knowledge about treatment policy, artesunate dosing, and artesunate preparation, respectively (Table 5.3–5.5).

Table 5. 2 Knowledge levels about artesunate treatment

Distribution of outcome variables		
Knowledge categories	**N=345**	
	n	**Percent (%)**
Treatment policy		
High	113	32.8
Medium	107	31.0
Low	125	36.2
Dosing		
High	255	73.9

Distribution of outcome variables		
Knowledge categories	**N=345**	
Medium	57	16.5
Low	33	9.6
Artesunate preparation*		
High	244	70.9
Medium	85	24.7
Low	15	4.4

* has one missing value

Table 5. 3 Distribution of the predictor variable in relation to the health workers' knowledge on malaria treatment policy with univariate ordinal logistic regression

	Knowledge on severe malaria treatment policy				
	N	**High**	**Medium**	**Low**	**OR(80%Credible Interval)**
Gender					
Male	139	48(34.5)	42(30.2)	49(35.3)	1 (ref)
Female	206	65(31.6)	65(31.6)	76(36.9)	0.87(0.68;1.14)
Health worker cadre					
Clinician	159	63(39.6)	47(29.6)	49(30.9)	1(ref)
Nurse	186	50(26.9)	60(32.3)	76(40.9)	**0.57(0.44;0.73)**
Age					
21-30	214	68(31.8)	72(33.6)	74(34.6)	1 (ref)
31-60	131	45(34.4)	35(26.7)	51(38.9)	0.90(0.69;1.17)
Years of experience					
>10years	60	17(28.3)	19(31.7)	24(40)	1 (ref)
<10years	285	96(33.7)	88(30.9)	101(35.4)	0.77(0.53;1.09)
Ward allocation					
Medical	170	55(32.4)	49(28.8)	66(38.8)	1 (ref)
Paediatric	175	58(33.1)	58(33.1)	59(33.7)	1.09(0.85;1.40)
Trained on artesunate					
No	218	68(31.2)	62(28.4)	88(40.4)	1 (ref)

	Knowledge on severe malaria treatment policy				
	N	**High**	**Medium**	**Low**	**OR(80%Credible Interval)**
Yes	127	45(35.4)	45(35.4)	37(29.1)	**1.42(1.08;1.87)**
Malaria treatment guidelines					
No	204	67(32.8)	55(27)	82(40.2)	1 (ref)
Yes	141	46(32.6)	52(36.9)	43(30.5)	1.00(0.76;1.29)
Paediatric protocol					
No	159	52(32.7)	44(27.7)	63(39.6)	1 (ref)
Yes	186	61(32.8)	63(33.9)	62(33.4)	1.18(0.91;1.54)
Exposure of artesunate poster					
No	59	22(37.3)	18(30.5)	19(32.2)	1 (ref)
Yes	286	91(31.8)	89(31.1)	106(37)	0.75(0.52;1.06)
Access of artesunate dosing wheel					
No	260	85(32.7)	78(30)	97(37.3)	1 (ref)
Yes	85	28(32.9)	29(34.1)	28(32.9)	1.20(0.89;1.60)
Availability of artesunate					
No	32	12(37.5)	7(21.9)	13(40.7)	1 (ref)
Yes	313	101(32.3)	100(31.9)	112(35.8)	0.95(0.61;1.55)
Endemicity					
Low	250	78(31.2)	75(30)	97(38.8)	1 (ref)
High	95	35(36.8)	32(33.7)	28(29.5)	1.27(0.95;1.70)

Table 5. 4 Distribution of the predictor variable in relation to the health workers' knowledge on artesunate dosing with univariate ordinal logistic regression

	Knowledge on artesunate dose				
	N	**High**	**Medium**	**Low**	**OR(80% Credible Interval (CI)**
Gender					
Male	139	109(78.4)	20(14.4)	10(7.2)	**1 (ref)**
Female	206	146(70.9)	37(18)	23(11.2)	**0.63(0.45; 0.87**)
Health worker cadre					

Knowledge on artesunate dose					
	N	**High**	**Medium**	**Low**	**OR(80% Credible Interval (CI)**
Clinician	159	134(84.3)	16(10.1)	9(5.7)	1 (ref)
Nurse	186	121(65.1)	41(22)	24(12.9)	**0.33(0.24; 0.47)**
Age					
21-30	214	173(80.8)	30(14)	11(5.1)	**1 (ref)**
31-60	131	82(62.6)	27(20.6)	22(16.8)	**0.34(0.24; 0.47)**
Years of experience					
>10years	60	33(55)	12(2.0)	15(25)	**1 (ref)**
<10years	285	222(77.9)	45(1.8)	18(6.3)	**1.58(1.05; 2.35)**
Ward allocation					
Medical	170	118(69.4)	34(20)	18(10.6)	**1 (ref)**
Paediatric	175	137(78.3)	23(13.1)	15(8.6)	**1.58(1.16; 2.17)**
Trained on artesunate					
No	218	158(72.5)	38(17.4)	22(10.1)	1 (ref)
Yes	127	97(76.4)	19(15)	11(8.7)	1.17(0.85; 1.64)
Malaria treatment guidelines					
No	204	146(71.6)	33(16.2)	25(12.3)	1 (ref)
Yes	141	109(77.3)	24(17)	8(5.7)	1.39(0.98; 1.93)
Paediatric protocol					
No	159	102(64.2)	37(23.3)	20(12.6)	1 (ref)
Yes	186	153(82.3)	20(10.8)	13(7)	**1.68(1.22; 2.29)**
Exposure of artesunate poster					
No	59	32(54.2)	15(25.4)	12(20.3)	**1 (ref)**
Yes	286	223(78)	42(14.7)	21(7.3)	**3.00(2.09; 4.37)**
Access of artesunate dosing wheel					
No	260	183(70.4)	48(18.5)	29(11.2)	**1 (ref)**
Yes	85	72(84.7)	9(10.6)	4(4.7)	**2.52(1.66; 4.01)**
Availability of artesunate					
No	32	16(50)	9(28.1)	7(21.9)	**1 (ref)**
Yes	313	239(76.4)	48(15.3)	26(8.3)	**3.07(1.97; 5.09)**

Knowledge on artesunate dose					
	N	**High**	**Medium**	**Low**	**OR(80% Credible Interval (CI)**
Endemicity					
Low	250	176(70.4)	44(17.6)	30(12)	**1 (ref)**
High	95	79(83.2)	13(13.7)	3(3.2)	**2.09(1.41; 3.13)**

Table 5. 5 Distribution of the predictor variable in relation to the health workers' knowledge on artesunate preparation with univariate ordinal logistic regression

Knowledge on artesunate preparation					
	N	**High**	**Medium**	**Low**	**OR(80% Credible Interval (CI)**
Gender					
Male	139	95(68.3)	32(23)	12(8.6)	**1 (ref)**
Female	205	149(72.7)	53(25.9)	3(1.5)	1.27(0.93;1.75)
Health worker cadre					
Clinician	159	111(69.8)	39(24.5)	9(5.7)	1 (ref)
Nurse	185	133(71.9)	46(24.9)	6(3.2)	1.12(0.83;1.52)
Age					
21-30	214	153(71.5)	54(25.2)	7(3.3)	1 (ref)
31-60	130	91(70)	31(23.8)	8(6.2)	0.82(0.60;1.13)
Years of experience					
>10years	59	44(74.6)	14(23.7)	1(1.7)	1 (ref)
<10years	285	200(70.2)	71(24.9)	14(4.9)	0.77(0.50;1.16)
Ward allocation					
Medical	170	116(68.2)	46(27.1)	8(4.7)	1 (ref)
Paediatric	174	128(73.6)	39(22.4)	7(4)	1.29(0.95;1.77)
Trained on artesunate					
No	217	156(71.9)	50(23)	11(5.1)	1 (ref)
Yes	127	88(69.3)	35(27.6)	4(3.1)	0.90(0.66;1.23)
Malaria treatment guidelines					
No	203	135(66.5)	59(29.1)	9(4.4)	1 (ref)

Knowledge on artesunate preparation					
	N	**High**	**Medium**	**Low**	**OR(80% Credible Interval (CI)**
Yes	141	109(77.3)	26(18.4)	6(4.3)	1.34(0.98;1.83)
Paediatric protocol					
No	159	105(66)	46(28.9)	8(5)	1 (ref)
Yes	185	139(75.1)	39(21.1)	7(3.8)	0.81(0.60;1.10)
Exposure of artesunate poster					
No	59	40(67.8)	17(28.8)	2(3.4)	1 (ref)
Yes	285	204(71.6)	68(23.9)	13(4.6)	1.19(0.79;1.76)
Access of artesunate dosing wheel					
No	259	179(69.1)	66(25.5)	14(5.4)	1 (ref)
Yes	85	65(76.5)	19(22.4)	1(1.2)	**1.57(1.09;2.30)**
Availability of artesunate					
No	32	21(65.6)	9(28.1)	2(6.3)	1 (ref)
Yes	312	223(71.5)	76(24.4)	13(4.2)	1.29(0.77;2.15)
Endemicity					
Low	249	176(70.7)	64(25.7)	9(3.6)	1 (ref)
High	95	68(71.6)	21(22.1)	6(6.3)	0.97(0.69;1.38)

Table 5.6-5.8 report the results of the three comparative hierarchical models that were fitted in multivariable analysis and their goodness of fit is compared using deviance information criterion (DIC). For the health workers' knowledge on treatment policy, the DIC for model 1, model 2, and model 3 were 762.71, 780.17, and 770.14, respectively (Table 5.7). Regarding health workers' knowledge on artesunate dosing, the DIC for model 1, model 2, and model 3 were 488.83, 496.19, and 497.80, respectively (Table 5.8). For the health workers' knowledge on artesunate preparation, the DIC for model 1, model 2, and model 3 were 503.31, 510.17, and 507.31, respectively (Table 5.9). Model 1 with spatially structured random effects provided a better fit for the three outcomes. Spatially structured random effects illustrate the necessity of

accounting for spatial autocorrelation, which, if ignored in the regression model, can lead to biased inferences.

The posterior means/odds ratio, quantiles, median, standard deviation, and the corresponding 95% credible interval (CI) were used to assess the significance of all parameters. The posterior estimates were similar across the three hierarchical models and the adjusted odds ratios (aOR) and 95% CI estimates from the best-fitting model are reported. For the outcome on the knowledge about artesunate treatment policy, the health workers' cadre was the only significant predictor. The likelihood of having a high knowledge of severe malaria treatment policy was significantly lower in nurses than in clinicians (aOR=0.59, 95% CI: 0.40 to 0.89). Regarding knowledge of the recommended artesunate dosing, health worker cadre, age, and exposure to artesunate administration poster were significant predictors. Nurses were 52% less likely to have high knowledge about dosing compared to the clinicians (aOR=0.48, 95% CI: 0.25 to 0.87). Health workers older than 30 years were 61% less likely to have high knowledge about dosing compared to younger health workers (aOR=0.39, 95% CI: 0.22 to 0.67), while health workers exposed to artesunate posters had 2.4-fold increased odds of higher knowledge about dosing compared to non-exposed health workers (aOR=2.38, 95% CI: 1.22 to 4.74). Finally, based on unadjusted univariate analysis (Table 5.6), the health workers who had access to an artesunate dosing wheel were 57% more likely to have higher knowledge of artesunate preparation compared to those who did not have access (OR=1.57, 80% CI: 1.09 to 2.30). However, the same predictor variable lost significance at the 95% CI, adjusted for multivariable analysis (aOR=1.58, 95% CI: 0.91 to 2.88), refer to Table 4.

Table 5. 6 Bayesian approach to multivariate ordinal logistic regression using odds ratio (OR), 95% Credible Interval (CI), knowledge on severe malaria treatment policy

		Posterior summary estimates based on 2.5% and 97.5% posterior quantiles		
		Knowledge on severe malaria treatment policy		
		Model 1	**Model 2**	**Model 3**
	n	**OR (95% Cred.int)**	**OR (95% Cred.int)**	**OR (95% Cred.int)**
Fixed effects				
Health worker cadre				
Clinician	159	1 (Ref)	1 (Ref)	1 (Ref)
Nurse	186	**0.59(0.40; 0.89)**	0.58(0.39; 0.86)	0.58(0.38; 0.87)
Trained on artesunate				
No	218	1 (Ref)	1 (Ref)	1 (Ref)
Yes	127	1.38(0.91; 2.08)	1.36(0.89; 2.04)	1.37(0.92; 2.07)
Random effects				
Spatially structured (τu)		313.20(1.36; 5185.57)		480.80(2.13;5184.00)
Spatially unstructured (τv)			111.85(2.62;5093.05)	331.50(3.85;3790.10)
Model fit				
DIC (PD)		**762.71 (20.34)**	780.17 (40.90)	770.14 (29.31)

Model 1: Spatially structured random effects; Model 2: Spatially unstructured random effects; Model 3: Convolution

Table 5. 7 Bayesian approach to multivariate ordinal logistic regression using odds ratio (OR), 95% Credible Interval (CI), knowledge on artesunate dose

		Posterior summary estimates based on 2.5% and 97.5% posterior quantiles		
		Knowledge on artesunate dose		
		Model 1	**Model 2**	**Model 3**
	n	**OR (95% Cred.int)**	**OR (95% Cred.int)**	**OR (95% Cred.int)**
Fixed effects				
Gender				
Male	139	1 (Ref)	1 (Ref)	1 (Ref)
Female	206	0.84(0.45; 1.53)	0.85(0.48; 1.50)	0.84(0.45; 1.58)
Health worker cadre				
Clinician	159	1 (Ref)	1 (Ref)	1 (Ref)
Nurse	186	**0.48(0.25; 0.87)**	0.48(0.26; 0.88)	0.47(0.26; 0.87)
Age				
21-30	214	1 (Ref)	1 (Ref)	1 (Ref)
31-60	131	**0.39(0.22; 0.67)**	0.39(0.23; 0.68)	0.39(0.22; 0.67)
Years of experience				
>10years	60	1 (Ref)	1 (Ref)	1 (Ref)
<10years	285	1.11(0.57; 2.13)	1.10(0.55; 2.29)	1.12(0.60; 2.22)
Ward allocation				
Medical	170	1 (Ref)	1 (Ref)	1 (Ref)
Paediatric	175	1.54(0.91; 2.60)	1.53(0.92; 2.63)	1.53(0.89; 2.54)
Paediatric protocol				
No	159	1 (Ref)	1 (Ref)	1 (Ref)
Yes	186	1.27(0.76; 2.21)	1.29(0.75; 2.18)	1.29(0.74; 2.19)
Artesunate poster				
No	59	1 (Ref)	1 (Ref)	1 (Ref)
Yes	286	**2.38(1.22; 4.74)**	2.33(1.18; 4.63)	2.44(1.22; 4.91)
Artesunate dosing wheel				
No	260	1 (Ref)	1 (Ref)	1 (Ref)

		Posterior summary estimates based on 2.5% and 97.5% posterior quantiles		
		Knowledge on artesunate dose		
		Model 1	**Model 2**	**Model 3**
	n	**OR (95% Cred.int)**	**OR (95% Cred.int)**	**OR (95% Cred.int)**
Yes	85	1.92(0.97; 4.04)	1.94(0.97; 3.98)	1.91(0.95; 4.01)
Availability of artesunate				
No	32	1 (Ref)	1 (Ref)	1 (Ref)
Yes	313	1.94(0.81; 4.35)	2.03(0.81; 4.52)	1.80(0.70; 4.03)
Endemicity				
Low	250	1 (Ref)	1 (Ref)	1 (Ref)
High	95	1.58(0.78; 3.24)	1.61(0.84; 3.24)	1.53(0.69; 3.12)
Random effects				
Spatially structured (τu)		423.15(1.24;5431.52)		482.00(0.60;4948.05)
Spatially unstructured (τv)			411.55(2.51;4587.05)	667.75(12.17;5402.00)
Model fit				
DIC (PD)		**488.83 (21.76)**	496.19 (29.86)	497.80 (33.04)

Model 1: Spatially structured random effects; Model 2: Spatially unstructured random effects; Model 3: Convolution

Table 5. 8 Bayesian approach to multivariate ordinal logistic regression using odds ratio (OR), 95% Credible Interval (CI), knowledge on artesunate preparation

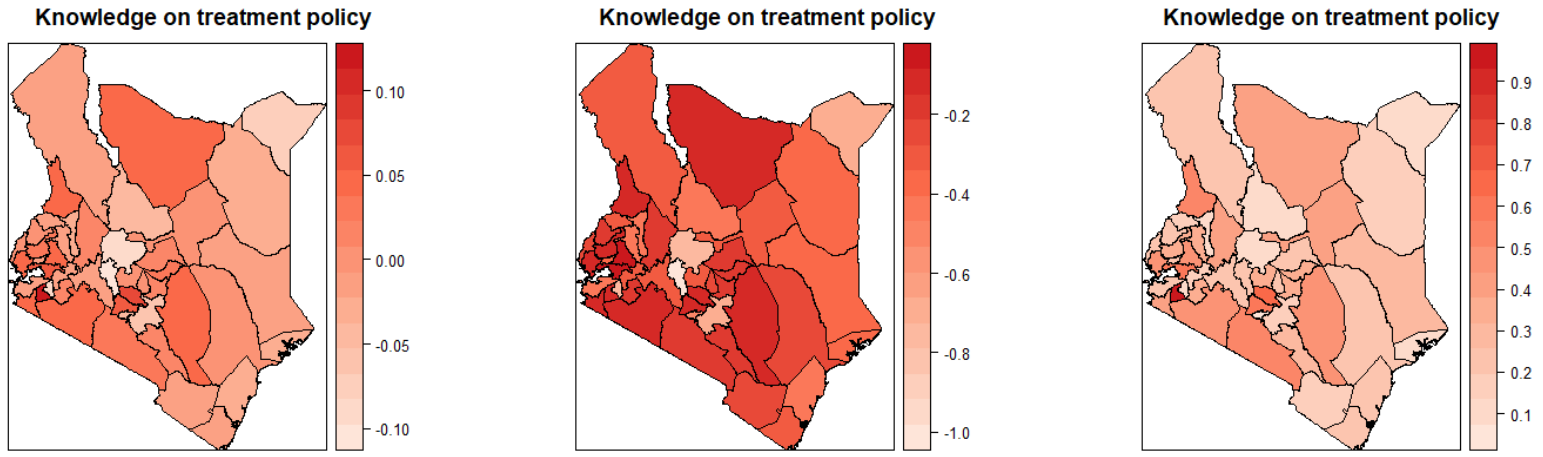
		Posterior summary estimates based on 2.5% and 97.5% posterior quantiles		
		Knowledge on artesunate preparation		
		Model 1	**Model 2**	**Model 3**
	n	**OR (95% Cred.int)**	**OR (95% Cred.int)**	**OR (95% Cred.int)**
Fixed effects				
Artesunate dosing wheel				
No	260	1 (Ref)	1 (Ref)	1 (Ref)
Yes	85	1.58(0.91; 2.88)	1.57(0.90; 2.85)	1.58(0.92; 2.83)
Random effects				
Spatially structured (τu)		549.40(6.96;4847.67)		561.90(7.37;5257.57)
Spatially unstructured (τv)			419.55(11.14;4565.10)	380.60(7.23; 4330.00)
Model fit				
DIC (PD)		**501.22 (4.4)**	**502.38 (6.29)**	504.25 (8.88)

Model 1: Spatially structured random effects; Model 2: Spatially unstructured random effects; Model 3: Convolution

Figures 5.2- 5.4 show the spatial random effects of the posterior means of the probability of health workers having high knowledge of severe malaria treatment policy, artesunate dosing, and preparation, respectively, overlaid on a map showing all counties in Kenya. The deep red colour denotes regions with strictly high knowledge, while the light red colour denotes strictly low knowledge. In Figure 5.2, the health workers in Kisii county had high knowledge levels (>10%) on severe malaria treatment policy, while those in Nyandarua, Nyamira, Laikipia, and Mandera counties had low knowledge levels (<10%). In Figure 5.3, the health workers in Muranga, Kisii, Embu, Uasin Gishu, Kiambu, and Kisumu counties had high knowledge levels (>10%) about artesunate doses, while those in Nyandarua, Nyamira, Garissa, Busia, and Nairobi

counties had low knowledge levels (<10%). In Figure 5.4, there were 17 counties with high knowledge levels (>10%), while 16 counties had low knowledge levels (<10%), on artesunate preparation.

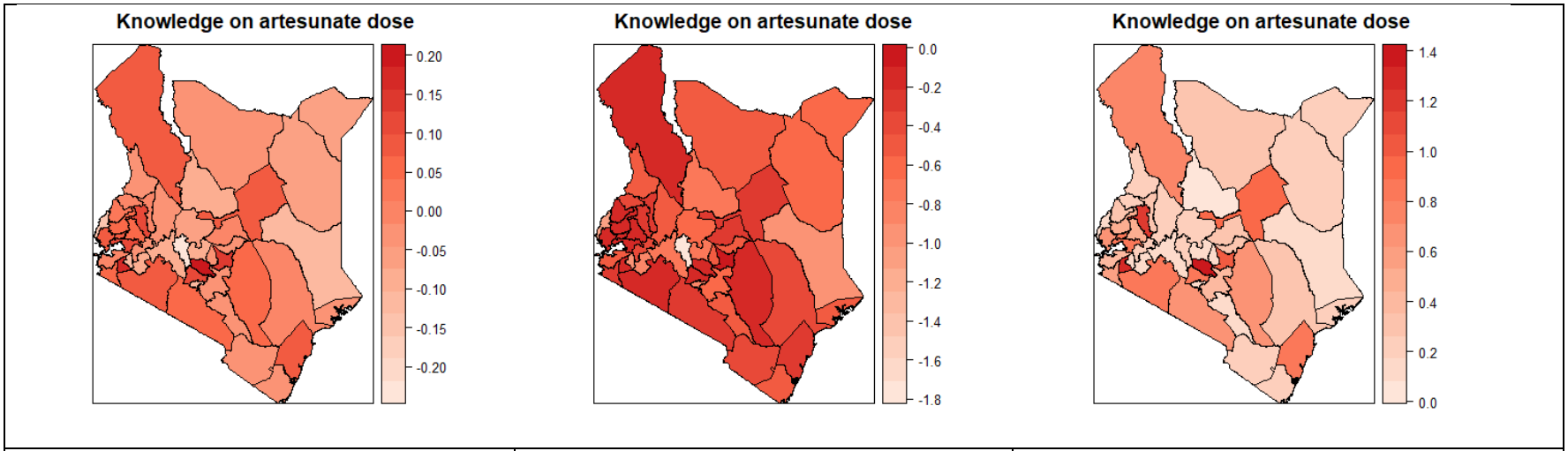
Knowledge on severe malaria treatment policy



(a) Posterior mean	(b) 2.5% quantiles of posterior mean	(c) 97.5% quantiles of posterior mean

Figure 5. 2 Spatially structured random effects on probability of health workers having high knowledge on the recommended treatment policy of severe malaria using artesunate: posterior mean (a), its 2.5% quantiles (b) and 97.5% quantiles (c)

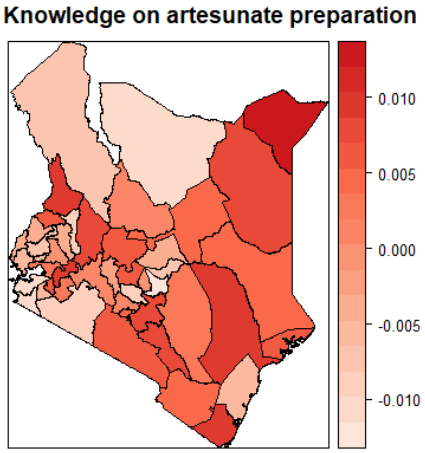
Knowlegde on artesunate dose



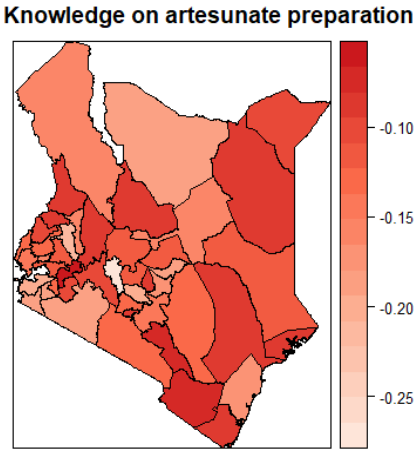
(a) Posterior mean	(b) 2.5% quantiles of posterior mean	(c) 97.5% quantiles of posterior mean

Figure 5. 3 Spatially structured random effects on probability of health workers having high knowledge on artesunate dose: posterior mean (a), its 2.5% quantiles (b) and 97.5% quantiles (c)

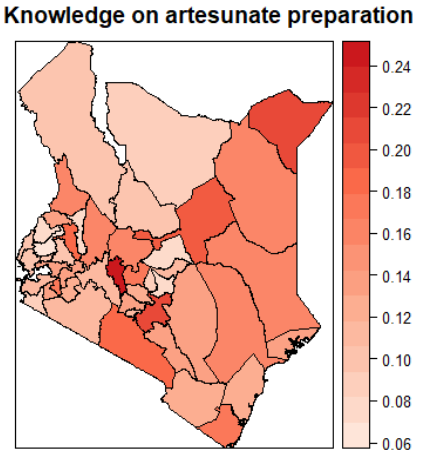
Knowledge on artesunate preparation



(a) Posterior mean



(b) 2.5% quantiles of posterior mean



(c) 97.5% quantiles of posterior mean

Figure 5. 4 Spatially structured random effects on probability of health workers having high knowledge on artesunate preparation: posterior mean (a), its 2.5% quantiles (b) and 97.5% quantiles (c)

5.5 Discussion

An extension to the traditional approach, this study used Bayesian hierarchical ecological spatial modeling in order to assess the national level effects on Kenyan health workers' knowledge of severe malaria treatment policy, artesunate dose, and pre-treatment preparation. A total of 12 individual and contextual factors were incorporated into the models for the treatment of severe malaria, including three ordinal response variables: policy, dose, and preparation of artesunate. Ordinal logistic regression with spatially structured random effects, spatially unstructured random effects, and convolution were modelled. The Bayesian method enabled us to look at factors related to health workers' knowledge levels and the spatial factors surrounding severe malaria treatment policies for targeted malaria interventions.

Spatially structured random effects illustrate the necessity of accounting for spatial autocorrelation among counties for accurate inferences (Dasgupta *et al*., 2014; Hanandita and Tampubolon, 2016). The similarity of health workers' responses in the same facility and the likelihood of similarity between health facility structures in neighbouring counties explain the neighbourhood influence on the spatially structured models.

In the country, not all health workers have received information they need to effectively treat severe malaria. The artesunate treatment policy was well understood by a third of health workers. Similar findings have been observed in other studies (Zurovac *et al*., 2018; Mikomangwa *et al*., 2019). This was associated with the low knowledge levels on the treatment policy for pregnant women in the first and second trimesters. The level of knowledge on the correct artesunate dose for patients weighing less than or more than 20 kilograms was high among nearly three-quarters of the health care providers surveyed. Health workers had suboptimal knowledge about artesunate treatment policies. More training avenues like seminars

and workshops should be organized, and previous methods of training delivery should be re-evaluated. According to previous studies, nurses were less knowledgeable about the WHO's treatment policy for severe malaria than clinicians (Mikomangwa *et al*., 2019; Reeves *et al*., 2017). Efforts to increase nurses' knowledge of severe malaria treatment policy and dosing must be based on interprofessional collaboration and new approaches. Artesunate dosing knowledge was low among the medical professionals over the age of 30 years. This age category is an important part of the workforce and should be retrained in managing severe malaria patients. The artesunate poster helps health workers to better understand the artesunate dose. In Tanzania, healthcare workers prepared injectable artesunate using posters, which led to a similar observation (Mikomangwa *et al*., 2019). To help people remember what they have learned in training, the program should include more posters that are updated, printed, and sent to all health facilities.

There is evidence that health workers' knowledge of severe malaria treatment policies, artesunate dose, and preparation at several county hospitals differ regionally. The best fitting model for severe malaria treatment policy, artesunate dosing, and preparation was fitted with spatially structured random effects. The similarity of health workers' responses in the same facility and the likelihood of similarity between health facility structures in neighboring counties explain the neighborhood influence on the spatially structured models. To account for clustering among health facilities and counties, a Bayesian hierarchical model was needed to account for the substantial heterogeneity among the health workers with high knowledge of treatment policy, artesunate dosing, and preparation at the county level (Hanandita and Tampubolon, 2016). Bayesian hierarchical spatial models assign a normal conditional autoregressive prior to the random effects to account for both the nesting of health workers within health facilities (vertical

dependence) and the geographical autocorrelation within counties (horizontal dependence) (Wang *et al*., 2018; Umer *et al*., 2019). A conditional autoregressive prior is used in Bayesian spatial models to examine small area variations and identify spatial patterns (Dasgupta *et al*., 2014). This shows the importance of taking into consideration the spatial variance among counties in order to draw accurate conclusions from the models. The non-spatial heterogeneity of the analysis units is captured by the spatially unstructured random variables (Mutua *et al*., 2019; Rashidi *et al*., 2016; Achia, 2014). The spatial maps show estimates of knowledge at the subnational level, which can be used to target interventions.

The findings of this study shed light on the knowledge levels of health workers and the factors that influence their knowledge at the subnational level. Using Bayesian modeling, researchers could draw solid conclusions from a wide range of data sources. There are a few limitations to this research. First, due to multiple exploratory data analyses and comparisons, some of the results may have been significant by chance. Second, the knowledge levels were self-reported by health workers in the inpatient departments of the sampled hospitals and should be taken with caution when generalizing results to other institutions. Third, the study determined the health workers' level of knowledge on artesunate treatment but not their actual practice; hence, the results cannot be extended to infer about actual clinical practice.

5.6 Conclusion

Bayesian hierarchical model accounted for the substantial heterogeneity among the health workers with high knowledge of treatment policy, artesunate dosing, and preparation at the county level, while spatial autocorrelation was addressed using Conditional autoregressive prior assigned to the random effects at the county level.

As a contextual factor, exposure to artesunate posters was related to health workers' understanding of severe malaria treatment policies, dose, and preparation. Job training, continuing medical education and case management training with an emphasis on dosage should be considered at the health facility level in order to increase health workers' knowledge of severe malaria management. It is recommended that artesunate posters be displayed at medical facilities as part of a regular education campaign.

A multidisciplinary strategy to bridge the information gaps revealed at the subnational level based on the spatial maps can be used to focus programmatic interventions. This methodology can be used to analyze similar data types and contexts in health surveys. Research into why the existing operational interventions have not increased the knowledge of health professionals about severe malaria treatment policies is suggested.

CHAPTER SIX

OBJECTIVE THREE: MODEL FACTORS ASSOCIATED WITH HOSPITAL LENGTH OF STAY AMONG SEVERE MALARIA PATIENTS USING COMPETING RISK SURVIVAL ANALYSIS APPROACH, ADJUSTING FOR HEALTH FACILITY STRUCTURES IN KENYA.

In order to assess this objective, a competing risk approach was applied to identify factors associated with hospital length of stay (LOS) for patients admitted with suspected severe malaria, based on usual clinical settings in Kenya, 2018. This chapter provides background information (Section 6.1), method (Section 6.2), statistical analysis (Section 6.3) and the results have been discussed in Section 6.4. It ends with discussion in Section 6.5. This chapter scientific work has been published by *BMJ Open.*

6.1 Background

Malaria is a leading cause of morbidity and mortality, disproportionately affecting children under five years of age and pregnant women in many developing countries. Severe malaria is associated with high mortality if untreated within 24 hours (MoH, 2015b). Comprehensive assessment of patients with suspected malaria is recommended on admission and during hospitalisation to optimise care and prevent further complications (MoH, 2015b; WHO, 2000). Patient triaging during routine admissions and monitoring of vital clinical and laboratory measurements in the wards such as temperature, blood pressure, pulse rate, respiratory rate, and assessment of the level of consciousness, oxygen saturation, blood glucose, haemoglobin level, and urine output (White, 2018; Lee *et al*., 2008; Lombardo *et al*., 2017; English *et al*., 1996; Gachot *et al*., 1998; Adebola, Babatunde and Bose, 2014; Taylor *et al*., 2012) are the basic

management standards. Moreover, effective malaria case management comprises appropriate antimalarial and supportive therapy. Finally, it is recommended that patients with suspected severe malaria should have a parasitological diagnosis irrespective of fever and should be managed in a facility with inpatient services with expertise and infrastructure for adequate management (MoH, 2015b).

The patient outcomes of being discharged home after treatment for malaria and the associated hospital LOS are not only dependent on the patient's clinical factors, but also on the quality of case management provided on admission and during hospitalisation (Keene *et al*.,2018; Hoffmeister, 2021). However, the discharge outcome can be interrupted by death as a competing risk (Pintillie, 2011). Competing risk analysis accurately evaluates LOS by estimating the marginal probability of an event in the presence of competing events using the cumulative incidence function (CIF). CIF avoids overestimation and bias resulting from applying general survival models that ignore competing risks (Schuster *et al*., 2020; Austin, Lee and Fine, 2016).

Predicting LOS for patients is an important measure in hospital service planning, resource allocation, and monitoring of the quality of health care (WHO, 2010). Assessing and modifying factors that influence hospital LOS for suspected malaria patients in the presence of competing risk events can lead to the optimisation of service delivery in resource-limited settings (Keene *et al*., 2018). The effect of the factors is determined using the cause-specific or subdistribution hazard function (Austin, Lee and Fine, 2016). The cause-specific hazard (CSH) is estimated by removing individuals from the risk set when they experience the competing event by treating them as censored observations. In addition, CSH can be estimated by fitting a standard Cox proportional hazards (CPH) model that determines the effect of factors on the survival function by assuming that hazard functions are proportional over time (Fine and Gray,

1999). The CSH model is considered more appropriate for etiologic research, as it directly quantifies subjects who are at risk of developing an event of interest (Austin, Lee and Fine, 2016). The subdistribution hazard (SDH) model is also retained within the risk set for subjects who are free of the event and those who experience the competing event. It relies on the precise accounting of the number of subjects who fail because of the event of interest, those who fail because of competing events, and those who are censored (Geskus, 2011). The SDH model is most appropriate for prediction research, given the direct relationship between these factors and CIF (Austin and Fine, 2017).

Identifying the factors that predict the time to discharge is the core of quality of care analysis. However, previous studies on the quality of care for inpatient malaria have assessed levels and trends in system readiness to implement the recommended malaria case management policies following a large clinical trial and change in therapeutic policies (Al Farsi *et al*., 2019; Amboko *et al.*, 2022; Dondorp *et al*., 2010). Studies investigating the factors influencing hospital LOS for malaria are scarce. Only one study was conducted in malaria-endemic areas, notably under controlled clinical trial conditions in areas with low malaria risk in Southeast Asia (Keene *et al*., 2018). Another observational study investigated malaria LOS predictors in a high-resource tertiary hospital in Germany (Hoffmeister, 2021). Studies based in routine clinical settings from low resource but high malaria risk areas in Africa have not been undertaken. In this study, the factors associated with hospital LOS for patients admitted with suspected malaria in the presence of competing risk events were examined in routine clinical settings in Kenya.

6.2 Methods

6.2.1 Description of the data

This was secondary data analysis as described in Section 3.2.1. According to the hospital's record keeping style, data was extracted from all available forms for each of the selected patients' files, including structured and unstructured forms for admission and follow-up and observation and treatment as well as nursing care and discharge forms and laboratory forms.

Hospital admission required a probable malaria diagnosis including any sort of diagnostic, test, or treatment for malaria. On admission or within 24 hours of being admitted to the ward, the presence of clinical criteria for severe malaria was reported. There were no results reported for any patients who had a malaria test ordered either on admission or post-admission and whose results had not been tracked down. It's possible to find out more about the study's approach elsewhere (Zurovac *et al*., 2018).

6.2.2 National standard case definitions for uncomplicated and severe malaria

The Kenyan recommendations (MoH, 2015b) state that if a patient has symptoms of malaria, a positive parasitological test, microscopy, or quick diagnostic test, but no signs of severe malaria, they are considered to have simple malaria. The presence of malaria parasitaemia in the presence of any of the following clinical and laboratory criteria constitutes severe malaria: Shock, convulsions (two or more), pulmonary oedema, abnormal bleeding, jaundice, haemoglobinuria, acute renal failure (oliguria/anuria), and other symptoms; severe anaemia (haemoglobin (Hb) <5 g/dL or haematocrit (HCT) <15%), hypoglycaemia (blood glucose <2.2 mmol/L) and hyperlactatemia.

6.2.3 Outcomes and factors examined

The length of stay (LOS) was the primary outcome for patients who were admitted to a hospital with suspected malaria. Perceived as a competing event, the time to death during hospitalization was used to model the time to discharge. The patients transferred, referred, or absent patients were removed from the final count. The length of stay (LOS) outcome was assessed against patient age, gender, and ward allocation, documentation of basic assessment tasks performed on admission (weight, temperature) and documentation of vital observations monitored during hospitalization (Table 6.1). The study also assessed for any relationship between confirmed severe malaria diagnosis and positive malaria test results and severity criteria or a health worker's admission diagnosis. Clinical severity parameters were used to help health professionals make better diagnoses of severe malaria so that biases in documentation are avoided.

6.3 Statistical Analysis

6.3.1 Explorative analysis and statistical inference

All of the variables in the study were summarized using exploratory data analysis and descriptive analysis. The medians and interquartile ranges of non-normally distributed variables are summarized (IQRs). The correlations between categorical variables were studied with the help of chi-square testing. As a result, hospital length of stay (LOS) and its associated determinants were studied using the competing risk technique in the survival analysis modeling approach.

The cumulative survival probability drops as the number of people at risk diminishes over time due to a competing event. Risk and cumulative incidence cannot be determined from one model in the context of conflicting risks; consequently, different models are needed to solve etiologic and prognostic epidemiological concerns (Austin, Lee and Fine, 2016; Andersen *et al*.,

2012). Subdistribution and cause-specific models were employed in this study to examine the cumulative chance of being discharged, taking into account the possibility that a patient may die during the hospitalization period.

Competing risk was evaluated using the CIF, which represents the likelihood of experiencing the event of interest before a given period and before the occurrence of any other type of event. The duration of hospitalization for patients admitted with probable malaria was examined using a cause-specific hazard ratio (CSHR) analysis. The SDHR was employed to account for competing hazards while examining the link between LOS and cumulative incidence. At each stage of hospitalization, variables from univariable CSHR and SDHR analyses were tested against the time to event (discharge), and only those determined to be significant (P<0.05) were taken into account in a multivariable model for determining the time to event. The CIs were reported at 95% accurate. P<0.05 was considered statistically significant. In accordance with the RECORD (REporting of Studies Conducted using Observational Routinely-collected Data) Statement. Analysis was performed using StataCorp.14 (Stata Statistical Software: Release 14. College Station, TX, StataCorp LP).

6.3.2 Competing risk modelling

The instantaneous rate at which the event of interest occurs in subjects who are still at risk is described by the hazard function, which is time-dependent (Schuster *et al*., 2020). The hazard function is defined as follows when there are no competing risks:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{Prob(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \qquad (6.1)$$

where T is the time from baseline time until the occurrence of the event of interest.

In the presence of competing risks, the cause-specific hazard function and the subdistribution hazard function are of importance.

Event-free subjects' instantaneous rate of occurrence of the k^{th} event is denoted by the CSHR (the subject is removed from the risk set the moment they experience the competing event or are censored). In the CSHR function (Austin and Fine, 2017), it is defined as:

$$\lambda_k^{cs}(t) = \lim_{\Delta t \to 0} \frac{Prob(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}. \quad (6.2)$$

where *D* is a variable denoting the type of event that occurred and the function

When a subject has not had an event of type *k* yet, the SDHR indicates the risk of failure from the k^{th} event. That is the subjects who were exposed to the competing event are still included in the risk set. In this study, both patients who had been discharged and those who had died of suspected severe malaria were included in the risk group. Subdistribution Hazard Function (Austin and Fine, 2017) was defined by Fine and Gray as follows:

$$\lambda_k^{sd}(t) = \lim_{\Delta t \to 0} \frac{Prob\big(t < T \leq t + \Delta t, D = k \big| T > t \cup (T < t \cap K \neq k)\big)}{\Delta t}. \quad (6.3)$$

Both models account for competing risks by modeling the effect of covariates on different hazard functions. There is a distinct cause-specific hazard function for each of the distinct types of events and a distinct subdistribution hazard function for each of the distinct types of events (Wolbers *et al*., 2014). The SDHR model is considered the right model for prediction research as it allows one to estimate the effect of covariates on the cumulative incidence function for the event of interest (Lambert, 2017) defined as:

$$CIF_k\ (t) = 1 - exp\{\ \hat{\mathrm{H}}_k(t)\}, \quad (6.4)$$

where,

$$\hat{\mathrm{H}}_k(t) = \int_0^t \hat{\mathrm{h}}_k(t)dt \text{ is a cumulative subhazard as } \hat{\mathrm{h}}_k(t) = \lambda_k^{sd}(t). \quad (6.5)$$

While taking competing risk into consideration, the CIF enables for estimation of the incidence of occurance of an event. A single event type can occur in a competing risks context, so that the subsequent appearance of other event types is impossible if one occurs. The following is a definition of the cumulative incidence function for the k^{th} cause:

$$CIF_k\,(t) = \Pr(T \leq t, D = K) \quad (6.6)$$

where, D is a variable denoting the type of event that occurred and the function $CIF_k\,(t)$ denotes the probability of experiencing the k^{th} event before time t and before the occurrence of a different type of event.

6.4 Results

6.4.1 Description of study population

A total of 2,396 patients' medical records from 90 health facilities with suspected malaria were examined (Table 6.1). Of the 2,396 reviewed files, 588 (24.5%) met the inclusion criteria based on the diagnosis of malaria at admission, 2,214 (92.4%) fulfilled the criterion based on the testing for malaria, 1,207 (50.4%) patients were admitted to the pediatric ward, while 1,189 (49.5%) were admitted to adult medical wards. Only 52.4% of the inpatient admissions were for male patients. Patients in the pediatric and medical wards had a median age of three years (IQR: 1–6 years) and 32 years (IQR: 22–37 years) respectively. The median duration of illness from admission was three days, and the median length of admission was four days. The factor estimates and the confidence interval spans between the SDH and CSH models were slightly varied.

Basic assessment tasks such as age (99.6%), weight (48.9%), pulse (69.9%), temperature (81.6%), respiration rate (53.3%), blood pressure (45.0%), and history of fever (87.1%) were assessed on admission by healthcare workers. In the course of hospitalization, vital signs were monitored at highest for temperature (83.8%) and lowest for oxygen saturation (20.2%). In the analyzed files, 33.4% of patients had at least one clinical characteristic of severe malaria, with a slightly greater frequency in the pediatric ward than the medical ward (36.6% vs. 30.0%; P=0.001). Malaria tests were performed on the majority of patients (92.4%), including children (92.1%) and those in the medical wards (92.7%). 25.2% of the patients had their glucose/RBS levels assessed (20.3% children vs. 30.2% adults; P<0.001), while 66.7% of the patients had their Hb or HCT determined.

There were 24.5% of patients with severe malaria on admission to the hospital, and 25.4% had confirmed severe malaria according to the study criteria. In the medical ward, injectable artesunate was prescribed to nearly half of the patients (50.8%), including 53.6% of children and 47.9% of adults (P=0.048). 2283 (95.3%) of the hospitalized patients were discharged, 49 (2.1%) died, 64 (2.6%) were either referred, absent, or discharged against medical advice (Table 6.1).

Table 6. 1 Description of study population, by admission ward

	Paediatric ward (N=1,207)	**Medical ward (N=1,189)**	**All patients (N=2,396)**
	n (%)	**n (%)**	**n (%)**
General information			
Sex (male)	689 (57.4)	561 (47.3)	1,250 (52.4)
Basic assessment performance on admission			

	Paediatric ward (N=1,207)	**Medical ward (N=1,189)**	**All patients (N=2,396)**
	n (%)	**n (%)**	**n (%)**
Age	1,201 (99.5)	1,186 (99.8)	2,387 (99.6)
Weight	922(76.4)	250(21.0)	1,172 (48.9)
Pulse	716 (59.3)	952 (80.1)	1,668 (69.6)
Temperature	1,106 (91.6)	849 (71.4)	1,955 (81.6)
Respiratory rate	682 (56.5)	593 (50.0)	1,275 (53.3)
Blood pressure	74 (6.2)	1,004 (84.4)	1,078 (45.0)
History of fever	1,145 (94.5)	941 (79.1)	2,086 (87.1)
Vital signs monitored during hospitalization			
Temperature	1,129 (93.5)	879 (73.9)	2,008 (83.8)
Respiratory rate	741 (61.4)	652 (54.8)	1,393 (58.1)
Blood pressure	89 (7.4)	1,032 (86.8)	1,121 (46.8)
Pulse rate	765 (63.4)	982 (82.6)	1,747 (72.9)
Oxygen saturation	338 (28.0)	147 (12.4)	485 (20.2)
Documented presence of severe malaria features			
Altered consciousness[a]	125 (10.4)	208 (17.5)	333 (13.9)
Convulsions (2 or more)[b]	152 (12.6)	32 (2.7)	184 (7.7)
Prostration[c]	112 (9.3)	58 (4.9)	170 (7.1)
Severe anaemia[d]	70 (5.8)	33 (2.8)	103 (4.3)
Respiratory distress[e]	64 (5.3)	26 (2.2)	90 (3.8)
Jaundice[f]	38 (3.2)	58 (4.9)	96 (4.0)
Shock[g]	25 (2.1)	24 (2.0)	49 (2.1)
Abnormal bleeding[h]	6 (0.5)	12 (1.0)	18 (0.8)
Renal failure[i]	9 (0.8)	8 (0.7)	17 (0.7)
Haemoglobinuria[j]	5 (0.4)	10 (0.8)	15 (0.6)
Hypoglycaemia[k]	6 (0.5)	4 (0.3)	10 (0.4)
Pulmonary oedema[l]	1 (0.1)	5 (0.4)	6 (0.3)
At least one of the above features of severe	442 (36.6)	357 (30.0)	799 (33.4)

	Paediatric ward (N=1,207)	Medical ward (N=1,189)	All patients (N=2,396)
	n (%)	n (%)	n (%)
malaria			
Laboratory testing practices			
Malaria test done on admission	1,112 (92.1)	1,102 (92.7)	2,214 (92.4)
Malaria test positive	587 (52.8)	560 (50.8)	1,147 (51.8)
Malaria test done post admission	77 (6.4)	67 (5.6)	144 (6.0)
Haemoglobin (Hb) or Heamatocrit (HCT) done	830 (68.8)	767 (64.5)	1,597 (66.7)
Glucose/Random Blood Sugar (RBS) test done	245 (20.3)	359 (30.2)	604 (25.2)
Malaria diagnosis			
Clinicians' diagnosis of severe malaria	322 (26.7)	266 (22.4)	588 (24.5)
Confirmed severe malaria[m]	347 (28.8)	262 (22.0)	609 (25.4)
Treatment during hospitalization			
Artesunate injection prescribed	647 (53.6)	570 (47.9)	1,217 (50.8)
Malaria outcome			
Discharged	1,167 (96.7)	1,116 (93.9)	2,283 (95.3)
Died	12 (1.0)	37 (3.1)	49 (2.1)
Absconded	1 (0.1)	4 (0.3)	5 (0.2)
Referred	26 (2.2)	30 (2.5)	56 (2.3)
Discharged against medical advice	1 (0.1)	2 (0.2)	3 (0.1)

[a] Documented "drowsiness, lethargy, confusion, unconsciousness, coma or GCS <15"/AVPU<A"; [b] Documented "convulsions, fits or seizures"; [c] Documented "unable to drink/breastfeed/sit/stand/walk, or prostrated"; [d] Documented "Hb <5 g/dL or HCT<15%". [e] Documented "acidotic/deep breathing, chest in-drawing, or respiratory distress"; [f] Documented "jaundice"; [g] Documented "capillary refill ≥3 sec, systolic BP<80 mmHg in adults/<70 mmHg in children or shock"; [h] Documented "bleeding"; [I] Documented "oliguria, anuria, reduced urine output, or renal failure"; [j] Documented "dark urine, blood in urine, haematuria"; [k] Documented "blood sugar <2.2 mmol"; [l] Documented "pulmonary oedema"; [m] Positive malaria test and severity criteria defined as documentation of at least one severe malaria feature or diagnosis of severe malaria made on admission by a health worker.

6.4.2 Multivariate factors associated with hospital LOS

The median length of stay was four days (IQR: 3-6 days; range: 1-46 days) among hospitalized patients. Cumulatively 82.6% of patients were discharged by the sixth day from the hospital,

indicating a right-skewed distribution of discharge days (Figure 6.1). A total of 2,332 patients were included in the final model, of which 2,283 were discharged alive and 49 died. However, 64 were censored because they had absconded, referred, or discharged against medical recommendation. In Table 6.2, the univariate analysis findings for the CSHR and SDHR are shown, identifying possible factors that may be related with hospital LOS. A multivariable model was used to examine the effects of significant covariates (P<0.05) identified in univariate analyses. In the univariable analysis, there was no connection between the outcomes studied and patient age, weight, sex, or ward allocation. Survival curves were estimated and cumulative incidence function curves to compare the risks illustrated by artesunate treatment for severe malaria.

Table 6. 2 Univariable analysis of factors associated with hospital length of stay

	Whole Sample N=2,332	**Cause-specific Hazard (Rate of Discharge)**		**Subdistribution-Hazard (Association With Cumulative Incidence of discharge)**	
Factors	**n(%)**	**CSHR(CI)**	***P* value**	**SDHR(CI)**	***P* value**
General information					
Age category (more 5 years)	1,525(65.7)	1.048(0.968; 1.135)	0.244	1.018(0.948; 1.094)	0.625
Age taken	2,323(99.6)	0.772(0.512; 1.163)	0.213	0.756(0.507; 1.126)	0.168
Sex (male)	1,211(52.1)	0.980(0.914; 1.051)	0.564	0.952(0.886; 1.022)	0.177
Ward (paediatric)	1,179(50.6)	0.992(0.918; 1.073)	0.843	1.039(0.967; 1.116)	0.302
Assessment on admission					
Weight	1,146(49.1)	1.023(0.925; 1.131)	0.657	1.063(0.967; 1.169)	0.206
Pulse	1,618(69.4)	0.845(0.752; 0.950)	**0.005**	0.869(0.777; 0.972)	**0.014**
Temperature	1,900(81.5)	0.816(0.722; 0.921)	**0.001**	0.905(0.796; 1.030)	0.130

	Whole Sample N=2,332	Cause-specific Hazard (Rate of Discharge)		Subdistribution-Hazard (Association With Cumulative Incidence of discharge)	
Factors	**n(%)**	**CSHR(CI)**	***P* value**	**SDHR(CI)**	***P* value**
Respiratory rate	1,232(52,9)	0.806(0.713; 0.911)	**0.001**	0.848(0.759; 0.946)	**0.003**
Blood pressure	1,046(44.9)	0.951(0.875; 1.034)	0.234	0.936(0.865; 1.013)	0.102
Fever complaint	2,030(87.1)	0.955(0.817; 1.117)	0.562	0.995(0.858; 1.154)	0.948
Monitoring during hospitalization					
Temperature	1,953(83.7)	0.791(0.687; 0.910)	**0.001**	0.862(0.742; 1.002)	0.053
Respiratory rate	1,349(57.8)	0.828(0.727; 0.943)	**0.005**	0.874(0.775; 0.985)	**0.027**
Blood pressure	1,089(46.7)	0.958(0.885; 1.037)	0.280	0.935(0.867; 1.007)	0.078
Pulse rate	1,696(72.7)	0.860(0.759; 0.973)	**0.017**	0.886(0.782; 1.004)	0.058
Oxygen saturation	469(20.1)	0.820(0.718; 0.938)	**0.004**	0.837(0.733; 0.955)	**0.008**
Laboratory testing					
Malaria test done on admission	2,155(92.4)	1.046(0.880; 1.243)	0.607	1.120(0.926; 1.354)	0.242
Hb/ HCT done	1,551(66.5)	0.708(0.650; 0.771)	**<0.001**	0.738(0.678; 0.803)	**<0.001**
Glucose/RBS test done	582(25.0)	0.769(0.695; 0.850)	**<0.001**	0.733(0.673; 0.799)	**<0.001**
Clinical features					
At least one feature of severe malaria[a]	762(32.7)	0.747(0.679;0.821)	**<0.001**	0.696(0.626;0.774)	**<0.001**
Diagnosis					
HW's severe malaria diagnosis on admission	571(24.5)	1.063(0.953;1.186)	0.267	1.054(0.937;1.185)	0.380
Confirmed severe malaria [b]	592(25.4)	1.219(1.090;1.362)	**0.001**	1.214(1.082;1.362)	**0.001**
Treatment during the					

	Whole Sample N=2,332	Cause-specific Hazard (Rate of Discharge)		Subdistribution-Hazard (Association With Cumulative Incidence of discharge)	
Factors	**n(%)**	**CSHR(CI)**	***P* value**	**SDHR(CI)**	***P* value**
hospitalization					
Artesunate injection prescribed	1,186(50.9)	1.365(1.206; 1.545)	**<0.001**	1.339(1.184; 1.514)	**<0.001**

The bold values are those that are significant results (P <0.05)

[a] Documentation of at least one of the clinical and laboratory features as specified and defined in Table 1.

[b] Defined as positive malaria test on admission and presence of severity criteria (either documentation of any clinical features of severe malaria or severe malaria diagnosis made by clinician); Clinical severity criteria were complemented with health workers diagnosis of severe malaria to protect correctness of severity classification from documentation biases.

The multivariable model (Table 6.3) revealed a significant (P<0.05) decrease in the discharge rate when temperature was measured on admission and during hospitalisation (by 10.9% and 13.3%, respectively), when the respiratory rate was assessed on admission (by 14.4%), when oxygen saturation was monitored during hospitalisation (by 13.1%), when Hb/HCT and glucose/RBS levels were measured (by 26.8% and 19.2%**,** respectively), and by 25.3% if patients had documentation of at least one clinical feature of severe malaria. Conversely, the adjusted discharge rate increased by 21.9% when patients presented with confirmed severe malaria and by 36.5% when patients were treated with injectable artesunate. With respect to the cumulative incidence of discharge, the multivariable model showed that assessment of respiratory rate decreased incidence by 12.7%, monitoring of oxygen saturation decreased it by 14.1%, and performance of Hb/HCT and glucose/RBS blood tests by 23.1% and 23.4%, respectively; if patients had documentation of at least one clinical feature of severe malaria, the cumulative incidence of discharge decreased by 30.4%. In contrast, patients with confirmed severe malaria and those treated with injectable artesunate had an increased cumulative incidence of discharge by 21.4% and 33.9%, respectively.

Table 6. 3 Multivariable analysis of factors associated with hospital length of stay

	Whole Sample N=2,332	**Cause-specific Hazard (Rate of Discharge)**		**Subdistribution-Hazard (Association with Cumulative Incidence of discharge)**	
Factors	**n (%)**	**Adjusted CSHR (CI)**	***P* value**	**Adjusted SDHR (CI)**	***P* value**
Assessment on admission					
Pulse	1,618 (69.4)	0.933 (0.848; 1.026)	0.152	0.932 (0.843; 1.031)	0.173
Temperature	1,900 (81.5)	0.891 (0.798; 0.994)	**0.039**		
Respiratory rate	1,232 (52,9)	0.856 (0.763; 0.959)	**0.008**	0.873 (0.789; 0.967)	**0.009**
Monitoring during hospitalization					
Temperature	1,953 (83.7)	0.867 (0.764; 0.984)	**0.028**		
Respiratory rate	1,349 (57.8)	0.896 (0.793; 1.013)	0.078	0.895 (0.795; 1.007)	0.064
Pulse rate	1,696 (72.7)	0.956 (0.862; 1.060)	0.390		
Oxygen saturation	469 (20.1)	0.869 (0.758; 0.998)	**0.046**	0.859 (0.754; 0.978)	**0.022**
Laboratory testing					
Hb/HCT done	1,551 (66.5)	0.732 (0.675; 0.794)	<**0.001**	0.769 (0.709; 0.833)	<**0.001**
Glucose/RBS test done	582 (25.0)	0.808 (0.733; 0.891)	<**0.001**	0.766 (0.704; 0.833)	<**0.001**
Clinical features					
At least one feature of severe malaria[a]	762 (32.7)	0.747 (0.679; 0.821)	<**0.001**	0.696 (0.626; 0.774)	<**0.001**
Diagnosis					
Confirmed severe malaria[b]	592 (25.4)	1.219 (1.090; 1.362)	**0.001**	1.214 (1.082; 1.362)	**0.001**
Treatment					
Artesunate injection treatment	1,186 (50.9)	1.365 (1.206; 1.545)	<**0.001**	1.339 (1.184; 1.515)	<**0.001**

Significant results ($P < 0.05$) are indicated by values in bold. [a]Documentation of at least one of the clinical and laboratory features as specified and defined in Table 1. [b]Defined as positive malaria test on admission and presence of severity criteria (either documentation of any clinical features of severe malaria or severe malaria diagnosis made by clinicians); clinical severity criteria were complemented with health workers' diagnosis of severe malaria to protect the correctness of severity classification from documentation biases.

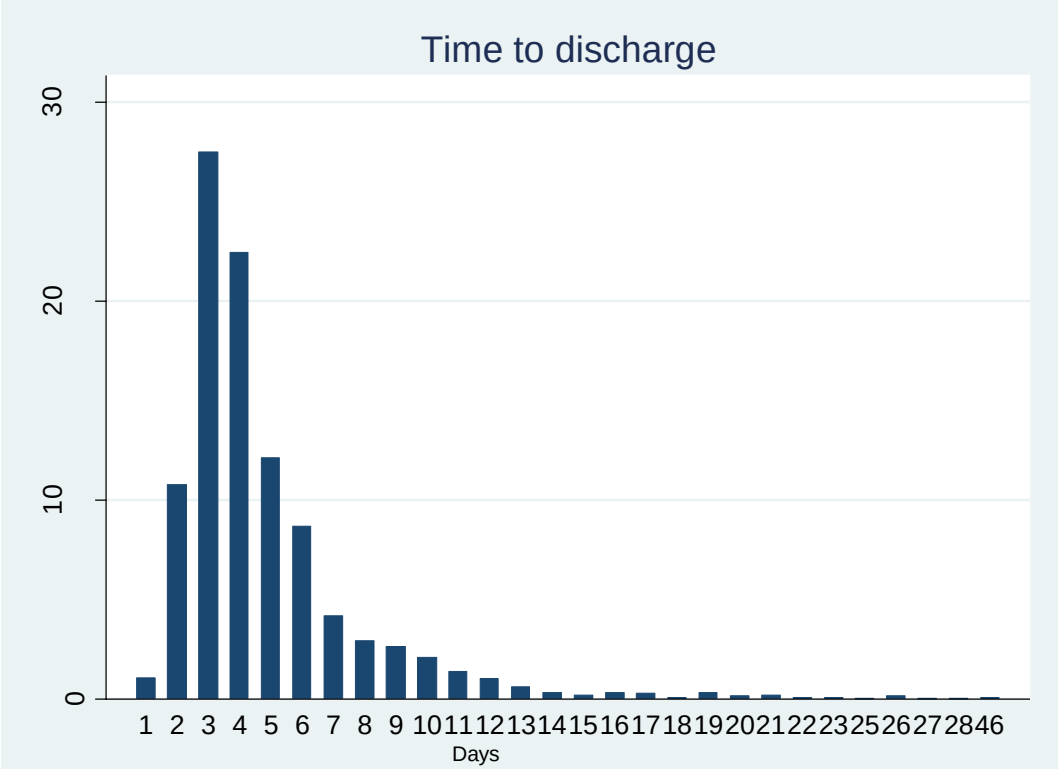


Figure 6. 1 Distribution of time to discharge for patients admitted with suspected malaria

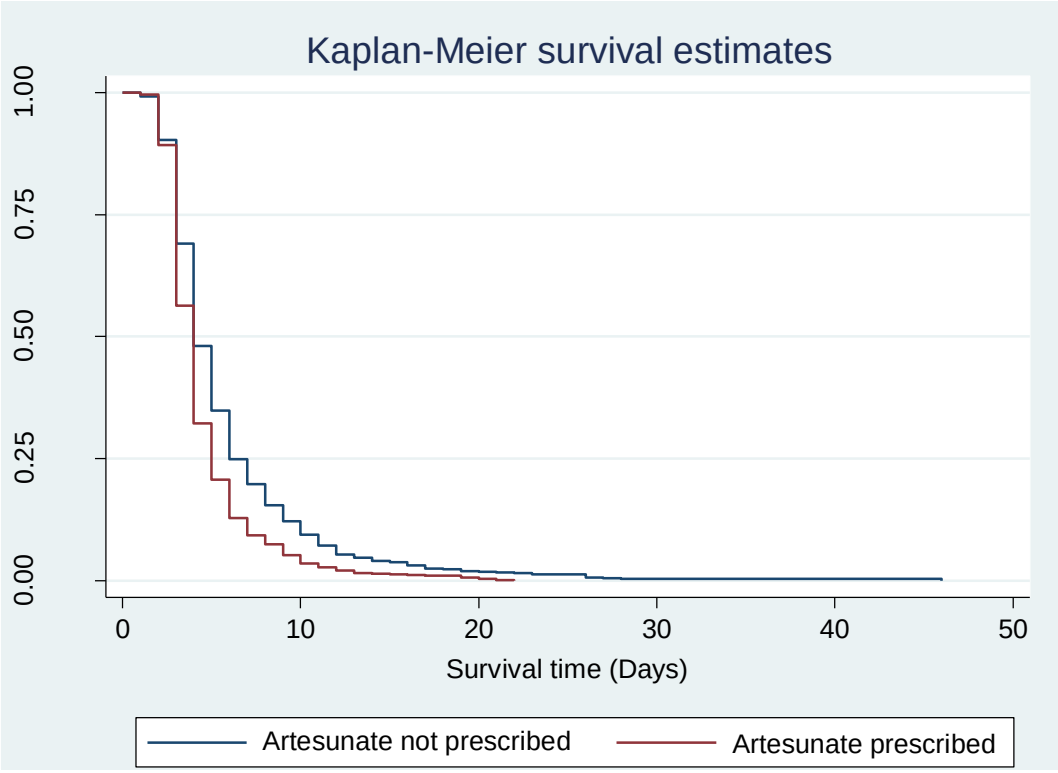


Figure 6. 2 Kaplan Meier Survival estimates

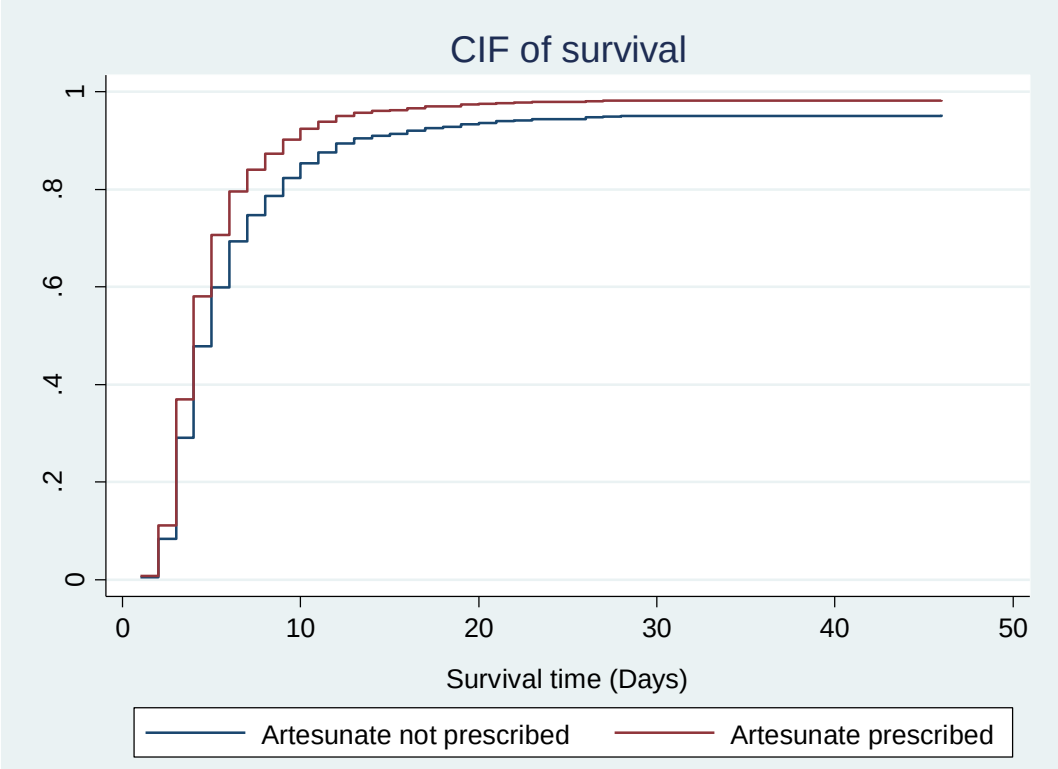


Figure 6. 3 Cumulative Incidence Function (CIF) of survival

6.5 Discussion

The median length of stay (LOS) for patients admitted to a Kenyan hospital with suspected malaria was four days, according to the findings of this study. When we looked at our data, we discovered a much lower length of stay than the seven days reported under trial settings in South East Asia (Keene *et al*., 2018) and slightly longer than the three-day LOS reported in an observational study from Germany's largest tertiary care hospital (Hoffmeister, 2021). We hypothesize that the LOS disparities between studies may be a result of different malaria populations studied, which range from a focus on severe malaria patients in Southeast Asia (Keene *et al*., 2018) to barely 10% of the malaria cases in Germany (Hoffmeister, 2021).

LOS for suspected malaria patients was found to be influenced by a number of factors, which can be used to improve normal hospital care in Kenya and to add to the growing body of information about malaria LOS worldwide. There was a clear correlation between longer LOS and sicker patients (those with at least one symptom of severe malaria), as previously documented in other malaria investigations (Keene *et al*., 2018; Hoffmeister, 2021). We did not investigate the relationship between the severity of malaria and specific features because of the prevalence of non-documentation of clinical signs and symptoms in routine information systems in African (Gathara *et al*., 2015; Aluvaala *et al*., 2015). However, we used a cumulative measure of malaria severity since most patient files have at least one documented sign of severity, and only one sign is necessary to classify a case as severe (Mace *et al*., 2014). Taking into account the study results and the widespread reports of substandard inpatient treatment, it is clear that there is a need for improvement across the continent (Zurovac *et al*., 2018; Amboko *et al*., 2016; Achan *et al*., 2011; Sears *et al*., 2015; Shah *et al*., 2016; Ojo *et al*., 2020). Healthcare implementers should focus on enhancing clinical practices for the early detection of severity signals and the proper management of well-established sets of severe malaria sequelae (MoH, 2015b; WHO, 2000).

Structured vital sign charts allowed us to investigate the relationship between LOS and the enactment of these tasks as quality-of-care indicators. Study participants were found to have shorter lengths of stay if vital signs (temperature, respiration rate, oxygen saturation) were measured. The low performance of vital sign monitoring, especially respiration counts and oxygen saturation in this study, as well as generally suboptimal performance of nursing care in this domain in Kenya (Ogero *et al*., 2018), is an area, requiring targeted interventionsfor all admitted patients, and not just those with malaria. In addition, we discovered that shorter LOS

was related to performance of laboratory test results, such as Hb/HCT and RBS values, in particular. Anaemia and hypoglycemia are common malaria complications that can be detected early with the help of a variety of laboratory tests, but they require significant quality improvement efforts in comparison to the widespread use of malaria testing in Kenya (Amboko *et al*., 2022). The widespread availability of laboratory services for anaemia and blood sugar testing within hospitals in Kenya is more behavioural than the availability of testing shortcomings. LOS was also significantly associated with confirmed severe malaria and treatment with injectable artesunate. While this pattern is intuitively expected for confirmed severe malaria and, together with previously shown association with severity features, simply shows that sicker patients require longer recovery, the association with artesunate treatment practice is less clear, though not previously unobserved (Keene *et al*., 2018). Malaria testing and treatment, on the other hand, is based on optimizing early recognition of severity indicators, quick malaria testing, and parenteral administration of artesunate in cases of severe malaria that test positive (MoH, 2015b; WHO, 2000).

Statistically, competing event (death) had an effect on the hazard of discharge based on estimates from CSH and SDH models (Austin, Lee and Fine, 2016; Fine and Gray, 1999; Geskus, 2011). The results showed that the factor estimates and the confidence interval spans between the SDH and CSH models were slightly varied. Thus, death had an impact on estimating the relevant event (discharge from hospital), according to these findings (Keene *et al*., 2018; Austin, Lee and Fine, 2016; Twabi and Mukaka, 2018). According to the findings, ignoring competing risks and applying standard survival models to data that includes competing events leads to biased estimates consequently biased conclusion. The SDH model is better at finding

prognostic factors when there are competing risks than the CSH model (Schuster *et al*., 2020; Gathara *et al*., 2015).

The study provides a national representation of hospitals and analyses of a large dataset of admissions under routine, real-world conditions of inpatient service delivery. There are certain limitations to the findings. The findings are limited to county referral and major faith-based hospitals and cannot be inferred for smaller health facilities where inpatient malaria care is also provided. The study did not account for severe malarial comorbidities that might influence LOS. Data extraction from routine hospital records is commonly subject to documentation biases, which prompted us to limit modelling to the set of basic clinical predictors that are routinely recorded in admission files such as age, sex, vital signs, testing, diagnosis, and treatment.

6.6 Conclusion

According to our research, the average LOS in Kenya for patients admitted with suspected malaria during regular hospitalization was four days. A competing risk approach was used to identify the seven inpatient clinical process characteristics that influence LOS and can be explicitly targeted during quality improvement interventions to improve service delivery in Kenyan health care facilities. Early recognition and appropriate management of the signs of malaria severity may have the greatest effect on beneficial outcomes. Strengthening clinical practices and nursing care according to national case management guidelines should be a priority for malaria control managers in Kenya.

CHAPTER 7

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

This last chapter provides a general discussion on statistical methods for correlated data (Section 7.1) and conclusion (Section 7.2) by study objective. It further highlights the study's strengths and limitations (Section 7.3) and ends with recommendations by study objective in Section 7.4.

7.1 Discussion

7.1.1 Investigate the impact of correlation on dichotomous or polytomous outcomes in terms of estimation of parameters, precision measures and prediction, adjusting for health facility and county structures

Analysing correlated data has been a statistical bother both in implementation and interpretation. Multiple statistical methods outlined in the literature review (Section 2.4) were utilized to examine correlated data. Multilevel mixed-effect ordinal and binary logistic regression was applied (Section 4.1). The multilevel models (Section 2.4.1.5) are aimed at analysing variables from multiple levels at the same time while taking intra-cluster correlation into account (Austin and Merlo, 2017; Zyzanski *et al*., 2004). The random effects in the model indicated the amount of variability within and between HFs and counties in the outcome. Considering intracluster correlation (ICC) in robust standard error estimations led to accurate inferences (Cameron & Miller 2015; Greenland *et al*., 2016).

In multilevel systems, people in similar group, like hospitals, may have observations that are correlated. This is related to the sharing same locality based on the outcome measure stated as clustering (Sainani, 2010). For example, health workers from a particular ward are likely to provide comparable answers than health workers randomly selected from diverse wards. In the

presence of clustering observations within a cluster, similar data produces an intra-cluster correlation between observations within a cluster. Applications of statistical approaches that overlook clustered data yield erroneous inferences. It is crucial to do suitable statistical studies that account for clustering (Schmidt-Catran and Fairbrother, 2016; Kanters, 2022). Multilevel modelling contained effects at all levels, taking care of hierarchical patterns in the data. The concept of the variance component model is based on the theory that parameters vary from cluster to cluster, revealing the natural heterogeneity induced by unmeasured causes (Liang and Zeger, 1993; Hedeker, Demirtas, and Mermelstein, 2009).

Statistical methodologies for correlated data must account for the inter-subject correlation of response measurements in order to get accurate results. Analyses were conducted using two statistical approaches. In the first model, correlation was ignored while in the second model clustering was considered and the two models were compared. The discrepancies in parameter estimation variability and precision between the two models were examined. The model that adjusted for structure design had a wide range of coefficients, robust standard errors that were greater than ordinary standard errors, and odds ratio confidence intervals that were slightly wider than usual. When intracluster correlation is considered in robust standard error estimations, the results are accurate (Liu, 1998). This is because data is analyzed at a single level while taking into consideration variance at other levels due to multilevel modeling. Accurate standard errors could further be achieved when using maximum likelihood algorithms in multilevel analyses that take into account several error factors (Kanters, 2022; Hox *et al*., 1998).

Some of the variables were no longer significant when the data was re-analyzed after accounting for clustering at the hospital and county levels of hierarchy, because the models ignored the possibility of homogeneity within clusters or the influence a cluster can exert on the

outcome. Cluster-level effects and item-level effects must be clearly understood, as well as the correlation between items in a single cluster. Analysis of data that is cluster-correlated results in skewed estimates of standard errors, which leads to inaccurate test statistics and confidence intervals, which in turn lead to incorrect conclusions (Desai and Begg, 2008). When analyzing data that is grouped together, methodologies that account for both individual and contextual effects should be considered.

Hierarchical regression models allow us to draw better conclusions from the data, with more precise estimations of the statistical outputs. Standard errors and confidence intervals calculated in classic ordinal logistic regression are influenced by the presence of clustering effects, leading to inaccurate interpretations of the relationships between variables (Cameron & Miller, 2015; Sainani, 2010; Greenland *et al*., 2016).

7.1.2 Bayesian hierarchical approach to analyze polytomous data, adjusting for county structures

In public health studies, clustered, hierarchical, non-independent, and geographical data are common. As a result of many levels of clustering, such as hospitals or wards, these data are frequently associated, and statistical analysis must account for this association. As detailed in the literature review in Section 2.4.2.8, Bayesian hierarchical techniques can be utilized to analyze correlated data. Cluster adjustment or correlation matrices that do not take into account geographical relatedness can lead to inaccurate predictions in the design of severe malaria health facility surveys. Based on the premise that locations nearer each other in space tend to have related outcomes, correlations between spatial units can be found in spatial application. In spatial analysis, autocorrelation must be tolerated in order to expose the sources of variation underlying spatial patterns for targeted interventions. Spatially structured random effects illustrate the

necessity of accounting for spatial autocorrelation among counties for accurate inferences (Dasgupta *et al*., 2014; Hanandita and Tampubolon, 2016).

Bayesian approach was implemented taking into account the individual and contextual characteristics, knowledge levels and spatial relatedness at the county level. The spatial heterogeneity of counties was described using Bayesian hierarchical models, and the conditional autoregressive model was used to model the data's spatial autocorrelation. Different spatial architectures of the model were examined using conditional autoregressive specifications and the best model was chosen. The similarity of health workers' responses in the same facility and the likelihood of similarity between health facility structures in neighbouring counties explain the neighbourhood influence on the spatially structured models.

Three models were utilized in the analysis: modelling with structured random effects, unstructured spatially random effects, and convolution. All three models used spatially random effects that were mapped. Using hierarchical disease-mapping (CAR) methods, the researcher was able to take into account the heterogeneity of the data and make risk maps that showed where health professionals needed to learn more about the policy, dose, and preparation for treating severe malaria (Coly *et al*., 2021). The Bayesian hierarchical model was used to take into account how health facilities and counties tend to group together and how healthcare workers' knowledge differ from one another (Hanandita and Tampubolon, 2016). Using the spatial maps, multidisciplinary interventions can be targeted at bridging knowledge gaps at county level.

7.1.3 Factors associated with hospital LOS among severe malaria patients using competing risk approach, adjusting for health facility structures

The LOS in the hospital for patients with severe malaria relies on the care they receive in the first 24 hours, when the danger of death is highest (MoH, 2015b). Patient outcomes after treatment for malaria are influenced by factors such as the patient's admission, diagnosis, and treatment as well as hospital length of stay (LOS). As a competing risk, death can interrupt the discharge outcome (Pintilie, 2011). One of the most significant aspects of healthcare delivery is predicting the LOS for patients. In resource-limited settings, optimizing service delivery will be made possible by assessing and changing parameters controlling the time malaria patients stay in the hospitals despite competing risk events (Keene *et al*., 2018). According to the findings, the average length of stay was four days. Patients who were assessed for respiratory rate and oxygen saturation at admission, Hb/HCT, glucose/RBS levels during hospitalization, and the presence of any one clinical characteristic were less likely to be released from the hospital. Patients hospitalised with suspected malaria were more likely to be discharged when their diagnosis of severe malaria was confirmed using the criteria and injectable artesunate was prescribed.

The CSH and SDH model estimations quantified the effect of factors on the hazard of discharge in the presence of a competing event (death) (Keene *et al*., 2018). The hierarchical likelihood estimation method was utilized to fit the models and draw inferences from clustered data (Ha *et al*., 2016; Austin, Lee, and Fine, 2016). The SDH model outperforms the CSH model in identifying prognostic factors when there is a competing risk (Taylor, 2015). Similar studies indicated that the SDH model explained the influence of variables better when there was a competing event (Twabi and Mukaka, 2018). To illustrate the impact of ignoring the competing risk using conventional survival models, Kaplan-Meier (KM) survival function was fitted and

compared it with the cumulative incidence function using artesunate treatment for severe malaria and its influence on hospital LOS. The CIF is superior to the KM survival function for calculating the crude incidence of events. The KM curve attempts to answer the chance of surviving before seven days. It fails to recognise that the patient can also die before the seven days are up, and the probability of survival is not equal to one. The results showed that administering injectable artesunate boosted survival in severe malaria patients while also increasing their LOS in the hospital. Previous research found that treating severe malaria patients with artesunate instead of quinine reduced mortality (Dondorp *et al*., 2010; Keene *et al*., 2018).

7.2 Conclusions

7.2.1 Investigate the impact of correlation on dichotomous or polytomous outcomes in terms of estimation of parameters, precision measures and prediction, adjusting for health facility and county structures

Multilevel mixed effect modelling addressed the issue of correlated data by having both fixed and random effects in the model. Based on multilevel statistical analysis, the coefficients were slightly varied and the standard errors of the coefficients and CI for the odds ratio in analysis done while adjusting for structures were slightly wider compared to the analysis without adjustment to structures. Ignoring the presence of clustering effect led to biased estimation of regression coefficients while, ordinary standard errors resulted to narrower CI and smaller *P* values leading to erroneous statistical inferences. Using fixed effects model in unadjusted analysis, it is difficult to isolate the effect of covariates at the group level. The random effects included in the adjusted model indicated the amount of variability within and between health facilities or counties in the outcome. Thus, adjusting for these structures of the hierarchy that considered robust standard errors estimations resulted in accurate inferences.

7.2.2 Bayesian hierarchical approach to analyze polytomous data, adjusting for county structures

Based on the DIC, Model 1 with spatially structured random effects provided a better fit for the three outcomes illustrating the necessity of accounting for spatial autocorrelation among counties inorder to draw accurate inferences. Spatial correlation was addressed using conditional autoregressive (CAR) prior assigned to the random effects at the county level. The similarity of health workers' responses in the same facility and the likelihood of similarity between health facility structures in neighbouring counties explain the neighbourhood influence on the spatially structured models. Bayesian hierarchical methods modelled the spatial heterogeneity among counties while, the CAR model, analyzed the spatial autocorrelation of data based on neighborhood relationships. The spatial maps identified knowledge gaps at subnational levels that can be targeted to bridge the gaps.

7.2.3 Factors associated with hospital LOS among severe malaria patients using competing risk approach, adjusting for health facility structures

The median hospital LOS for patients admitted with suspected malaria was 4 days. Multivariable analysis of factors associated with hospital length of stay adjusting for health facility structures was applied using conventional Cox regression model to obtain a cause-specific hazard ratio and, Fine and Gray competing risks method to obtain a subdistribution hazard ratio. The factor estimates and the confidence interval spans between the SDH and CSH models were slightly varied. The competing event affected the estimation of the factors of the event of interest. Ignoring competing risks and applying standard survival models to data leads to biased estimates and conclusions.

7.3 Strengths and limitations

As a representative sample of the nation's hospital systems, the study examined a huge dataset of inpatient admissions under typical, real-world conditions allowing for generalization of results. This study also shed light on the knowledge and factors impacting subnational health workers' knowledge of targeted treatments. Multiple sources of data could be merged in a systematic fashion using Bayesian modeling to arrive at accurate conclusions. Understanding of the new treatment policy is an essential prerequisite for the implementation of drug policy. There are, however, some drawbacks to be aware of. First, the results apply to major government and FBO hospitals nationwide. However, smaller facilities with lower inpatient capacities are not included in this analysis. This finding can be applied to other health institutions with similar settings and levels of care because the sample design was representative of the study population. Second, several exploratory analyses and comparisons may have resulted in some results being significant. Third, severe malaria comorbidity that could affect LOS was not taken into account in the study. Fourth, there is a possibility that data cleaned from routine hospital records is skewed due to incomplete or inaccurate documentation of the criteria for identifying severe malaria. To make up for these problems, clinical severity criteria were added to the health professionals' admission diagnosis of severe malaria and the clinical features written in the patients' files.

7.4 Recommendations

7.4.1 Investigate the impact of correlation on dichotomous and polytomous outcomes in terms of estimation of parameters, precision measures (SE & CI), and prediction, adjusting for health facility and county structures in Kenya

Investigating the impact of correlation, multilevel mixed effect logistic regression modelling was used to examine the predictors of health workers' knowledge about artesunate-based severe malaria treatment recommendations. During the predictor analysis the data was analyzed while ignoring correlation and subsequently, analyzed adjusting for clustering at county and health facility levels. Then, the variance within the statistical estimates was examined and found to be slightly varied. Multilevel modelling allowed data to be analyzed at one level while accounting for variance at other levels resulting in more accurate estimates. This model accounted for the hierarchical structure of the data while, the mixed effect ordinal logistic regression identified factors associated with health workers' level of knowledge. The model is recommended in analyzing similar data to address the correlated nature of data.

Based on prediction of knowledge, key influential interventions like simple job aids like displayed poster may enhance severe malaria case management. It is recommended that health facilities exhibit appropriate posters on a regular basis to remind health workers about artesunate preparation, dosage, dosing interval, and preferred route of administration. These posters are aligned to the national malaria treatment guidelines that provide detailed information on malaria case management. FBOs should put a lot of effort into making artesunate available in the health facilities besides capacity build their nurses' skills. In both sectors, future interventions should pay greater attention to the health workers managing adult patients in the medical wards

7.4.2 Bayesian hierarchical approach to analyse dichotomous or polytomous data, adjusting for county structures in Kenya

To address this objective, Bayesian hierarchical ecological spatial model beyond predictor analysis was applied to test for the best fitting model to predict subnational artesunate knowledge levels across the 47 counties. Expending Bayesian framework, ordinal logistic regression modelling adjusted for clustering at the county was fitted using three hierarchical models; spatially structured random effects, unstructured spatially random effects and convolution. Spatially structured random effects offered the best fitting model. Overdispersion and geographical dependence were addressed by combining individual, contextual, and random variables into a single model. As a result of this strategy, countries with significant knowledge gaps could be pinpointed for further study and action. The Bayesian approach offered another tool to examine aspects related to health workers' knowledge in correlated data. This methodology can be used to analyze similar data types and contexts in health surveys.

There is a need for interprofessional collaboration and innovative measures to bridge the knowledge gap difference between the nurses and clinicians. The spatial analyses highlighted sources of heterogeneity underlying spatial patterns for focused interventions. Spatial maps have shown that there are gaps in subnational knowledge for focussed interventions.

7.4.3 Factors associated with the LOS among severe malaria patients using a competing risk survival analysis approach, adjusting for health facility structures in Kenya

In relation to quality of care practices, competing risk approach was used to identify the factors associated with LOS for patients admitted with suspected severe malaria based on usual clinical setting. Cumulative incidence function approach, sub-distribution (SDH) model and cause-specific (CSH) model were used to evaluate the effects covariates have on the cumulative

probability of being discharged taking into account that a patient can die during the hospitalization period. The CSH model analysis explored the factors related to the duration of hospitalization for patients admitted with suspected severe malaria. While, SDH model examined the association with cumulative incidence accounting for competing risk in the presence of correlation. The SDH model can be used to analyse correlated malaria survey data and other similar research problems.

The identified factors shortening hospital LOS like the basic vital observations (temperature, respiration and oxygen saturation), simple laboratory tests (haemoglobin and glucose levels) for suspected malaria patients can be emphasized during quality improvement interventions to improve health service delivery.

7.4.4 Implication for future research

Based on this study design methodologies, generalizability of findings is assured in similar research settings. In addition, further study is recommended to establish the effectiveness of some of the operational programmatic interventions employed by the program to improve malaria case management. Also, determining the challenges associated with severe malaria case management, and assessing the effect of comorbidity on LOS for severe malaria patients could be beneficial. Lastly, spatial analysis on malaria prevalence in Kenya can pinpoint the hotspots for focused interventions.

REFERENCES

Achan, J., Tibenderana, J., Kyabayinze, D., Mawejje, H., Mugizi, R., Mpeka, B., Talisuna, A. and D'Alessandro, U., 2011. Case management of severe malaria-a forgotten practice: experiences from health facilities in Uganda. *PLoS One*, *6*(3), p.e17053.

Achia, T.N., 2014. Spatial modelling and mapping of female genital mutilation in Kenya. *BMC public health*, *14*(1), pp.1-14.

Adebola, O., Babatunde, O. and Bose, O., 2014. Hypoxemia predicts death from severe falciparum malaria among children under 5 years of age in Nigeria: the need for pulse oximetry in case management. *African health sciences*, *14*(2), pp.397-407.

Agresti, A., 2003. *Categorical data analysis* (Vol. 482). John Wiley & Sons.

Agresti, A., 2007. An Introduction to Categorical Data Analysis Second Edition Second edi., Hoboken, New Jersey, Canda: JohnWiley & Sons. *Inc., Hoboken, New Je*.

Akech, S., Chepkirui, M., Ogero, M., Agweyu, A., Irimu, G., English, M. and Snow, R.W., 2020. The clinical profile of severe pediatric malaria in an area targeted for routine RTS, S/AS01 malaria vaccination in Western Kenya. *Clinical Infectious Diseases*, 71(2), pp.372-380.

Al Farsi, F., Chandwani, J., Mahdi, A.S. and Petersen, E., 2019. Severe imported malaria in an intensive care unit: A case series. *IDCases*, *17*, p.e00544.

Allison, P.D., 2012. *Logistic regression using SAS: Theory and application*. SAS institute.

Aluvaala, J., Nyamai, R., Were, F., Wasunna, A., Kosgei, R., Karumbi, J., Gathara, D. and English, M., 2015. Assessment of neonatal care in clinical training facilities in Kenya. *Archives of disease in childhood*, *100*(1), pp.42-47.

Amboko, B.I., Ayieko, P., Ogero, M., Julius, T., Irimu, G. and English, M., 2016. Malaria investigation and treatment of children admitted to county hospitals in western Kenya. *Malaria journal*, *15*(1), pp.1-9.

Amboko, B., Stepniewska, K., Macharia, P.M., Machini, B., Bejon, P., Snow, R.W. and Zurovac, D., 2020. Trends in health workers' compliance with outpatient malaria case-management guidelines across malaria epidemiological zones in Kenya, 2010-2016. *Malaria Journal*, *19*(1), pp.1-14.

Amin, A.A., Zurovac, D., Kangwana, B.B., Greenfield, J., Otieno, D.N., Akhwale, W.S. and Snow, R.W., 2007. The challenges of changing national malaria drug policy to artemisinin-based combinations in Kenya. *Malaria journal*, *6*(1), pp.1-11.

Amorim, L.D. and Cai, J., 2015. Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology*, *44*(1), pp.324-333.

Ampadu, H.H., Asante, K.P., Bosomprah, S., Akakpo, S., Hugo, P., Gardarsdottir, H., Leufkens, H.G., Kajungu, D. and Dodoo, A.N., 2019. Prescribing patterns and compliance with World Health Organization recommendations for the management of severe malaria: a modified cohort event monitoring study in public health facilities in Ghana and Uganda. *Malaria journal*, *18*(1), pp.1-8.

Andersen, P.K., Geskus, R.B., De Witte, T. and Putter, H., 2012. Competing risks in epidemiology: possibilities and pitfalls. *International journal of epidemiology*, *41*(3), pp.861-870.

Andersen, P.K., Klein, J.P. and Rosthøj, S., 2003. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, *90*(1), pp.15-27.

Andersen, P.K. and Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pp.1100-1120.

Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, *27*(3), pp.247-267.

Arfan, M. and Sherwani, R.A.K., 2017. Ordinal logit and multilevel ordinal logit models: an application on wealth index MICS-survey data. *Pakistan Journal of Statistics and Operation Research*, pp.211-226.

Aswi, A., Cramb, S., Duncan, E. and Mengersen, K., 2020. Evaluating the impact of a small number of areas on spatial estimation. *International journal of health geographics*, *19*(1), pp.1-14.

Austin, P.C., 2017. A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review*, *85*(2), pp.185-203.

Austin, P.C. and Fine, J.P., 2017. Practical recommendations for reporting F ine-G ray model analyses for competing risk data. *Statistics in medicine*, *36*(27), pp.4391-4400.

Austin, P.C., Lee, D.S. and Fine, J.P., 2016. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*(6), pp.601-609.

Austin, P.C., Lee, D.S. and Fine, J.P., 2016. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*(6), pp.601-609.

Austin, P.C. and Merlo, J., 2017. Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in medicine*, *36*(20), pp.3257-3277.

Austin, P.C., Steyerberg, E.W. and Putter, H., 2021. Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1. *Statistics in Medicine*.

Ayele, D., Zewotir, T. and Mwambi, H., 2014. Multiple correspondence analysis as a tool for analysis of large health surveys in African settings. *African health sciences*, *14*(4), pp.1036-1045.

Bae, H., Monti, S., Montano, M., Steinberg, M.H., Perls, T.T. and Sebastiani, P., 2016. Learning Bayesian networks from correlated data. *Scientific reports*, *6*(1), pp.1-14.

Banner, K.M., Irvine, K.M. and Rodhouse, T.J., 2020. The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, 11(8), pp.882-889.

Bender, R., Augustin, T. and Blettner, M., 2005. Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine*, *24*(11), pp.1713-1723.

Berger, J.O., De Oliveira, V. and Sansó, B., 2001. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, *96*(456), pp.1361-1374.

Berhe, D.F., Taxis, K., Haaijer-Ruskamp, F.M. and Mol, P.G., 2018. Healthcare professionals' level of medication knowledge in Africa: a systematic review. *British journal of clinical pharmacology*, *84*(12), pp.2729-2746.

Besag, J., York, J. and Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, *43*(1), pp.1-20.

Bewick, V., Cheek, L. and Ball, J., 2005. Statistics review 14: Logistic regression. *Critical care*, *9*(1), pp.1-7.

Box, G.E. and Tiao, G.C., 2011. *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.

Brant, R., 1990. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, pp.1171-1178.

Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), pp.9-25.

Cairns, M., Cheung, Y.B., Xu, Y., Asante, K.P., Owusu-Agyei, S., Diallo, D., Konate, A.T., Dicko, A., Chandramohan, D., Greenwood, B. and Milligan, P., 2015. Analysis of preventive interventions for malaria: exploring partial and complete protection and total and primary intervention effects. *American journal of epidemiology*, *181*(12), pp.1008-1017.

Cameron, A.C. and Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. *Journal of human resources*, *50*(2), pp.317-372.

Clayton, D. and Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, pp.671-681.

Coly, S., Garrido, M., Abrial, D. and Yao, A.F., 2021. Bayesian hierarchical models for disease mapping applied to contagious pathologies. *Plos one*, *16*(1), p.e0222898.

Corani, G., Benavoli, A., Demšar, J., Mangili, F. and Zaffalon, M., 2017. Statistical comparison of classifiers through Bayesian hierarchical modelling. *Machine Learning*, *106*(11), pp.1817-1837.

Cox, D.R. and OAKES, D., 2018. Analysis of Survival Data, vol. 21 of Chapman and Hall. *CRC Monographs on Statistics and Applied Probability. Routledge, Boca Raton*.

Cox, D.R., 1959. The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society: Series B (Methodological)*, *21*(2), pp.411-421.

Cressie, N., 1993. Aggregation in geostatistical problems. In *Geostatistics Troia'92* (pp. 25-36). Springer, Dordrecht.

Darmawan, A.S., Anggraeni, D. and Tirta, I.M., 2020, May. Fitting the rice production model using generalized additive mixed model and generalized estimating equation with shiny web application. In *Journal of Physics: Conference Series* (Vol. 1538, No. 1, p. 012032). IOP Publishing.

Davis, B., Ladner, J., Sams, K., Tekinturhan, E., de Korte, D. and Saba, J., 2013. Artemisinin-based combination therapy availability and use in the private sector of five AMFm phase 1 countries. *Malaria journal*, 12(1), pp.1-9.

Das, S. and Rahman, R.M., 2011. Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition journal*, *10*(1), pp.1-11.

Dasgupta, P., Cramb, S.M., Aitken, J.F., Turrell, G. and Baade, P.D., 2014. Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: a case study. *International journal of health geographics*, *13*(1), pp.1-14.

De Oliveira, V., 2012. Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, *64*(1), pp.107-133.

Dell-Kuster, S., Droeser, R.A., Schäfer, J., Gloy, V., Ewald, H., Schandelmaier, S., Hemkens, L.G., Bucher, H.C., Young, J. and Rosenthal, R., 2018. Systematic review and simulation study of ignoring clustered data in surgical trials. *Journal of British Surgery*, *105*(3), pp.182-191.

Desai, M., Buff, A.M., Khagayi, S., Byass, P., Amek, N., van Eijk, A., Slutsker, L., Vulule, J., Odhiambo, F.O., Phillips-Howard, P.A. and Lindblade, K.A., 2014. Age-specific malaria

mortality rates in the KEMRI/CDC health and demographic surveillance system in western Kenya, 2003–2010. *PloS one*, *9*(9), p.e106197.

Dickinson, L.M. and Basu, A., 2005. Multilevel modeling and practice-based research. *The Annals of Family Medicine*, *3*(suppl 1), pp.S52-S60.

Dienes, Z., 2011. Bayesian versus orthodox statistics: Which side are you on?. *Perspectives on Psychological Science*, *6*(3), pp.274-290.

Dienes, Z., 2016. How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, pp.78-89.

Ding, C., Wang, D., Liu, C., Zhang, Y. and Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, *100*, pp.65-80.

Division of National Malaria Programme (DNMP) [Kenya] and ICF, 2021. *2020 Kenya Malaria Indicator Survey Summary Report*. Nairobi, Kenya and Rockville, Maryland, USA: DNMP and ICF.

Dixon, S.N., Darlington, G.A. and Desmond, A.F., 2011. A competing risks model for correlated data based on the subdistribution hazard. *Lifetime data analysis*, *17*(4), pp.473-495.

Do Ha, I. and Lee, Y., 2021. A review of h-likelihood for survival analysis. *Japanese Journal of Statistics and Data Science*, pp.1-22.

Dohoo, I.R., Martin, S.W. and Stryhn, H., 2012. *Methods in epidemiologic research*.

Dondorp, A.M., Fanello, C.I., Hendriksen, I.C., Gomes, E., Seni, A., Chhaganlal, K.D., Bojang, K., Olaosebikan, R., Anunobi, N., Maitland, K. and Kivaya, E., 2010. Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): an open-label, randomised trial. *The Lancet*, *376*(9753), pp.1647-1657.

Eboreime, E.A., Eyles, J., Nxumalo, N., Eboreime, O.L. and Ramaswamy, R., 2019. Implementation process and quality of a primary health care system improvement initiative in a decentralized context: a retrospective appraisal using the quality

implementation framework. *The International journal of health planning and management*, *34*(1), pp.e369-e386.

Edwards, W., Lindman, H. and Savage, L.J., 1963. Bayesian statistical inference for psychological research. *Psychological review*, *70*(3), p.193.

Eliades, M.J., Alombah, F., Wun, J., Burnett, S.M., Clark, T., Ntumy, R., Chikoko, A., Onditi, S., Mkomwa, Z., Makanka, D. and Hamilton, P., 2019. Perspectives on implementation considerations and costs of malaria case management supportive supervision. *The American journal of tropical medicine and hygiene*, 100(4), p.861.

Elnour, F.A., Alagib, M.E., Bansal, D., Abd Farag, E.A.B. and Malik, E.M., 2019. Severe malaria management: current situation, challenges and lessons learned from Gezira State, Sudan. *Malaria journal*, *18*(1), pp.1-8.

English, M., Waruiru, C., Amukoye, E., Murphy, S., Crawley, J., Mwangi, I., Peshu, N. and Marsh, K., 1996. Deep breathing in children with severe malaria: indicator of metabolic acidosis and poor outcome. *The American journal of tropical medicine and hygiene*, *55*(5), pp.521-524.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W. and Kühn, I., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), pp.609-628.

Feleke, D.G., Tarko, S. and Hadush, H., 2017. Performance comparison of CareStart™ HRP2/pLDH combo rapid malaria test with light microscopy in north-western Tigray, Ethiopia: a cross-sectional study. *BMC infectious diseases*, *17*(1), pp.1-7.

Ferrari, F. and Dunson, D.B., 2021. Bayesian factor analysis for inference on interactions. *Journal of the American Statistical Association*, *116*(535), pp.1521-1532.

Fine, J.P. and Gray, R.J., 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, *94*(446), pp.496-509.

Fitzmaurice, G.M., 2005. Overview of methods for analyzing cluster-correlated data. *Boston: Harvard School of Public Health*.

Force, R.B.M.A.T., 2007. Affordable Medicines Facility-malaria (AMFm): Technical Design. *Rollbackmalaria. org*, November.

Gachot, B. and Ringwald, P., 1998. Paludisme pernicieux: Paludisme. *La Revue du praticien (Paris)*, *48*(3), pp.273-278.

Galbraith, S., Daniel, J.A. and Vissel, B., 2010. A study of clustered data and approaches to its analysis. *Journal of Neuroscience*, *30*(32), pp.10601-10608.

Gathara, D., Nyamai, R., Were, F., Mogoa, W., Karumbi, J., Kihuba, E., Mwinga, S., Aluvaala, J., Mulaku, M., Kosgei, R. and Todd, J., 2015. Moving towards routine evaluation of quality of inpatient pediatric care in Kenya. *PloS one*, *10*(3), p.e0117048.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2013. *Bayesian data analysis*. CRC press.

Geskus, R.B., 2011. Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, *67*(1), pp.39-49.

Ghosh, S.K. and Kyung, M., 2009. Bayesian inference for directional conditionally autoregressive models. *Bayesian Analysis*, *4*(4), pp.675-706.

Gorfine, M. and Hsu, L., 2011. Frailty-based competing risks model for multivariate survival data. *Biometrics*, *67*(2), pp.415-426.

Government of Kenya (2008). Kenya vision 2030.

Government of Kenya (2018). The Big Four.

Gray, R.J., 1988. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of statistics*, pp.1141-1154.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. and Altman, D.G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, *31*(4), pp.337-350.

Gumulira, A., 2021. Malaria vaccine approval: a step change for global health. *N Engl J Med*, *385*, pp.1005-17.

Guo, Z., Gill, T.M. and Allore, H.G., 2008. Modeling repeated time-to-event health conditions with discontinuous risk intervals. *Methods of information in medicine*, *47*(02), pp.107-116.

Gute, C.G., Reddy, C., Taye, A., and Rao, K.V., 2015. Statistical Analysis of Risk Factors of Malaria related in-hospital Mortality: A Case Study at Bushulo Major Health Center, Hawassa. *International journal of latest research in science and technology*. 4(5), pp.32-42.

Gwitira, I., Murwira, A., Zengeya, F.M. and Shekede, M.D., 2018. Application of GIS to predict malaria hotspots based on Anopheles arabiensis habitat suitability in Southern Africa. *International journal of applied earth observation and geoinformation*, *64*, pp.12-21.

Ha, I.D., Christian, N.J., Jeong, J.H., Park, J. and Lee, Y., 2016. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. *Statistical methods in medical research*, *25*(6), pp.2488-2505.

Hahs-Vaughn, D.L., 2005. A primer for using and understanding weights with national datasets. *The Journal of experimental education*, *73*(3), pp.221-248.

Hanandita, W. and Tampubolon, G., 2016. Geography and social distribution of malaria in Indonesian Papua: a cross-sectional study. *International journal of health geographics*, *15*(1), pp.1-15.

Hanley, J.A., Negassa, A., Edwardes, M.D.D. and Forrester, J.E., 2003. Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology*, *157*(4), pp.364-375.

Harrell, F.E., 2015. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis* (Vol. 3). New York: springer.

Hedeker, D., Demirtas, H. and Mermelstein, R.J., 2009. A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Statistics and its Interface*, *2*(4), p.391.

Hoffmeister, B., 2021. Factors Associated with Prolonged Hospital Length of Stay in Adults with Imported Falciparum Malaria—An Observational Study from a Tertiary Care University Hospital in Berlin, Germany. *Microorganisms*, *9*(9), p.1941.

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

Hosmer, D.W., Taber, S. and Lemeshow, S., 1991. The importance of assessing the fit of logistic regression models: a case study. *American journal of public health*, *81*(12), pp.1630-1635.

Hox, J., Balderjahn, I., Mathar, R. and Schader, M., 1998. Classification, data analysis, and data highways. *Multilevel modeling: When and why. New York: Springer Verlag*, pp.147-154.

Huang, A. and Wand, M.P., 2013. Simple marginally noninformative prior distributions for covariance matrices. Bayesian Analysis, 8(2), pp.439-452.

Hu, J., Chen, Y., Leng, C. and Tang, C.Y., 2021. Regression Analysis of Correlations for Correlated Data. *arXiv preprint arXiv:2109.05861*.

Idro, R. and Aloyo, J., 2004. Manifestations, quality of emergency care and outcome of severe malaria in Mulago Hospital, Uganda. *African health sciences*, *4*(1), pp.50-57.

Jinadu, K.A., Adebiyi, A.O., Sekoni, O.O. and Bamgboye, E.A., 2018. Integrated disease surveillance and response strategy for epidemic prone diseases at the primary health care (PHC) level in Oyo State, Nigeria: what do health care workers know and feel?. *Pan African Medical Journal*, *31*(1).

Kanters, S., 2022. Fixed-and Random-Effects Models. In *Meta-Research* (pp. 41-65). Humana, New York, NY.

Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), pp.457-481.

Kartsonaki, C., 2016. Survival analysis. *Diagnostic Histopathology*, 22(7), pp.263-270.

Kassahun, W., Neyens, T., Molenberghs, G., Faes, C. and Verbeke, G., 2014. Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros. *Statistics in medicine*, *33*(25), pp.4402-4419.

Katsahian, S. and Boudreau, C., 2011. Estimating and testing for center effects in competing risks. *Statistics in medicine*, *30*(13), pp.1608-1617.

Kazembe, L.N., Chirwa, T.F., Simbeye, J.S. and Namangale, J.J., 2008. Applications of Bayesian approach in modelling risk of malaria-related hospital mortality. *BMC medical research methodology*, *8*(1), pp.1-14.

Keene, C.M., Dondorp, A., Crawley, J., Ohuma, E.O. and Mukaka, M., 2018. A competing-risk approach for modeling length of stay in severe malaria patients in South-East Asia and the implications for planning of hospital services. *Clinical Infectious Diseases*, *67*(7), pp.1053-1062.

Khiari, M. and ben Rejeb, J., 2015. Determination of the regional impact on innovation with an ordinal logit and a multilevel analysis. *Procedia-Social and Behavioral Sciences*, *195*, pp.592-602.

Khuu, D., Eberhard, M.L., Bristow, B.N., Javanbakht, M., Ash, L.R., Shafir, S.C. and Sorvillo, F.J., 2018. Risk factors for severe malaria among hospitalized patients in the United States, 2000–2014. *Infection, Disease & Health*, *23*(2), pp.93-106.

Kiguba, R., Karamagi, C. and Bird, S.M., 2021. Quality of care for adult in-patients with malaria in a tertiary hospital in Uganda. *Malaria journal*, *20*(1), pp.1-15.

Kim, S. and Kim, Y., 2019. Spatially filtered multilevel analysis on spatial determinants for malaria occurrence in Korea. *International Journal of Environmental Research and Public Health*, *16*(7), p.1250.

Klein, J.P. and Andersen, P.K., 2005. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, *61*(1), pp.223-229.

Kleinbaum, D.G. and Klein, M., 2010. Introduction to logistic regression. In *Logistic regression* (pp. 1-39). Springer, New York, NY.

Kruschke, J.K. and Liddell, T.M., 2018. Bayesian data analysis for newcomers, Psychon. B. Rev., 25, 155–177.

Kruschke, J.K., 2010. Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), pp.658-676.

Kruschke, J.K. and Vanpaemel, W., 2015. Bayesian estimation in hierarchical models. *The Oxford handbook of computational and mathematical psychology*, *279*.

Kurtz, E., 2016. Factors Impacting the Effectiveness of Health Care Worker Behavior Change: A Literature Review. *Baltimore: Johns Hopkins Center for Communication Programs*.

LaHuis, D.M., Hartman, M.J., Hakoyama, S. and Clark, P.C., 2014. Explained variance measures for multilevel models. *Organizational Research Methods*, 17(4), pp.433-451.

Lambert, P.C., 2017. The estimation and modeling of cause-specific cumulative incidence functions using time-dependent weights. *The Stata Journal*, *17*(1), pp.181-207.

Larson, M.G. and Dinse, G.E., 1985. A mixture model for the regression analysis of competing risks data. *Journal of the royal statistical society: series C (Applied Statistics)*, *34*(3), pp.201-211.

Lau, B., Cole, S.R. and Gange, S.J., 2009. Competing risk regression models for epidemiologic data. *American journal of epidemiology*, *170*(2), pp.244-256.

Lawson A, Lee D. Bayesian disease mapping for public health. In Handbook of statistics 2017 Jan 1 (Vol. 36, pp. 443-481). Elsevier.

Lawson, A.B., 2018. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press.

Lee, S.J., Stepniewska, K., Anstey, N., Ashley, E., Barnes, K., Binh, T.Q., D'Alessandro, U., Day, N.P., de Vries, P.J., Dorsey, G. and Guthmann, J.P., 2008. The relationship between the haemoglobin concentration and the haematocrit in Plasmodium falciparum malaria. *Malaria journal*, *7*(1), pp.1-4.

Lee, Y. and Nelder, J.A., 2004. Conditional and marginal models: another view. *Statistical Science*, *19*(2), pp.219-238.

Lennon, J.J., 2000. Red-shifts and red herrings in geographical ecology. *Ecography*, *23*(1), pp.101-113.

Lesaffre, E., Steyerberg, E.W., Lingsma, H.F. and Li, B., 2011. Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes.

Levitz, L., Janko, M., Mwandagalirwa, K., Thwai, K.L., Likwela, J.L., Tshefu, A.K., Emch, M. and Meshnick, S.R., 2018. Effect of individual and community-level bed net usage on malaria prevalence among under-fives in the Democratic Republic of Congo. *Malaria journal*, *17*(1), pp.1-8.

Li, W. and Fearnhead, P., 2018. Convergence of regression-adjusted approximate Bayesian computation. *Biometrika*, *105*(2), pp.301-318.

Liang, K.Y. and Zeger, S.L., 1993. Regression analysis for correlated data. *Annual review of public health*, *14*(1), pp.43-68.

Lin, D.Y., 1994. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, *13*(21), pp.2233-2247.

Liu, I. and Agresti, A., 2005. The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, *14*(1), pp.1-73.

Liu, H. and Zhu, X., 2016. Exploring the influence of neighborhood characteristics on burglary risks: A Bayesian random effects modeling approach. *ISPRS International Journal of Geo-Information*, *5*(7), p.102.

Liu, H., 1998. Robust standard error estimate for cluster sampling data: a SAS/IML macro procedure for logistic regression with Huberization. In *Proceedings of the Twenty-Third Annual SAS Users Group International*.

Liu, X. and Koirala, H., 2013. Fitting proportional odds models to educational data with complex sampling designs in ordinal logistic regression. *Journal of Modern Applied Statistical Methods*, *12*(1), p.26.

Lomas, J., Anderson, G.M., Domnick-Pierre, K., Vayda, E., Enkin, M.W. and Hannah, W.J., 1989. Do practice guidelines guide practice?. *New England Journal of Medicine*, *321*(19), pp.1306-1311.

Lombardo, P., Vaucher, P., Rarau, P., Mueller, I., Favrat, B. and Senn, N., 2017. Hemoglobin levels and the risk of malaria in Papua New Guinean infants: a nested cohort study. *The American journal of tropical medicine and hygiene*, *97*(6), p.1770.

Luo, W., Li, H., Baek, E., Chen, S., Lam, K.H. and Semma, B., 2021. Reporting Practice in Multilevel Modeling: A Revisit After 10 Years. *Review of Educational Research*, *91*(3), pp.311-355.

Ma, Z. and Chen, G., 2018. Bayesian methods for dealing with missing data problems. *Journal of the Korean Statistical Society*, *47*(3), pp.297-313.

Mace KE, Gueye AS, Lynch MF, *et al*. An evaluation of methods for assessing the quality of case management for inpatients with malaria in Benin. Am J *Trop Med Hyg*. 2014 Aug;91(2):354-60.

Macharia, P.M., Giorgi, E., Noor, A.M., Waqo, E., Kiptui, R., Okiro, E.A. and Snow, R.W., 2018. Spatio-temporal analysis of Plasmodium falciparum prevalence to understand the past and chart the future of malaria control in Kenya. *Malaria journal*, *17*(1), pp.1-13.

Mahase, E., 2021. Malaria vaccine becomes first to achieve 75% efficacy goal in trial of children. BMJ: *British Medical Journal* (Online), 373.

Mansournia, M.A., Nazemipour, M., Naimi, A.I., Collins, G.S. and Campbell, M.J., 2021. Reflection on modern methods: demystifying robust standard errors for epidemiologists. *International Journal of Epidemiology*, *50*(1), pp.346-351.

Marteau, A., Ouedraogo, E., Van der Meersch, G., Akhoundi, M., Souhail, B., Cohen, Y., Bouchaud, O. and Izri, A., 2021. Severe long-delayed malaria caused by Plasmodium malariae in an elderly French patient. *Malaria Journal*, *20*(1), pp.1-5.

McCullagh, P., 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*(2), pp.109-127.

Menard, S., 2002. *Applied logistic regression analysis* (Vol. 106). Sage. Paper series on quantitative applications in the social sciences.

Mikomangwa, W.P., Kaaya, C., Kilonzi, M., Mlyuka, H., Marealle, A.I. and Mutagonda, R., 2019. Level of knowledge among health care providers on preparation of injectable artesunate for treatment of severe malaria in public health facilities in Tanzania. *BMC research notes*, *12*(1), pp.1-5.

Ministry of Health, 2012a. *National guidelines for diagnosis, treatment and prevention of malaria for health workers.* 4th ed. Nairobi, Kenya: Division of Malaria Control.

Ministry of Health, 2012b. *Participants manual for diagnosis, management and prevention of malaria for health workers*. Nairobi, Kenya: National Malaria Control Programme.

Ministry of Health, 2015a. *Guidelines for administration of injectable artesunate for severe malaria.* Nairobi, Kenya: National Malaria Control Programme.

Ministry of Health, 2015b. *National Guidelines for Diagnosis, Treatment and Prevention of Malaria for Health Workers in Kenya*. Nairobi, Kenya: Division of National Malaria Programme.

Ministry of Health, 2016. *Diagnosis, management, and prevention of malaria in Kenya: Participants manual*. Nairobi, Kenya: National Malaria Control Programme.

Ministry of Health, 2019a. *Kenya Malaria Monitoring and Evaluation Plan 2019-2023*. Nairobi, Kenya: National Malaria Control Programme.

Ministry of Health, 2019b. *Kenya Malaria Strategy 2019-2023*. Nairobi, Kenya: Nairobi, Kenya: National Malaria Control Programme.

Ministry of Health, 2020a. *National Guidelines for Diagnosis, Treatment and Prevention of Malaria for Health Workers in Kenya*. Nairobi, Kenya: Division of National Malaria Programme.

Ministry of Health, 2020b. *Manual for malaria support supervision*. Nairobi, Kenya: Division of National Malaria Programme.

Moen, E.L., Fricano-Kugler, C.J., Luikart, B.W. and O'Malley, A.J., 2016. Analyzing clustered data: why and how to account for multiple observations nested within a study participant?. *Plos one, 11*(1), p.e0146721.

Moody, A., 2002. Rapid diagnostic tests for malaria parasites. *Clinical microbiology reviews, 15*(1), pp.66-78.

Moon, A.M., Biggs, H.M., Rubach, M.P., Crump, J.A., Maro, V.P., Saganda, W. and Reddy, E.A., 2014. Evaluation of in-hospital management for febrile illness in Northern Tanzania before and after 2010 World Health Organization Guidelines for the treatment of malaria. *PLoS One, 9*(2), p.e89814.

Muff, S., Held, L. and Keller, L.F., 2016. Marginal or conditional regression models for correlated non-normal data?. *Methods in Ecology and Evolution, 7*(12), pp.1514-1524.

Müller, S., Scealy, J.L. and Welsh, A.H., 2013. Model selection in linear mixed models. *Statistical Science, 28*(2), pp.135-167.

Murray, C.J., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., Fullman, N., Naghavi, M., Lozano, R. and Lopez, A.D., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet, 379*(9814), pp.413-431.

Mutua, M.M., Achia, T.N., Manderson, L. and Musenge, E., 2019. Spatial and socio-economic correlates of effective contraception among women seeking post-abortion care in healthcare facilities in Kenya. *PloS one, 14*(3), p.e0214049.

National Malaria Control Programme (NMCP), 2016. Kenya National Bureau of Statistics (KNBS) and ICF, International. Kenya Malaria Indicator Survey 2015. *Nairobi, Kenya, and Rockville, Maryland, USA: NMCP, KNBS, and ICF International*.

Nicolaie, M.A., van Houwelingen, H.C. and Putter, H., 2010. Vertical modeling: a pattern mixture approach for competing risks modeling. *Statistics in medicine, 29*(11), pp.1190-1205.

Obaromi, D., 2019. Spatial modelling of some conditional autoregressive priors in a disease mapping model: the Bayesian approach. *Biomedical Journal of Scientific & Technical Research, 14*(3).

O'Brien, L.M. and Fitzmaurice, G.M., 2005. Regression models for the analysis of longitudinal Gaussian data from multiple sources. *Statistics in medicine*, *24*(11), pp.1725-1744.

Odey, F., Esu, E., Effa, E., Udoh, E., Oduwole, O., Chibuzor, M., Oyo-Ita, A. and Meremikwu, M., 2013. Management of severe malaria in children under 5 years of age in private and public health facilities in Cross River State, southeastern Nigeria: an audit of current practices. *Clinical Audit*, *5*, p.43.

Ogero, M., Ayieko, P., Boniface Makone, T.J., Malla, L., Oliwa, J., Irimu, G. and English, M., 2018. An observational study of monitoring of vital signs in children admitted to Kenyan hospitals: an insight into the quality of nursing care?. *Journal of global health*, *8*(1).

Ojo, A.A., Maxwell, K., Oresanya, O., Adaji, J., Hamade, P., Tibenderana, J.K., Abubakar, S.S., Audu, B.M., Njidda, A., Gubio, A.B. and Snow, R.W., 2020. Health systems readiness and quality of inpatient malaria case-management in Kano State, Nigeria. *Malaria Journal*, *19*(1), pp.1-12.

Ouma, P.O., Maina, J., Thuranira, P.N., Macharia, P.M., Alegana, V.A., English, M., Okiro, E.A. and Snow, R.W., 2018. Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health*, *6*(3), pp.e342-e350.

Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. and Fisher, M.R., 1996. A survey of methods for analyzing clustered binary response data. *International Statistical Review/Revue Internationale de Statistique*, pp.89-118.

Perezgonzalez, J.D., 2016. Commentary: How Bayes factors change scientific practice. *Frontiers in psychology*, *7*, p.1504.

Phillips, A., Bassett, P., Szeki, S., Newman, S. and Pasvol, G., 2009. Risk factors for severe disease in adults with falciparum malaria. *Clinical Infectious Diseases*, *48*(7), pp.871-878.

Pintilie, M., 2011. An introduction to competing risks analysis. *Spanish Journal of Cardiology (English Edition)*, *64* (7), pp.599-605.

Plucinski, M.M., Ferreira, M., Ferreira, C.M., Burns, J., Gaparayi, P., João, L., da Costa, O., Gill, P., Samutondo, C., Quivinja, J. and Mbounga, E., 2017. Evaluating malaria case

management at public health facilities in two provinces in Angola. *Malaria journal*, *16*(1), pp.1-10.

Prentice, R.L., Williams, B.J. and Peterson, A.V., 1981. On the regression analysis of multivariate failure time data. *Biometrika*, *68*(2), pp.373-379.

Prentice, R.L., Kalbfleisch, J.D., Peterson Jr, A.V., Flournoy, N., Farewell, V.T. and Breslow, N.E., 1978. The analysis of failure times in the presence of competing risks. *Biometrics*, pp.541-554.

Rashidi, P., Wang, T., Skidmore, A., Mehdipoor, H., Darvishzadeh, R., Ngene, S., Vrieling, A. and Toxopeus, A.G., 2016. Elephant poaching risk assessed using spatial and non-spatial Bayesian models. *Ecological modelling*, *338*, pp.60-68.

Reeves, S., Pelone, F., Harrison, R., Goldman, J. and Zwarenstein, M., 2017. Interprofessional collaboration to improve professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*, (6).

Rek, J., Musiime, A., Zedi, M., Otto, G., Kyagamba, P., Asiimwe Rwatooro, J., Arinaitwe, E., Nankabirwa, J., Staedke, S.G., Drakeley, C. and Rosenthal, P.J., 2020. Non-adherence to long-lasting insecticide treated bednet use following successful malaria control in Tororo, Uganda. *PloS one*, *15*(12), p.e0243303.

Rowe, A.K., Hamel, M.J., Flanders, W.D., Doutizanga, R., Ndoyo, J. and Deming, M.S., 2000. Predictors of correct treatment of children with fever seen at outpatient health facilities in the Central African Republic. *American Journal of Epidemiology*, *151*(10), pp.1029-1035.

Sainani, K., 2010. The importance of accounting for correlated observations. *PM&R*, *2*(9), pp.858-861.

Salinelli, E. and Tomarelli, F., 2014. Markov chains. In *Discrete Dynamical Models* (pp. 227-255). Springer, Cham.

Sánchez, G. and StataCorp, L.L.C., 2017. Introduction to Bayesian Analysis in Stata.

Schmidt-Catran, A.W. and Fairbrother, M., 2016. The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review*, *32*(1), pp.23-38.

Schuster, N.A., Hoogendijk, E.O., Kok, A.A., Twisk, J.W. and Heymans, M.W., 2020. Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *Journal of clinical epidemiology*, *122*, pp.42-48.

Sears, D., Mpimbaza, A., Kigozi, R., Sserwanga, A., Chang, M.A., Kapella, B.K., Yoon, S., Kamya, M.R., Dorsey, G. and Ruel, T., 2015. Quality of inpatient pediatric case management for four leading causes of child mortality at six government-run Ugandan hospitals. *PLoS One*, *10*(5), p.e0127192.

Shah, M.P., Briggs-Hagen, M., Chinkhumba, J., Bauleni, A., Chalira, A., Moyo, D., Dodoli, W., Luhanga, M., Sande, J., Ali, D. and Gutman, J., 2016. Adherence to national guidelines for the diagnosis and management of severe malaria: a nationwide, cross-sectional survey in Malawi, 2012. *Malaria journal*, *15*(1), pp.1-11.

Sharifi-Malvajerdi, S., Zhu, F., Fogarty, C.B., Fay, M.P., Fairhurst, R.M., Flegg, J.A., Stepniewska, K. and Small, D.S., 2019. Malaria parasite clearance rate regression: an R software package for a Bayesian hierarchical regression model. *Malaria journal*, *18*(1), pp.1-16.

Shayo, E.H., Våga, B.B., Moland, K.M., Kamuzora, P. and Blystad, A., 2014. Challenges of disseminating clinical practice guidelines in a weak health system: the case of HIV and infant feeding recommendations in Tanzania. *International breastfeeding journal*, *9*(1), pp.1-13.

Shor, B., Bafumi, J., Keele, L. and Park, D., 2007. A Bayesian multilevel modeling approach to time-series cross-sectional data. *Political Analysis*, *15*(2), pp.165-181.

Snow, R.W., Kibuchi, E., Karuri, S.W., Sang, G., Gitonga, C.W., Mwandawiro, C., Bejon, P. and Noor, A.M., 2015. Changing malaria prevalence on the Kenyan coast since 1974: climate, drugs and vector control. *Plos one*, *10*(6), p.e0128792.

Snow, R.W., 2015. Global malaria eradication and the importance of Plasmodium falciparum epidemiology in Africa. *BMC medicine*, *13*(1), pp.1-3.

Sommet, N. and Morselli, D., 2017. Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*, pp.203-218.

Sperandei, S., 2014. Understanding logistic regression analysis. *Biochemia medica*, *24*(1), pp.12-18.

Sserwanga, A., Sears, D., Kapella, B.K., Kigozi, R., Rubahika, D., Staedke, S.G., Kamya, M., Yoon, S.S., Chang, M.A., Dorsey, G. and Mpimbaza, A., 2015. Anti-malarial prescription practices among children admitted to six public hospitals in Uganda from 2011 to 2013. *Malaria journal*, *14*(1), pp.1-10.

Tabachnick, B.G. and Fidell, L.S., 2014. Using multivariate statistics. Harlow. *Essex: Pearson Education Limited*.

Taylor, S.L., Sen, S., Greenhalgh, D.G., Lawless, M., Curri, T. and Palmieri, T.L., 2015. A competing risk analysis for hospital length of stay in patients with burns. *JAMA surgery*, *150*(5), pp.450-456.

Taylor, W.R., Hanson, J., Turner, G.D., White, N.J. and Dondorp, A.M., 2012. Respiratory manifestations of malaria. *Chest*, *142*(2), pp.492-505.

Thomas, S.L. and Heck, R.H., 2001. Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in higher education*, *42*(5), pp.517-540.

Toda, M., Zurovac, D., Njeru, I., Kareko, D., Mwau, M. and Morita, K., 2018. Health worker knowledge of Integrated Disease Surveillance and Response standard case definitions: a cross-sectional survey at rural health facilities in Kenya. *BMC Public Health*, *18*(1), pp.1-8.

Tutz, G., 2011. *Regression for categorical data* (Vol. 34). Cambridge University Press.

Tutz, G. and Oelker, M.R., 2017. Modelling clustered heterogeneity: Fixed effects, random effects and mixtures. *International statistical review*, *85*(2), pp.204-227.

Twabi, H.S. and Mukaka, M., 2018. Modelling Length of Hospital Stay for Tuberculosis Treated in-Patients at Queen Elizabeth Central Hospital: A Competing Risk Perspective. *Open Access Biostatistics & Bioinformatics, 1 (2)*.

Ugah, U.I., Alo, M.N., Owolabi, J.O., Okata-Nwali, O.D., Ekejindu, I.M., Ibeh, N. and Elom, M.O., 2017. Evaluation of the utility value of three diagnostic methods in the detection of malaria parasites in endemic area. *Malaria journal*, *16*(1), pp.1-8.

Umer, M.F., Zofeen, S., Majeed, A., Hu, W., Qi, X. and Zhuang, G., 2019. Effects of socio-environmental factors on malaria infection in Pakistan: a Bayesian spatial analysis. *International journal of environmental research and public health*, *16*(8), p.1365.

United States Agency for International Development (USAID), 2019. *President's Malaria Initiative. Malaria Operational Plan FY 2019*. Washington, DC: USAID.

Van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J. and Yau, C., 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), pp.1-26.

Villegas, R., Julià, O. and Ocaña, J., 2013. Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring. *BMC medical research methodology*, *13*(1), pp.1-10.

Waller, L.A., Carlin, B.P., Xia, H. and Gelfand, A.E., 1997. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, *92*(438), pp.607-617.

Wang, X., Shao, C., Yin, C., Zhuge, C. and Li, W., 2018. Application of bayesian multilevel models using small and medium size city in China: the case of Changchun. *Sustainability*, *10*(2), p.484.

Wang, X., Shao, C., Yin, C. and Zhuge, C., 2018. Exploring the influence of built environment on car ownership and use with a spatial multilevel model: A case study of Changchun, China. *International journal of environmental research and public health*, *15*(9), p.1868.

Wei, L.J., Lin, D.Y. and Weissfeld, L., 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, *84*(408), pp.1065-1073.White NJ. Anaemia and malaria. *Malaria journal*. 2018 Dec;17(1):1-7.

Wilairatana, P., Tangpukdee, N. and Krudsood, S., 2013. Practical aspects of artesunate administration in severe malaria treatment. *Tropical Medicine & Surgery*, *1*(7).

Williams, R.A. and Quiroz, C., 2020. *Ordinal regression models*. SAGE Publications Limited.

Wolbers, M., Koller, M.T., Stel, V.S., Schaer, B., Jager, K.J., Leffondre, K. and Heinze, G., 2014. Competing risks analyses: objectives and approaches. *European heart journal*, *35*(42), pp.2936-2941.

Wondaya, S., Kifle, Y.G., Tereda, A.B. and Seyoum, D., 2016. Modeling Time to First Malaria Using Spatially Correlated Conditional Autoregressive Frailty Model.

Worges, M., Whitehurst, N., Yamo, E., Moonga, H., Yukich, J. and Benavente, L., 2018. Outreach training and supportive supervision for malaria case management in Zambia: the effects of focused capacity building on indicators of diagnostic and clinical performance. *Malaria journal*, *17*(1), pp.1-16.

World Health Organization, 2000. *Management of severe malaria: a practical handbook*. World Health Organization.

World Health Organization, 2007. Insecticide-Treated Mosquito Nets: a WHO Position Statement. *Geneva*: World Health Organization.

World Health Organization, 2009. *Malaria case management: Operations manual*. World Health Organization.

World Health Organization, 2010. Parasitological confirmation of malaria diagnosis: report of a WHO technical consultation, Geneva, 6-8 October 2009.

World Health Organization, 2012. Management of severe malaria: a practical handbook 3rd ed. 2012. *Geneva*: World Health Organization.

World Health Organization, 2015a. *Global technical strategy for malaria 2016-2030*. World Health Organization.

World Health Organization, 2015a. *Global technical strategy for malaria 2016-2030*. World Health Organization.

World Health Organization, 2015b. *Guidelines for the treatment of malaria*. World Health Organization.

World Health Organization, 2016. *World malaria report 2015*. World Health Organization.

World Health Organization, 2017a. *World malaria report 2017*. Geneva: World Health Organization.

World Health Organization, 2017b. *Malaria in Pregnant Women*. World Health Organization.

World Health Organization, 2018. *Malaria surveillance, monitoring & evaluation: a reference manual*. World Health Organization.

World Health Organization, 2019a. *World malaria report 2019*. World Health Organization.

World Health Organization, 2019b. *Intermittent Preventive Treatment in Pregnancy*. World Health Organization.

World Health Organization, 2021. *World malaria report 2021*. World Health Organization

Ye, Y., Arnold, F., Noor, A., Wamukoya, M., Amuasi, J., Blay, S., Mberu, B., Ren, R., Kyobutungi, C., Wekesah, F. and Gatakaa, H., 2015. The Affordable Medicines Facility-malaria (AMFm): are remote areas benefiting from the intervention?. *Malaria journal*, 14(1), pp.1-11.

Ye, Y. and Duah, D., 2019. The President's Malaria Initiative contributed to reducing malaria burden in sub-Saharan Africa between 2004 and 2014: evidence from generalized estimating equation analysis. *PLoS one*, *14*(5), p.e0217103.

Yoo, J.Y., Kim, J.H., Kim, J.S., Kim, H.L. and Ki, J.S., 2019. Clinical nurses' beliefs, knowledge, organizational readiness and level of implementation of evidence-based practice: The first step to creating an evidence-based practice culture. *PloS one*, *14*(12), p.e0226742.

Yu, H., Jiang, S. and Land, K.C., 2015. Multicollinearity in hierarchical linear models. *Social science research*, *53*, pp.118-136.

Zeger, S.L., Liang, K.Y. and Albert, P.S., 1988. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, pp.1049-1060.

Zekar, L. and Sharman, T., 2020. Plasmodium Falciparum Malaria.

Zhou, B., Fine, J., Latouche, A. and Labopin, M., 2012. Competing risks regression for clustered data. *Biostatistics*, *13*(3), pp.371-383.

Zyzanski, S.J., Flocke, S.A. and Dickinson, L.M., 2004. On the nature and analysis of clustered data. *The Annals of Family Medicine*, *2*(3), pp.199-200.

Zurovac, D., Githinji, S., Memusi, D., Kigen, S., Machini, B., Muturi, A., Otieno, G., Snow, R.W. and Nyandigisi, A., 2014. Major improvements in the quality of malaria case-management under the "test and treat" policy in Kenya. *PLoS One*, *9*(3), p.e92782.

Zurovac, D., Machini, B., Kiptui, R., Memusi, D., Amboko, B., Kigen, S., Njiri, P. and Waqo, E., 2018. Monitoring health systems readiness and inpatient malaria case-management at Kenyan county hospitals. *Malaria journal*, *17*(1), pp.1-15.

APPENDICES

Appendix 1. Permission to nest the study

MINISTRY OF HEALTH
NATIONAL MALARIA CONTROL PROGRAMME

Telephone: Nairobi 2716934/5 Fax 2716935
E-Mail kenyadomc@gmail.com
All correspondence should be addressed to the Head, NMCP
When replying please quote:

KENYATTA HOSPITAL GROUNDS
P. O. BOX 19982 – 00202
KNH NAIROBI

DATE: 26th March, 2018.

Beatrice Machini
Reg. No W83/52/2017
PhD Candidate
Institute of Tropical and Infectious Diseases (UNITID)
College of Health Sciences
University of Nairobi

Dear Beatrice,

RE: Permission to nest PHD study on the ongoing Monitoring Quality of Inpatient Malaria Case Management at Kenyan County Hospitals (KNH/ERC/R/160: P643/10/2015)

We acknowledge receipt of your letter dated 21st March, 2018 on the above subject matter.

You may use the Malaria Inpatient Quality of Care data for your PHD studies as indicated in your letter of request, with the understanding that this will be in collaboration with the National Malaria Control Programme (NMCP).

The NMCP will thereafter be informed of your progress and the findings of your research.

Yours sincerely,

Dr. Waqo Ejersa
Head - National Malaria Control Programme

ISO 9001:2008
Certified

Appendix 2. Ethical approval

UNIVERSITY OF NAIROBI
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KENYATTA NATIONAL HOSPITAL APPROVED 22 JUN 2018 KNH/UoN-ERC P.O. Box 20723-00202 NAIROBI

KNH-UON ERC
Email: uonknh_erc@uonbl.ac.ke
Website: http://www.erc.uonbl.ac.ke
Facebook: https://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC https://twitter.com/UONKNH_ERC

KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref: KNH-ERC/A/245

June 22, 2018

Beatrice Machini
Reg. No. W83/52076/2017
PhD Candidate
Institute of Tropical and Infectious Diseases (UNITID)
College of Health Sciences
University of Nairobi

Dear Beatrice,

RESEARCH PROPOSAL – STATISTICAL METHOD FOR CORRELATED DATA; APPLICATION TO SEVERE MALARIA CASE MANAGEMENT EVALUATION IN KENYA, 2018 (P233/04/2018)

This is to inform you that the KNH- UoN Ethics & Research Committee (KNH- UoN ERC) has reviewed and **approved** your above research proposal. The approval period is from 22nd June 2018 – 21st June 2019.

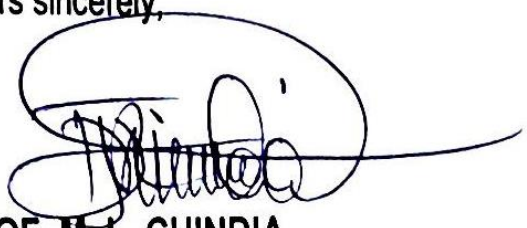
This approval is subject to compliance with the following requirements:

a) Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
b) All changes (amendments, deviations, violations etc) are submitted for review and approval by KNH-UoN ERC before implementation.
c) Death and life threatening problems and serious adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH-UoN ERC within 72 hours of notification.
d) Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
e) Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
f) Submission of an *executive summary* report within 90 days upon completion of the study. This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/ or plagiarism.

Protect to discover

For more details consult the KNH- UoN ERC website **http://www.erc.uonbi.ac.ke**

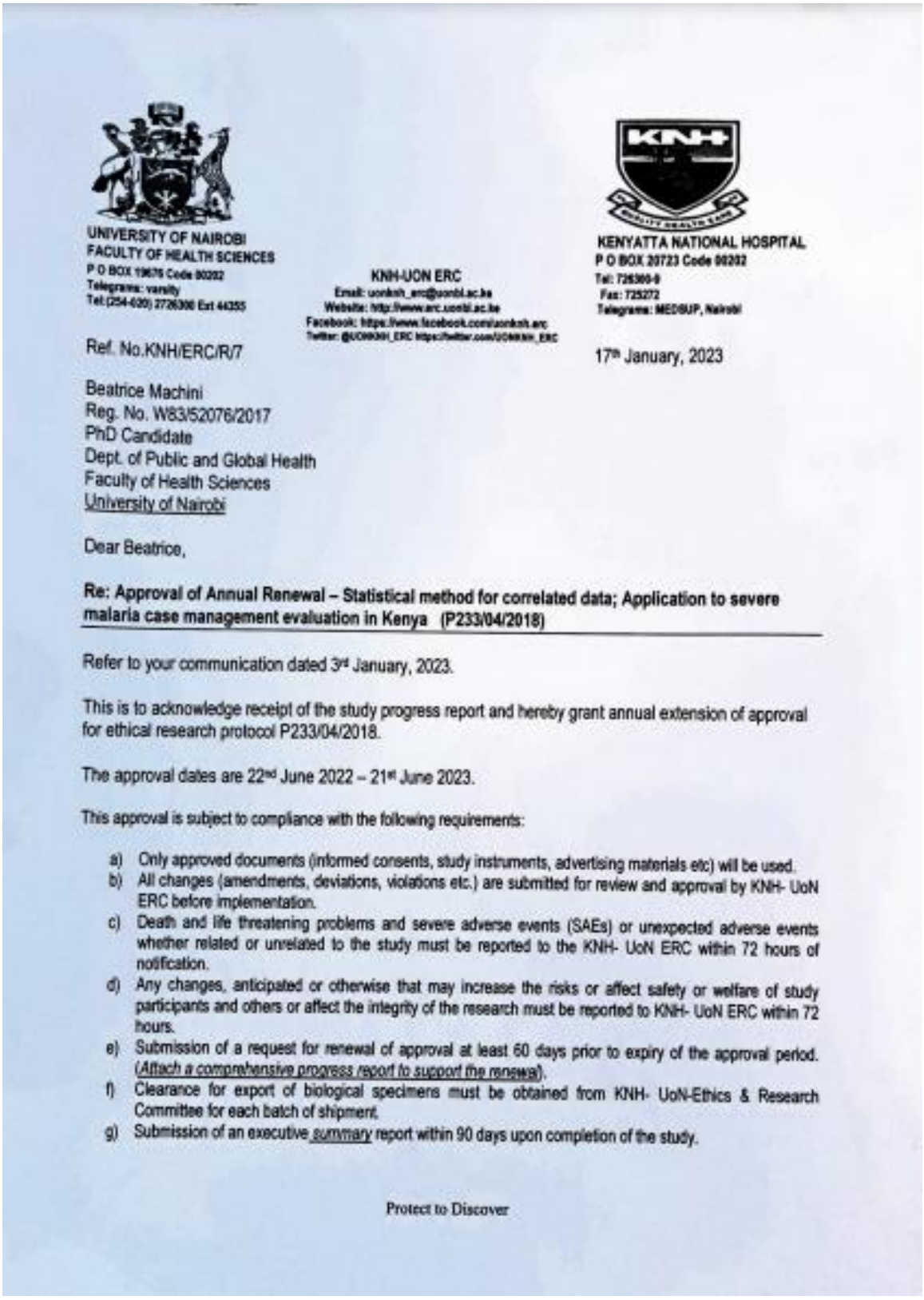
Yours sincerely,



PROF. M. L. CHINDIA
SECRETARY, KNH-UoN ERC

c.c. The Principal, College of Health Sciences, UoN
The Deputy Director, CS, KNH
The Chairperson, KNH-UON ERC
The Assistant Director, Health Information, KNH
The Director, Institute of Tropical and Infectious Diseases (UNITID), UoN
Supervisors: Prof. Thomas Achia, Dr. Hillary Kipruto, Prof. Dejan Zurovac

Appendix 3. Ethical renewal



UNIVERSITY OF NAIROBI
FACULTY OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KNH-UON ERC
Email: uonknh_erc@uonbi.ac.ke
Website: http://www.erc.uonbi.ac.ke
Facebook: https://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC https://twitter.com/UONKNH_ERC

KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref. No.KNH/ERC/R/7

17th January, 2023

Beatrice Machini
Reg. No. W83/52076/2017
PhD Candidate
Dept. of Public and Global Health
Faculty of Health Sciences
University of Nairobi

Dear Beatrice,

Re: Approval of Annual Renewal – Statistical method for correlated data; Application to severe malaria case management evaluation in Kenya (P233/04/2018)

Refer to your communication dated 3rd January, 2023.

This is to acknowledge receipt of the study progress report and hereby grant annual extension of approval for ethical research protocol P233/04/2018.

The approval dates are 22nd June 2022 – 21st June 2023.

This approval is subject to compliance with the following requirements:

a) Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
b) All changes (amendments, deviations, violations etc.) are submitted for review and approval by KNH- UoN ERC before implementation.
c) Death and life threatening problems and severe adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH- UoN ERC within 72 hours of notification.
d) Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
e) Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
f) Clearance for export of biological specimens must be obtained from KNH- UoN-Ethics & Research Committee for each batch of shipment.
g) Submission of an executive *summary* report within 90 days upon completion of the study.

Protect to Discover

This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/or plagiarism.

Yours sincerely,

DR. BEATRICE K.M. AMUGUNE
SECRETARY, KNH- UoN ERC

cc. The Dean, Faculty of Health Sciences, UoN
The Senior Director, Clinical Services, KNH
The Chair, KNH-UoN ERC

Protect to Discover