INTERRUPTED TIME SERIES ANALYSIS OF TRENDS IN PAEDIATRIC ADMISSIONS WITH DIARRHEA AND DEHYDRATION FOLLOWING INTRODUCTION OF ROUTINE ROTAVIRUS VACCINE IN THIRTEEN KENYAN HOSPITALS

DAISY CHELANGAT

A Thesis Submitted to the Board of Graduate Studies in Partial Fulfillment of the Requirement for the Conferment of the Degree of Master of Science in Applied Statistics of the University of Kabianga

UNIVERSITY OF KABIANGA

JANUARY, 2022

DECLARATION AND APPROVAL

Declaration

This thesis is my original work and has not been presented for the conferment of a degree or for the award of a diploma in this or any other university.

Signature……  ……Date: 17/01/2022

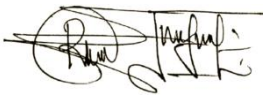
Daisy Chelangat

PGC/AST/004/17

Approval

This thesis has been submitted for examination with our approval as university supervisors.

Supervisors:

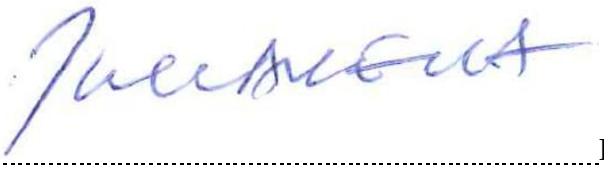


Sign……Date: 17/01/2022

Dr. Reuben Cheruiyot Lang'at

Department of Mathematics and Computer Science

University of Kabianga



Sign Date: 17/01/2022

Dr. Samuel Akech -KEMRI-Wellcome Trust

Health Services Unit

Kenya Medical Research Institute - Wellcome Trust Research Program

Sign Date: 17/01/2022

Dr. Lucas Malla -KEMRI-Wellcome Trust

Infectious Disease Epidemiology

London School of Hygiene and Tropical Medicine

COPYRIGHT

DEDICATION

I would like to dedicate this thesis to my siblings: Glorious Cherotich, Linet Chebet Emanuel Kipkorir and Owen Kipkoech. I also dedicate it to my lovely daughters; Angela and Favour.

AKNOWLEDGMENT

I would love to thank God for the blessing of life and a chance to do this work. Gratitude to Dr. Reuben Langat for his supervision, my parents: Margaret Chepkwony and Benard Chepkwony, my spouse: Mr. Andarson Koech and other family members who supported this work. Further, I also wish to appreciate the KEMRI Wellcome Trust/Initiative to develop African Leaders (IDeAL) for hosting me and funding the project; Dr Samuel Akech, Dr. Lucas Malla and the entire Clinical Information Network team for their supervision and contribution.

ABSTRACT

Diarrhea and dehydration has been reported to be among the top causes of hospitalization and mortality in children aged under 5 years. Most cases of diarrhea in childhood are caused by rotavirus and routine introduction of rotavirus vaccine has been promoted to potentially reduce incidence and severity of diarrhea and dehydration in vaccinated infants. I examined changes in admissions of all clinical cases of diarrhea and dehydration following introduction of routine vaccination with rotavirus vaccine in 2014 in Kenya. Previously studies have mostly examined changes in admissions with stools positive for rotavirus. This study assessed changes in admissions due to all-cause diarrhea and dehydration without considering whether the patient was tested for rotavirus or not.

This was a retrospective observational study that used data from 13 public hospitals currently involved in a clinical network (Clinical Information Network (CIN)) set up to ensure improved collection of routine data to improve inpatient care in Kenya. The hospitals were purposefully selected by the ministry of health to represent different geographical locations in Kenya. I included data for children aged 2-36 months, the age most vulnerable to rotavirus infection. Simulations were used to determine whether the sample size yielded enough power to detect changes in admissions to diarrhea and dehydration. I used interrupted time series analysis model following a negative binomial distribution to assess changes in the burden of diarrhea and dehydration. I used 3 pairs of Fourier terms to account for seasonality of infectious diarrhea admissions. Non febrile admissions (surgical or burns) were used as controls There were 29,231 patients who were classified to be having diarrhoea as well as dehydration. The average DAD admissions per month before the vaccine was introduced (July 2014) was 35 (standard deviation (SD): ±22) and 17 (SD: ±12) after vaccine introduction. Fitting a Segmented regression analysis model revealed a 28.32% (95% C.I, 0.786 to 0.950) decrease in hospital admissions immediately after July 2014 when the vaccine was introduced to the Kenya routine childhood immunization program. This was followed by a 3.00% (95% C.I, 0.786 to 0.950) decrease in month to month hospital admissions due to all-cause diarrhea and dehydration after vaccine introduction. There was statistically significant change in admissions from non-febrile admissions before and after vaccine introduction. In conclusion, the introduction of the rotavirus vaccine resulted in a reduction in public hospital admissions because of all-cause diarrhoea and dehydration. It is therefore recommended that continuous monitoring be done to ensure that its performance over time is known.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVEATIONS AND ACRONYMS

AR- Autoregressive

ARMA- Auto Regressive Moving Average

CIN – Clinical Information Network

DAD- Diarrhea and Dehydration

EDA- Exploratory Data Analysis

FSC – Fully Conditional Specification

ITS – Interrupted Time Series

KEMRI – Kenya Medical Research Institute

MOH – Ministry of Health

MAR – Missing At Random

MNAR- Missing Not At Random

MCAR- Missing Completely At Random

PAR – Pediatric Admission Form

WHO – World Health Organization

DEFINITION OF TERMS

Diarrhea: Passing loose stool more than three times in a day leading to loss of fluids in the body.

Dehydration: When the body loses more water than that being taken in by the patient.

Rotavirus: Rotavirus is an infectious virus that causes inflammation in the gastrointestinal linings with some of the symptoms being vomiting, watery diarrhea, high fever and abdominal pains.

Rotavirus Vaccine: Vaccine that protects against rotavirus

Missingness: Missingness in clinical data arises when a clinician misses to record some symptoms of a patient I the record sheets provided to then.

Multiple imputation: Substitution of missing data with plausible values under a specific missing data mechanism

Interrupted Time Series analysis: Time series analysis is used to evaluate sequential datapoints over time. Interrupted time series analysis involves the interruption of the analysis period by a significant phenomenon that could have caused change.

CHAPTER ONE

INTRODUCTION

1.1 Overview

This section gives the background of our study, problem statement, objectives, hypothesis, justification, and significance of the study.

1.2 Background of the Study

World Health Organization defines diarrhea as passing three or more loose stools in one day. Severity increases with the frequency and volume of stools. Dehydration results when loss of fluid exceeds intake or replacement. Diarrhea may be caused by infection of the gastrointestinal lining by various varieties of bacteria, viruses or parasites but infection with rotavirus predominates in early childhood (Kirk, Angulo, Havelaar, & Black, 2017).

Globally, approximately 1.7 billion cases of diarrhea are reported every year among children aged less than five years (Heaton & Ciarlet, 2007). The Kenya Demographic and Health Survey (KDHS) in 2014, categorized diarrhea as the second leading cause of death in under-fives in Kenya and is among the top causes of death in children especially in the sub-Saharan African countries. Fifteen percent of the children who took part in the survey had a history of diarrhea two weeks just before the survey started and 58 % of these cases were reported to a clinician for treatment. In addition, diarrhea is a leading cause of malnutrition and is among the top causes of death in children especially in the sub-Saharan African countries.

Exclusive breastfeeding, hand and food hygiene, clean water and sanitation, vaccination are measures recommended by WHO for reducing diarrhea and dehydration (Kirk *et al*., 2017; Schwartz *et al*., 2019) that have had huge impact globally but despite progress, diarrhea in children still remains a significant burden amongst children in Kenya (Health, 2007).

1.2.1 Rota virus and Rotavirus Vaccine

Rotavirus is an infectious virus that causes inflammation in the gastrointestinal linings with some of the symptoms being vomiting, watery diarrhea, high fever and abdominal pains. The virus is spread through ingestion of contaminated food or contaminated fingers. It is a major cause of severe diarrhea in children between ages from 2 to 36 months. Rotavirus vaccine was introduced to routine vaccination to decrease incidences of severe diarrhea and death due to diarrhea and dehydration (Wandera *et al*., 2017). It is administered orally to children at age of six and ten weeks and was implemented as part of routine Kenya Expanded Immunization Program (EPI) in July 2014.

Studies in many parts of the world that have implemented the vaccine into their routine immunization programs have shown it to be effective using data monitoring cases of rotavirus positive cases determined from stool sampling. A 2016 study in Rwanda, which was the first low income African country to implement the vaccine, showed that admissions to hospitals for rotavirus after the introduction of the vaccine had decreased substantially (61-70%) (Ngabo *et al*., 2016). Raes and his team in 2013 also found out that even though there were strong seasonal trend in hospital admissions due to diarrhea, there were fewer hospital admissions after the vaccine was introduced. A recent interrupted time series analysis on how the vaccine had impacted admissions to two Kenyan hospitals in a period spanning three years after its introduction showed a significant decrease in rotavirus positive admissions (Raes *et al*., 2016). The study involved 3,165 children aged between 2 and 59 months with their stool samples used to determine their rotavirus status.

1.2.2 Time Series Analysis

Time series analysis is used to evaluate sequential datapoints over time. Interrupted time series analysis involves the interruption of the analysis period by a significant phenomenon that could have caused change (López Bernal, 2018).

As an example, trends in hospitalizations due to a certain disease could be interrupted by introduction of a vaccine expected to affect the disease.

1.3 Statement of the Problem

The efficacy of a vaccine is tested in randomized controlled trial (RCT). However, a randomized trial has strict inclusion and exclusion criteria (López Bernal, 2018). It is thus important to examine the effectiveness in routine care. The rotavirus vaccine, introduced in Kenya Expanded Immunization Program (EPI) in July 2014, has been shown to reduce rotavirus positive admissions. This has been shown in randomized controlled trials that uses lab stool samples to test for rotavirus in a patient (Otieno *et al*., 2020). However, its impact on all cause diarrhea and dehydration admissions in children has not been shown. In this study I use routinely collected data to assess the changes in all cause diarrhea admissions following introduction of the vaccine. The use of routinely collected data is a cost-effective tool for assessment of impact of introduction of public health interventions (Tuti *et al*., 2016).

1.4 General Objective

To investigate trends in hospital admissions due to all-cause diarrhea following the introduction of rotavirus vaccine in Kenya

1.5 Specific Objectives

This study was guided by the following objectives:

i. To investigate the effect of introduction of rotavirus vaccine on all cause severe diarrhea admissions using interrupted time series models.

ii. To investigate the nature of missingness of routine data used for the analysis of impact of rotavirus vaccine on diarrhea admissions collected from 13 hospitals participating in the clinical information network

iii. To set simulations for power calculations for interrupted time series designs.

1.6 Research Hypotheses

The following null hypothesis were tested:

H_0: There is no decrease in hospital admissions due to diarrhea and dehydration immediately and following the introduction of the rotavirus vaccine.

1.7 Justification of the Study

The burden of diarrhea and dehydration is still high in Kenya and regionally remains among the top causes of death in children under five years of age despite many interventions being put into place (Kirk *et al*., 2017). This study, which uses interrupted time series analysis to evaluate changes in hospitalization due to diarrhea and dehydration (DAD), was necessary to assess the overall impact of introduction of the rotavirus vaccine on all admissions with DAD. This study will also highlight the utility of interrupted time series analysis for the evaluation of impact of interventions that are administered at a population level.

1.8 Significance of the Study

This study analyses the current trend in hospitalization due to diarrhea and dehydration and may guide appropriate measures to be implemented by the Ministry of Health in Kenya and the research community towards lessening incidence and fatality as a result of diarrhea in children.

Furthermore, the research could be replicated in other countries and at future times to have more information on the performance of interventions especially the rotavirus vaccine.

In addition, since the study is not specific to laboratory tested positive rotavirus cases, it gives the bigger picture of changes in hospitalizations due to DAD. This study also adds to the benefits of using routinely collected data which is cost effective

1.9 Scope of the Study

This study uses routinely collected hospital data belonging to children aged between 1 to 36 months admitted with a history of diarrhea to inspect the trends of hospital admissions due to diarrhea and dehydration. Pediatric treatment protocol is used to classify patients as having diarrhea and dehydration.

1.10 Limitations of the Study

There is however an analytical challenge of missing data in the clinical records which is handled by use of multiple imputation.

1.11 Assumptions of the Study

This study assumed that all the data recorded were correct as indicated by the clinician and that no manipulation was made whatsoever.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter will give a brief overview of the previous studies done on the research topic and methodological approaches that researchers have used in addressing the trends in the impact of the rotavirus vaccine with a focus on interrupted time series. Furthermore, I shall highlight the gaps in these studies, which will be addressed in the analysis.

2.2 Theoretical Framework

The Rota virus vaccine, having been first licensed and used in the United States of America (Heaton & Ciarlet, 2007), has been shown to have commendable results in reducing rotavirus infections in both developed and developing countries. A 2014 report by the World Health Organization (WHO) showed that majority of the countries that have implemented the vaccine in their vaccination programs have witnessed fewer deaths due to diarrhea and dehydration. Currently, two types of the vaccine (Rotarix® and Rotateq®) are available globally which are administered to children from 6 weeks of age (Folorunso & Sebolai, 2020).

Children obtain natural immunity from the rotavirus vaccine at the age of 36 months. This has been shown by a study conducted in Guinea-Bissau evaluating immunity in children against rotavirus after receiving vaccination. The study involved weekly collection of stool from children and retrospectively following the infants for two years after which the data was analyzed (Akech *et al*., 2018). As the ages increased, the symptoms were becoming asymptomatic implying natural immunity from the virus.

Globally, there has been a 59% (interquartile range: 46-74) decrease in rotavirus positive hospitalization and a 36% (Interquartile range: 28-46) reduction in deaths. This is according to a systematic review seeking to evaluate how the vaccine have been performing worldwide since it was licensed up to 2019. The review used 105 publications involving 49 countries and the admissions belonged to children of less than 5 years of age (Rahajamanana *et al.*, 2018).

Among the sub-Saharan countries, only 32 out of 47 WHO member countries had implemented the vaccine in their national immunization programs by the year 2017 (Rahajamanana *et al*., 2018). According to WHO, by 2010, 42% of hospitalizations of children aged less than 5 years, were due to diarrhea and dehydration (Otieno *et al*., 2019). These cases were reduced after the introduction of the vaccine by 61 – 67%.

A recent systematic review evaluating the effectiveness of the rotavirus vaccine in different settings post its licensing in 2006, found the median effectiveness in 24 selected low, high and medium child mortality countries to be 84%, 57% and 75% respectively. According to this study, the effectiveness of the vaccine reduced at the age of 2 and this was more evident in medium to high mortality countries. The results were based on rotavirus lab test positive cases spanning a decade since the vaccine was licensed (Jonesteller, Burnett, Yen, Tate, & Parashar, 2017).

(Khagayi *et al.*, 2020) in their study also found the vaccine to have been able to reduce hospitalization related hto rotavirus in children. In the study, stool was taken from children admitted to three Kenyan hospitals and tested for rotavirus after confirmation that they had history of rotavirus vaccination. The study included 677 children as the treatment group and 567 as controls and the overall vaccine effectiveness was found to be 64% (95% confidence interval: 35% to 80%).

There were also varying effectiveness for different age groups: 67% for children aged more than 12 months and 72% for those whose age was more than 12 months (Khagayi *et al*., 2020).

Rotavirus negative cases he been used as controls and effectiveness checked basing adjusted odds ratios after fitting multivariate regression models (Banajeh & Abu-Asba, 2015). The model could not however deal with confounding factors that could have led to the reduction in hospitalizations and deaths. Furthermore, the trend in this rotavirus admissions could have already been reducing due to other measures that were in place at that time which the model could not capture.

A Kenyan study involving two county hospitals showed a significant reduction in hospitalization due to rotavirus positive cases. The study reported a 48% decline in hospital admissions in relation to Rotavirus infections (Wandera *et al.*, 2018). However, rotavirus still remain a significant burden despite efforts to reduce its effect in children under the age of five years in Kenya (Gikonyo *et al*., 2019). Routine data from health facilities in admissions with diarrhea and dehydration are lacking and most studies reporting on impact of introduction of rotavirus vaccine are from research centers conducting surveillance of rotavirus positive cases. Routine data used in this study allow for generalization of findings and assessment of wider impact.

2.3 Conceptual Framework

Variety of methodological approaches for assessing the impact of interventions exists and have been used widely by researchers. Randomized controlled trials (RCTs), whenever feasible, have been shown to yield inferences that are less biased (López Bernal, 2018). They are, however, sometimes expensive and might have limited generalizability due to restricted inclusion and exclusion criteria.

Furthermore, randomized controlled trials might sometimes be unethical or infeasible which calls for the use of observational studies utilizing quasi experimental designs.

Quasi experimental designs are used to study the causal impact of an intervention to the population of target without randomly assigning participants to the study. The designs, just like traditional experimental design have the quasi dependent and quasi-independent variables. The researcher assigns participants into the study by using a set criterion such as age. Quasi experimental types include pre-posttest, regression discontinuity, panel analysis, nonequivalent control group and case control designs, and interrupted time series design (López Bernal, 2018).

Interrupted time series (ITS) type of quasi experimental design fits regression equation models into data belonging to different periods interrupted by a significant event. The model fitted id as shown in equation 2.1.

$$y = \beta_0 + \beta_1(T) + \beta_2(X_t) + \beta_3(TX_t) + \epsilon ij \qquad (2.1)$$

where

y Represents the dependent variable

T: Time that has passed since a study begun.

X_t : dummy variable representing the pre and post intervention periods

β_0: baseline at the beginning of the study, $T = 0$

β_1: is the change in response when there is a unit increase in time.

β_2: change in trend level after the intervention

β_3 change in slope after the intervention as a result of the interaction between the intervention and time elapsed TX_t

ϵ_{ij} Is the error term

Several studies have used ITS to evaluate the impact of interventions especially the impact of rotavirus vaccine. A study conducted in 2017 which aimed to inspect how the rotavirus vaccine had impacted hospitalization of children less than 5 years of age in Madagascar, compared the admissions pre and post vaccination (Rahajamanana *et al*., 2018). The study was retrospective and analyzed data from 2010 to 2016 with the pre-vaccination period being from January 2010 to December 2015 and the post- vaccination period from January 2014 to December 2016. They used simple before and after analysis to do their comparison. This methodology, however, did not address matters concerning internal and external validity which includes seasonality, autocorrelation among others. Other sources of bias such as history was not also addressed.

Another study conducted in Yemen, also wishing to evaluate vaccine performance before and after its introduction using surveillance data from a hospital used simple before-after analysis to do the comparison. This study did not also capture any methodological issues that could have limited their study such as seasonality, autocorrelation and other sources of bias encountered when handling routine data (Banajeh & Abu-Asba, 2015).

They analyzed data belonging to 5161 children aged less than five years and admitted due to acute gastroenteritis over a period of 7 years (2007-2014). The pre-intervention period was between 2007 and 2011 and the post-intervention period between 2013 and 2014. Comparisons were done using chi-square tests with the odds ratios generated using generalized linear models.

2.4 Identification of Knowledge Gap

There is little literature on the trends of overall hospital admissions due to diarrhea and dehydration after rotavirus vaccine introduction in Kenya using controlled interrupted time series.

Moreover, there are a few studies addressing the performance of the vaccine using methodological approaches that take into consideration possible sources of bias such has history bias, autocorrelation, and seasonality. I used a combination of four words: rotavirus, vaccine, admission and effect to search through several journals and most studies reported only on rotavirus positive cases while our study examines on all admission with DAD from routinely collected data where rotavirus stool results are not available. In this study I used data collected from a larger number of hospitals distributed across Kenya as compared to other that have used at most two hospitals.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the study design used, scope of the study, the study population, procedures for data collection and presentation, and sample size/power calculation. It also describes methodologies for dealing with missing data and interrupted time series analysis.

3.2 Research Design

The study was a retrospective study using data obtained from routine clinical records from 13 public hospitals in Kenya that were purposefully selected to participate in a clinical information network.

3.3 Location of the Study

Data from inpatient pediatric wards collected from 2013 to 2019 were included in the study. There were no admissions to the hospitals due to health worker strikes for eight months between December 2016 to march 2017 and June 2017 to November 2017, therefore, data for this period was excluded from the analysis (Muendo *et al*., 2018). The study period was divided into pre- and post-intervention periods with the intervention period being from July 2014 to December 2019. Time points before the intervention period were categorized as pre-intervention period while those after July 2014 as post-intervention period. Data was aggregated by months.

3.4 Target Population

Previous studies indicate that after 36 months of age most children survivors obtain natural acquired immunity from rotavirus infection even if they had not been immunized (Fischer *et al.*, 2002).

The study population comprised of children between the age of 2 and 36 months admitted into hospital with a history or clinician diagnosis of diarrhea and dehydration following the protocol given by the Ministry of Health.

3.5 Sample Size and Sampling Procedures

I conducted power simulations to determine if there was sufficient power to detect any significant differences. The following assumptions for the simulations were made:

The number of hospital admissions does not change during the period before vaccine introduction and has constant reduction after vaccine introduction as shown in Figure 3.1

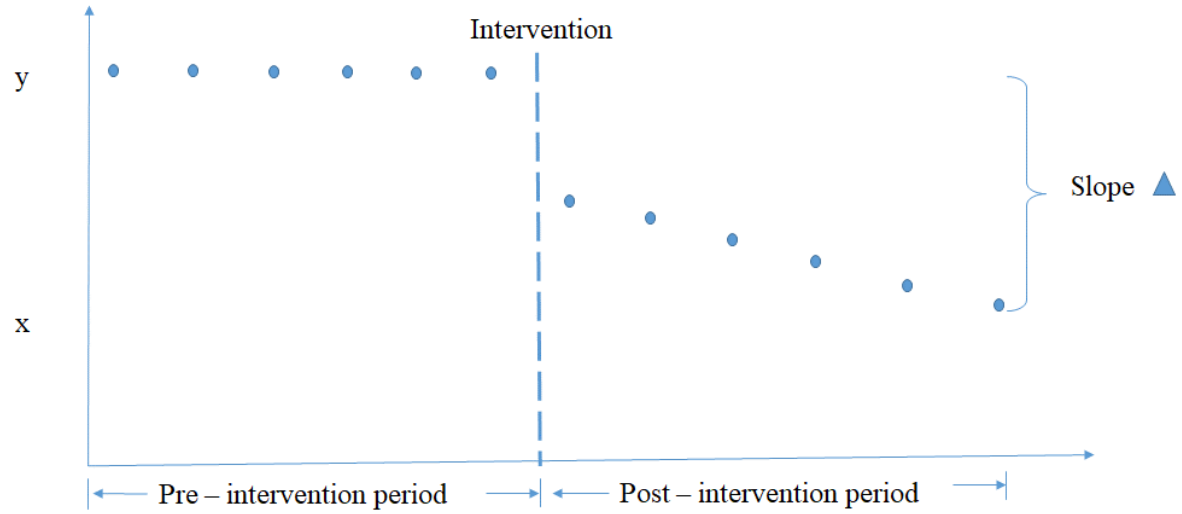


Figure 3.1: Impact model

i. The pre and post intervention periods were allowed to vary.

ii. Segmented autoregressive (AR (1)) model of order 1 was fitted with the response (hospital admissions) following a negative binomial distribution.

iii. Sample size per month was allowed to vary in relation to different effect sizes to obtain a sample size that would yield a power of approximately 80%.

iv. Allowed variation of the effect size in such a way that the estimates of event rates at baseline would decrease by a half, a quarter or an eighth by end of study.

v. Performed simulation for 1000 datasets.

3.6 Data Collection Procedures

Data used in this study was retrieved from the Clinical Information Network (CIN), a collaboration to improve collection of routine medical data for improvement of quality of care provided in admitted children (Irimu *et al*., 2018). The network is a partnership of several hospitals, KEMRI-Wellcome Trust Research Program (KWTRP), the Ministry of Health (MoH) and Kenya Pediatric Association. It is composed of hospitals from 12 counties which were selected purposefully by MoH (Ayieko *et al*., 2016; Tuti *et al*., 2016) and fourteen hospitals had joined the network by 2019.

Standardized pediatric admission record (PAR) forms are used to capture the patient's demographic and clinical details during admission and discharge summary forms captures all the patient's discharge details including diagnosis and whether they are alive or dead. These forms, which are structured according to the basic pediatric treatment protocols provided by MoH are filed together with laboratory reports and other notes as recorded by the clinician and form part of patients' medical notes. The partner hospitals had agreed to adopt these forms as part of their routine medical records.

3.6.1 Validity of the Instruments

The clerks synchronize data to the KEMRI-Wellcome trust server every day after running it through a data cleaning code written in R programming language. Once at the server, the data is assessed for quality and validity in terms of whether it is complete and within the accepTable ranges, for example patient temperature cannot be recorded as 100 degrees Celsius. This is done by the data management team who also run the data cleaning code on the data for the second time. Where there are errors, clerks are contacted by phone and asked to make corrections where possible and thereafter resubmit the data. Where necessary, clerks are asked to send pictures of the forms to aid in verification and correction of errors.

3.6.2 Reliability of the Instruments

To ensure reliability of the collected data, the data team audits the data every three months by visiting the hospitals and randomly picking used files, entering the data, and marching it with the previously sent data.

3.6.3 Missing Values

The data being obtained from routine hospital admission records are often plagued with missing data, that if not addressed adequately might bias regression estimates (Nicholls, Langan, & Benchimol, 2017).

As mentioned earlier, clinicians are provided with pediatric admissions record (PAR) forms where they record the patients' history and diagnosis at admission. During discharge, diagnosis is recorded in the discharge summary forms with the treatment history being recorded in the pediatric treatment chart. Missing data arises when a clinician misses to record some symptoms

as indicated in the forms with reference to the pediatric treatment protocol or in some cases, the symptoms such as temperature are not measured due to device failures or unavailability.

Missing data is common in most research and is always necessary to handle. I use routinely collected data from medical records to classify patients with diarrhea and dehydration using clinical signs outlined in the pediatric treatment protocol (Health, 2007). Missing data in any of these clinical signs will make it impossible to do the classification.

3.6.4 Mechanisms of Missing Data

Missing data mechanism guides how missing data should be handled. These mechanisms are Missing Completely at Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR).

Data is termed MCAR when the missingness probability is not related to data that is either observed or unobserved for example when a patient enrolled in a study relocates to another city and is therefore no longer available for follow up (Carpenter & Kenward, 2012). MCAR type of missing data can be handled using list wise deletion or pairwise deletion without the risk of obtaining biased results.

MAR data are those whose probability of missingness given observed data does not depend on missing data. For example, a patient with no history of diarrhea or vomiting is less likely to be dehydrated. This kind of missing data can be handled using multiple imputation, expectation-maximization algorithm among others.

Finally, data is termed as MNAR when the missingness probability depends on the missing observation itself. It is neither missing at random nor missing completely at random, for example, if a variable recording history of diarrhea in patients has missing values for those patients that had diarrhea and it is not known why they are missing.

3.6.5 Multiple Imputation

Multiple imputation is an approach to dealing with missing observations that entails creation of multiple datasets with plausible imputed values then combining the estimates using a set of rules (Enders, Keller, & Levy, 2018; Little & Rubin, 2019). It is impossible to know the exact value of a missing observation and therefore, multiple imputation accounts for uncertainties as it predicts missing values by allowing some variability into the multiply imputed observations (Carpenter & Kenward, 2012; Van Buuren, 2011).

Multiple imputation was not commonly used until recently when many researchers in the epidemiological field resorted to this method as reported in a systematic review by Penteha (2015).

The review, published in 2015, identified 103 papers that had used multiple imputation in handling missing data compared to 2008 when only 11 publications, and 26 by 2013, had used this method (Rezvan, Lee, & Simpson, 2015).

When the proportion of missing data is less than 30% and the Missing at random assumption has been taken, multiple imputation produces unbiased results (Rezvan *et al.*, 2015). Other methods of handling missing data such as piecewise and pairwise deletion are used when the missingness pattern is not MAR.

The general algorithm for multiple imputation is that, given a matrix of data say X, then denote observed values with X_0 and missing values as X_m.

Multiple imputation takes into consideration the distribution of the unobserved conditioned by the observed values then draws several times from the distribution which results into say M 'complete' sets of data. Then:

i. Let $m = 1,...,M$ then, imputation is done on missing data from the distribution of the unobserved data given the observed data $f(X_m|X_o)$ and taking into account possible uncertainty to obtain M datasets of complete data.

ii. A substantive model is formed and fitted to the M data sets, $m = 1,...,M$. This yields M estimates for the model parameters say, β_m and Var(β_m) estimates for variance.

iii. To make inferences, the imputations are combined following Rubin's rules (Little & Rubin, 2019) .

3.6.6 Rubin's Rules

Scalar element of interest was denoted by β which is associated with variance σ^2. Then, letting the m imputed data sets obtained after fitting the model and treating them the same way as we would in the absence of missingness be, β_m ,σ^2_m. The multiple imputation estimator for β (β_{mi}) is:

$$\hat{\beta}_{mi} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_m \tag{3.1}$$

The variance estimator is given by:

$$\hat{V}_{mi} = \widehat{W} + (1 + \frac{1}{M})\hat{\beta} \tag{3.2}$$

And the pooled variance constitutes of variability within ($\widehat{\boldsymbol{W}}$) and between imputed datasets ($\widehat{\boldsymbol{\beta}}$)

Where

$$\hat{W} = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma_m^2} \tag{3.3}$$

And

$$\hat{\beta} = \frac{1}{M-1}\sum_{m-1}^{M}(\hat{\beta}_m - \hat{\beta}_{mi}) \tag{3.4}$$

then test the null hypothesis that the actual β is equal to the estimated β^0

and T is calculated as:

$$T = \frac{\hat{\beta}_{mi} - \beta^0}{\sqrt{\hat{V}_{mi}}} \quad (3.5)$$

Which is compared to a *t* tabulated with *v* degrees of freedom calculated as:

$$v = (M-1)[1 + \frac{\hat{W}}{(1 + \frac{1}{M}\hat{B}}]^2 \quad (3.6)$$

3.6.7 Fully Conditional Specification (FSC) Imputation for Data at Two Levels

A variety of methods exists for multiple imputation and this includes multivariate normal imputation and fully conditional specification imputation. The multivariate normal imputation method assumes that the variables follows a multivariate normal distribution while the FSC method is more flexible as it allows the user to conditionally choose the distribution of the variables. The FSC is therefore, due to its flexibility more preferable to the multivariate normal imputation method (Rubin, 1988).

The multilevel FSC framework does imputation for each variable one at a time. A multilevel model is given by:

$$Y_{1ij} = \beta_0 + \beta_1 Y_{2ij} + \beta_2 Y_{3ij} + u_{0j} + u_{1j} + u_{1j} Y_{2ij} + \epsilon_{ij} \quad (3.7)$$

where

Y_{1ij} - outcome value for the i^{th} observation and the j^{th}cluster.

Y_{2ij},Y_{3ij} - predictors at level-1 , is denoted by β_0- intercept

β_1 andβ_2 - slope coefficients for Y_1,Y_2 respectively.

μ_{0j} -residual for Variation in the responses between clusters u_{1j}- variation of Y_2 across clusters.

ij-Unexplained Level-1 variation within clusters.

ϵ_{ij} is the error term

The algorithm draws imputation values from: formerly imputed data sets that are complete, univariate distributions conditioned on model parameters at multilevel basis and residual terms at level-2 (Mistler & Enders, 2017). It takes the assumption that level-1 variables are normally distributed. During the first imputation step, Y_1 is treated as a response and Y_2 and Y_3 as explanatory variables to predict Y_1. Y_1 imputations are generated from the resultant residual terms and parameter values defined from a normal distribution. Next, Y_2 is treated as the response and Y_1 and Y_3 as the explanatory variables. This process continues until all the variables with missing values have been imputed in a single iteration say t to form one complete data set(Van Buuren, 2011). M imputations are carried out to obtain M complete datasets as described earlier.

The number of imputations M is chosen in such a way that the error in p-value estimation is satisfactorily small and the inferences are precise after a few imputations.

Rubin suggested an average of 3 to 10 imputations to be sufficient however, if the inferences are not clear-cut then more imputations will be required (Rubin, 1988). The M data sets will then be analyzed and combined using Rubin's rules as stated earlier.

3.6.8. Handling Missing Data in Addressing the Thesis Objective

Missingness in the data is handled using multilevel multiple imputation with chained equations which considers the fact that the data is hierarchical clustered in different hospital (Van Buuren, 2011).

Since contexts between hospitals are different, multilevel structure with individual admissions at level 1 who belongs to different hospitals at level 2 was used. Data to be imputed was obtained using the selection criteria and different variables were used to diagnose identify children with DAD. These variables include history of diarrhea, pulse rate, cap refill, skin temperature, sunken eyes, skin pinch, alertness, ability to drink, age in months and whether a patient is vomiting. These variables were included in the imputation model as level one variables and hospital as level 2 variable.

3.6.9 Justification for Multilevel Multiple Imputation

As earlier stated, the data comprised observations from children admitted to 13 different hospitals and 5.1% of the study participants had at least one missing variable. I reasoned that since hospitals are in different in context (local practices, geographical locations) this may contribute to differences in missingness across hospitals. There are *adhoc* methods for dealing with missing data and this includes listwise deletion, last observed observation carried forward and mean imputation. These methods works under assumption of Missing Completely At Random (MCAR) and produce biased inferences when this assumption is violated (Buhi, Goodson, & Neilands, 2008).

Other methods which can be considered are substitution where missing values are substituted by values of individuals who were not initially selected into the sample, hot deck imputation where values for imputation are selected at random from observations with complete data, cold deck imputation where values are chosen systematically from individuals with like values on other variables, regression imputation where missing values are predicted by regressing on other

variables, stochastic regression imputation and interpolation and extrapolation imputation (Carpenter & Kenward, 2012).

There are two types of imputation, multiple and single; single imputation involves obtaining one estimate of the unobserved data by making use of any of the methods available for dealing with missing data. This method can yield biased results when it comes to estimating parameters especially when the missingness mechanism is either MAR or MNAR and when the proportion of missingness is high. Furthermore, single imputation gives an under estimation of standard errors and as a result one may obtain p-values that are too small (Tang, Song, Belin, & Unützer, 2005; Waljee *et al*., 2013).

Multiple imputation yields more estimates which in return, leads to less bias, improved validity, increased precision and more robust results which are less affected by outliers (Schwartz *et al*., 2019). In this study, multiple imputation was used for dealing with missing values and was implemented using MICE package in R statistical software version 4.0.0 (Van Buuren & Groothuis-Oudshoorn, 2011). To check for model (model 3.7) convergence, plots of observed verses imputed datasets were plotted to check their distributions.

3.7 Data Collection Procedures

Data is collected soon after the patient is discharged by abstracting data from the medical records into a dedicated database built into Research Electronic Data capture (REDcap) platform, an open source platform for capturing data (Ihaka & Gentleman, 1996). There is a dedicated clerk in each hospital who does data entry(Ayieko *et al*., 2016; Tuti *et al*., 2016).

Two categories of datasets are captured: minimum dataset and full dataset. Minimum datasets are collected as required for reporting by the routine ministry of health's health management information system (HMIS) and consists of the demographic information of the patient, final diagnosis, and outcome (dead/alive). Full dataset consists of specific disease symptoms as recorded by the clinician in the pediatric admission and discharge summary forms. Minimum datasets are collected when the study clerks are on leave otherwise clerks enter full datasets during other times.

3.8 Data Analysis and Presentation

Observations were classified according to severity of diarrhea and dehydration following the pediatric treatment protocol and then aggregated monthly since the mean stay of children below 36 months of age was estimated to be less than one month.

The resulting data set was then converted into a time series thereafter plotting trends and checking for seasonality and auto-correlation. Mean hospital admissions pre- and post-intervention were also obtained as part of EDA.

3.8.1 Interrupted Time Series Analysis (ITS)

Interrupted time series analysis was used to evaluate the effect of the Rotavirus vaccine on hospital admissions due to diarrhea and dehydration.

Interrupted time series (ITS) design is a powerful quasi experimental design especially in assessing the impact of interventions at a population level. Observations are taken on a subject over time before and after an intervention. The interruption period is that period when an intervention is introduced. Intervention impact is thereafter determined by examining the change in trend patterns after the intervention.

A major difference between ITS studies and simple analysis of outcome of interest before and after an intervention is its ability to model trends during the two periods.

This enables ITS analysis to display more in-depth results that would better inform decision making. The design is also able to deal with internal validity, a major threat to the validity of observational studies, by having a sequence of observations over time before an intervention. This is necessary for maturation and regression effects to be detected.

History bias, caused by other events that could possibly take place during the time when the intervention is introduced and affect the outcome of interest, is a major threat to the validity of ITS designs. This is however overcome when there is a short time span between time points and short lags before intervention effects are seen.

Moreover, the design makes use of the pre-intervention period observations as the control group in a situation where it is impossible to have a control group thus reducing threats to external validity. When a control group is available, the analysis can control for the confounding factors (Lerman, 1980).

Selecting a control group for ITS studies can sometimes be difficult because the design might involve the whole population. A variety of options to be used in selecting the controls exists which are location-based control group, characteristic based, behavior based, historical cohort, control outcome and control time period. This control groups has to be selected carefully so as not to bias results.

ITS analysis requires the time series data to have a clear pre and post-intervention periods. The data comprised of hospital admissions from 2013 to 2019 with the vaccine introduced in July 2014. This met the first requirement to perform ITS analysis. Secondly, the outcomes were counts of hospital admissions aggregated monthly across the years.

Total monthly hospital admissions were used as an offset to account for population changes because different hospitals joined the CIN at different times between October 2013 and March 2014. The control group, used in sensitivity analysis, comprised of patients admitted due to surgical burns.

I fitted a segmented regression analysis model following a negative binomial distribution, which involves partitioning the independent variable into different intervals (segments) then regression lines fitted to each segment separately. The segments in the data were the pre and post vaccination periods with vaccine introduction acting as the breakpoint (Waljee et al., 2013).
The response variable represented counts of diarrhea and dehydration admissions per month. Equation 3.8shows the regression model used.

$$y = \beta_0 + \beta_1(T) + \beta_2(X_t) + \beta_3(TX_t) + \beta_2(X_t) + \epsilon ij \quad (3.8)$$

where

y Represents hospital admissions due to diarrhea and dehydration

T: time elapsed since the beginning of the study.

X_t : dummy variable representing the pre and post intervention periods coded as 0 and 1 respectively.

X_c: dummy variable coded 0 for control group and 1 for treatment group

β_0: baseline at the beginning of the study, $T = 0$

β_1: is the change in response when there is a unit increase in time.

β_2: change in trend level after the intervention

β_3 change in slope after the intervention as a result of the interaction between the intervention and time elapsed TX_t

ϵ_{ij} Is the error term

3.8.2 Autoregressive Moving Average (ARMA (p, q))

Autoregressive moving averages of order 1 (ARMA) were also fitted using model 3.8 and the order of p (the autoregressive function) and q (the moving average function) was chosen by means the Akaike's Information Criterion (AIC), a criterion that estimates the quality of a model used with respect to the others.

A model with the lowest AIC was therefore chosen as it implied that the lowest amount of information was lost by that model.

3.8.3 Seasonality

The data, having been split into total monthly hospital admissions, exhibited seasonality and infectious diarrhea is expected to be seasonal. Cosine and sine time functions, (Fourier terms) were used to adjust for the seasonality.

3.8.4 Model Diagnostics

To assess the fit of the model selected using AIC, plots of residuals were generated. These were basically meant to check if the negative binomial regression assumptions had not been violated. These assumptions include:

i. As a generalized linear model, it is assumed that a linear relationship exists between model paramaters

ii. No serial correlation - Autocorrelation and partial autocorrelation plots were also generated to check for serial correlation in the residuals

iii. Observations are independent

iv. Conditional variance is greater than the conditional mean

3.8.5 Sensitivity Analysis

To determine the consistency of results obtained from the analyses, several sensitivity analyses were conducted. Since multiple imputation was based on the Missing at Random assumption, the possibility of Missing not at random was explored under the Missing Not At Random Assumption (MNAR). In deriving the missingness patterns, I first inspected change in hospital admissions due to DAD using imputed data but with the restriction of cases with any one of the variables being missing. Secondly, I considered a case where any two variables were missing. The third scenario took cases where any three variables under consideration were missing and finally case where any five variables were missing.

Further, change in hospital admissions due to burns using data from the same clinical information network was inspected to determine if change in DAD admissions had any other confounding factor different from the rotavirus vaccine. The data belonged to children aged less than 36 months and in the same hospitals as used in DAD admissions. Data was aggregated monthly and negative binomial regression model fitted.

3.8.6 Software

All the analyses were conducted using R statistical software version 4.0.0 (Ihaka & Gentleman, 1996)

3.9 Ethical Considerations

Data that was used in this research was collected as part of routine medical records and individual patients' consent was not obtained.

The Ministry of Health (Kenya) and participating hospitals have given permission for CIN collaboration, which involves sharing routine data with the research group. Clinical Information Network study has been approved by the Kenya Medical Research Institute (KEMRI) Scientific and Ethical Review Unit (SERU), which has approved use CIN data for observational research without individual consenting.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

This chapter will outline the results of the analysis and a discussion of the same. It begins with an exploratory data analysis, sample size and power calculation, ITS analysis and finally results from the sensitivity analysis.

4.2. Presentation of Results

This section will present the exploratory data analysis results, diagnostics for multiple imputation, ITS analysis results and sensitivity analysis.

4.2.1 Exploratory Data Analysis

I use imputed datasets for all the analyses. A total of 79,784 patients aged between 2 and 36 months (before imputation) were admitted to 13 CIN hospitals from 2013 to 2019. This is 64.27% (79784/128148) of all the patients admitted to all the CIN hospitals. An exploratory complete case analysis of the data showed pneumonia to be the leading cause of hospital admissions (40.9%, 32628/79784) followed by diarrhea and dehydration (38.0%, 29231/76784) then malaria (21.1%, 16803/79784). Out of the 32628 patients admitted due to pneumonia, 6% (1949/32628) died, 9.05% (2647/29231) of those diagnosed with diarrhea and dehydration also died. Moreover, 3.5% (600/16803) were classified as having died from malaria.

Patients admitted due to either diarrhea or dehydration comprised 40% (31813/79784) of all age-eligible admissions and 36.6% (29231/79784) had both diarrhea and dehydration and forms the study population as shown in the exclusion criteria in Figure 4.1. When classified according to gender, 12983/29231 (44.4%) were female and 16079/29231 (55.6%) were male.

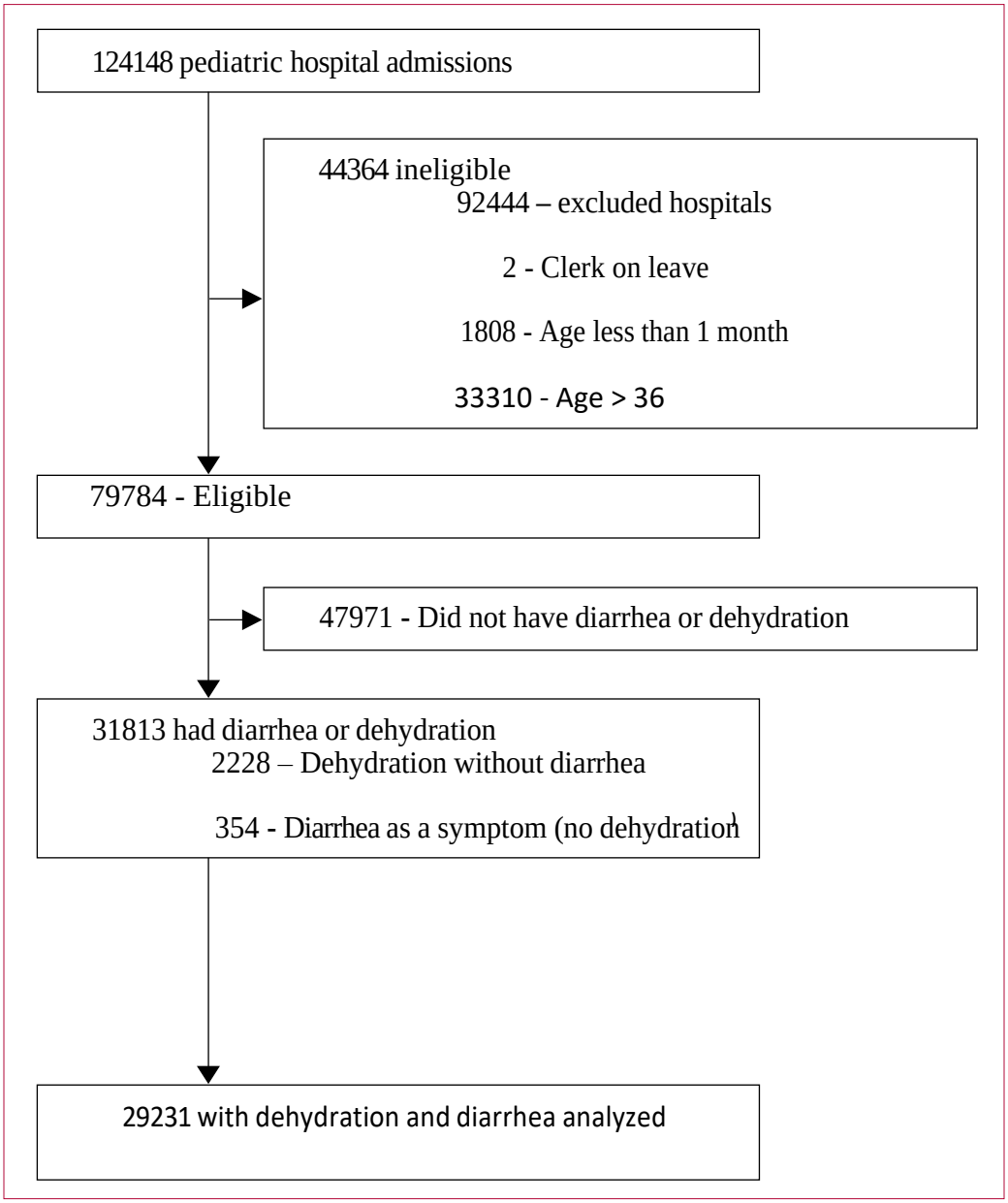


Figure 4.1: Patients selection criteria

A total of 66 months were included in the study with 11 months before vaccine introduction and 55 months after the vaccine. The average DAD admissions per month before the vaccine was introduced (July 2014) was 35 (standard deviation (SD): ±22) and 17 (SD: ±12) after vaccine introduction. A median number of DAD admissions before July 2014 was 570 and 429 after. Furthermore, the inter-quartile range between the two periods was 278 and 171 respectively.

Approximately 5% of the dataset had missing values. Appendix 2 summarizes the distribution of missing values in every hospital included in the study for every variable of interest.

4.2.2 Diagnostics for Multiple Imputation

The results are presented using imputed data and therefore begin with the diagnostics for multiple imputation to ensure that the imputation model was able to yield plausible values

Appendix 2 shows the percentages of missing observations in the variables per hospital. Different hospitals had different percentages of incomplete data for the different variables. Multiple imputation yielded values which when plotted, gave similar distributions as the observed values as shown in Figure 4.2. These suggest that the imputation model was able to give us plausible values which were used in the ITS analysis.

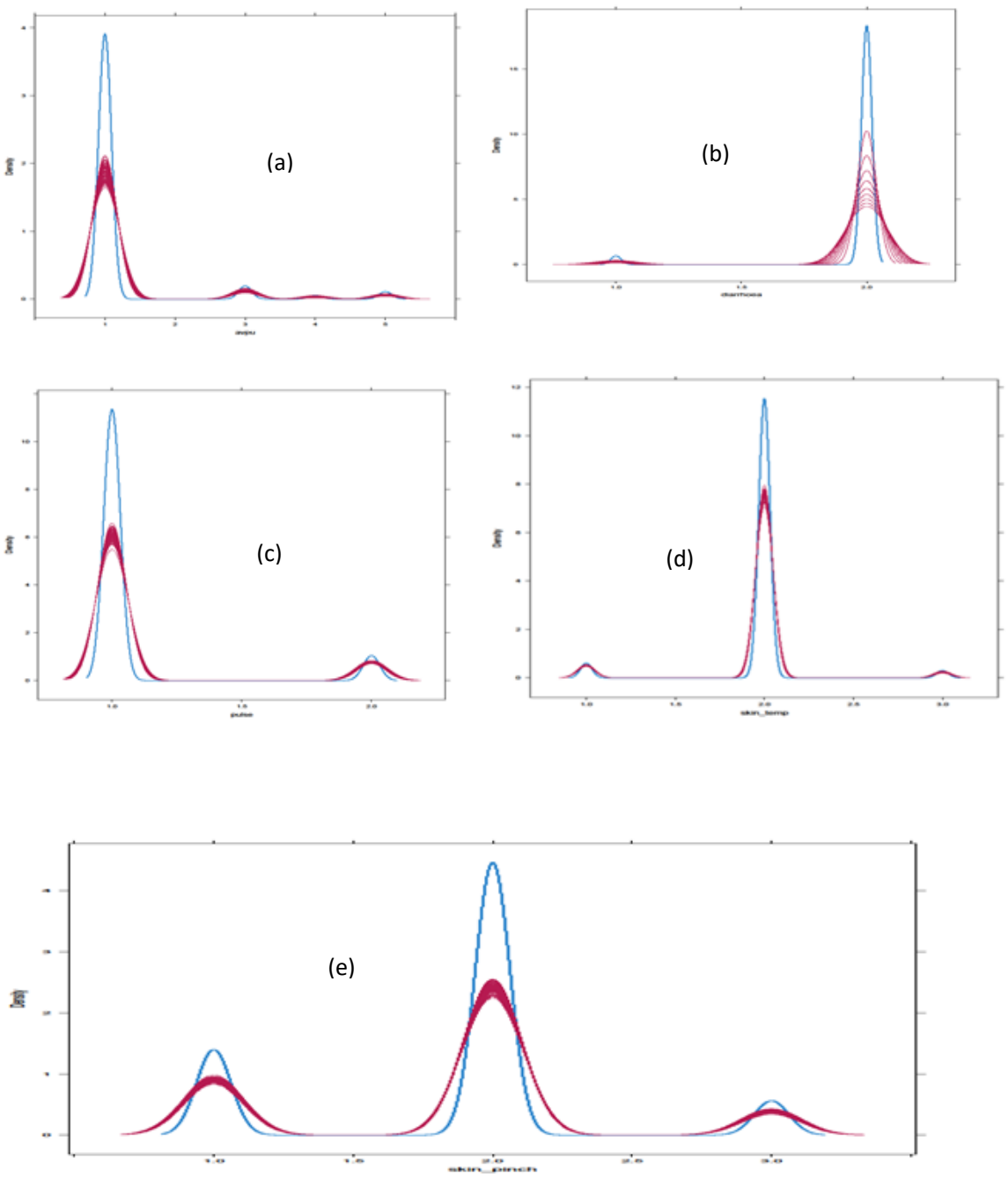


Figure 4.2: Distribution of imputed datasets versus observed

(a)- alertness, (b)- diarrhea, (c)- pulse rate,(d) – skin temperature, (e) – minutes till skin return when pinched

4.2.3. ITS Analysis Results

The fitted models yielded AICs: 625, 805 and 725. The model with the lowest AIC (625) was chosen and used to obtain coefficients reported. Low AIC implies that the lowest amount of information was lost by the model when compared to the other models used in the analysis.

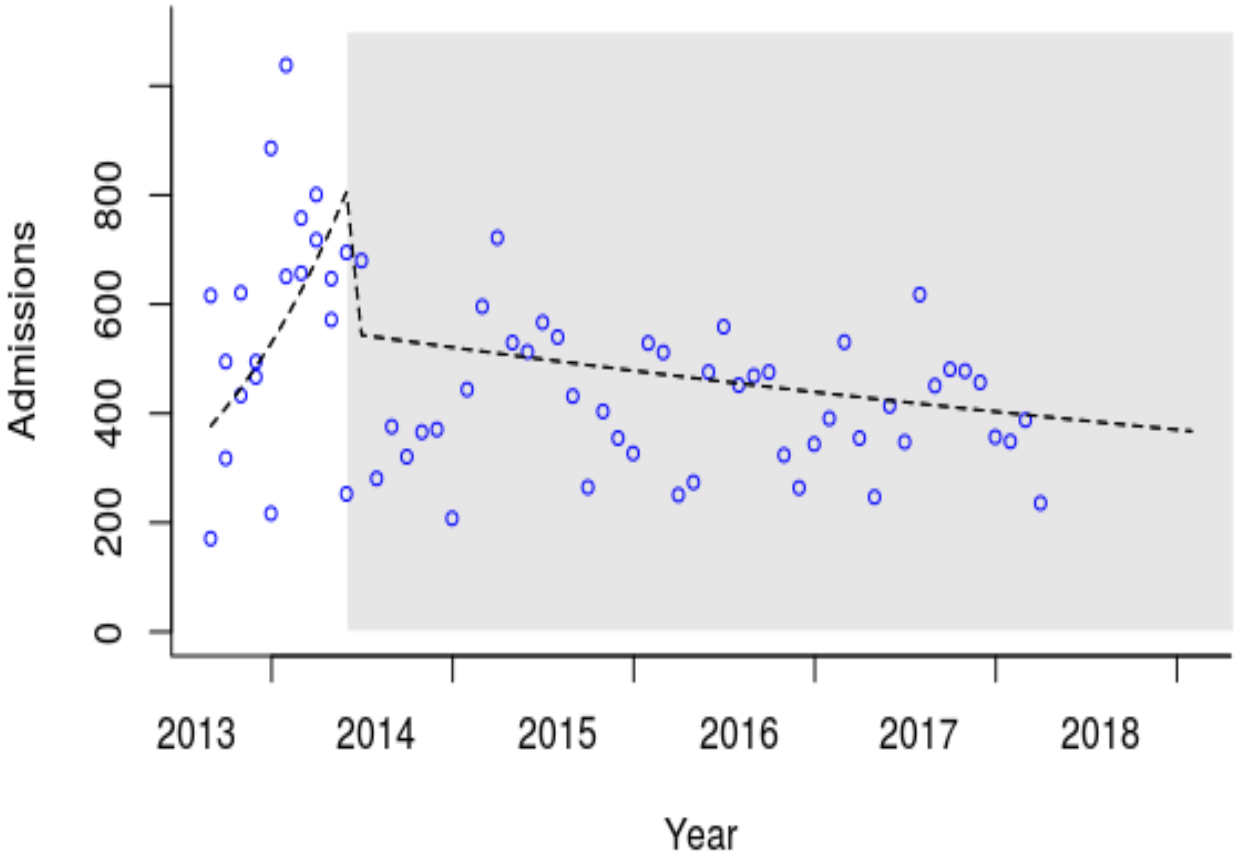
There was a 28.32% (95% C.I, 0.786 to 0.950) decrease in hospital admissions immediately after July 2014 when the vaccine was introduced to the Kenya routine childhood immunization program. This was followed by a 3.00% (95% C.I, 0.786 to 0.950) decrease in month to month hospital admissions due to all-cause diarrhea and dehydration after vaccine introduction as shown in Table 4.2.

Table 4.1: Interrupted time series analysis coefficients. *significant at p-value less than 0.1. ¨ p value not significant

	Coefficients	Exp (coefficients)	(Exp-1)*100	95% confidence Interval	P-Value
β_1	-0.0006	0.9994	-0.100%	0.984 to 1.116	0.2399¨
β_2	-0.3329	0.7168	-8.320%	0.945 to 2.185	0.090*
β_3	-0.030*	0.970	-3.000%	0.786 to 0.950	0.099*

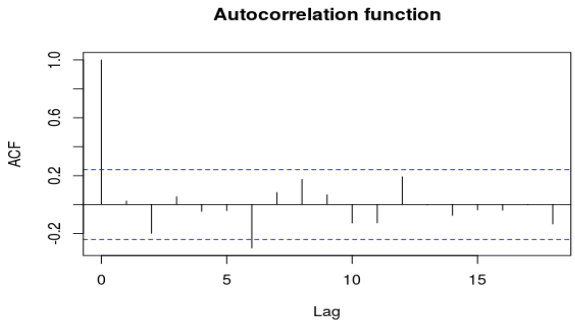
Note: β_1 - change in slope of DAD admissions before July 2014; β_2 - change in admissions immediately after July 2014; β_3 - change in slope of admissions after July 2014

Table 4.1 shows a summary of regression coefficients for change in admissions due to DAD following the introduction of rotavirus vaccine



. *Figure 4.3*: Trends in hospital admissions due to diarrhea and dehydration over time

Diagnostics of the chosen model 3.3 showed absence of serial correlation as illustrated by the autocorrelation and partial autocorrelation plots of Figure 4.4 a and b. Plots of residuals against time also showed good model fit as shown by Figure 4.6



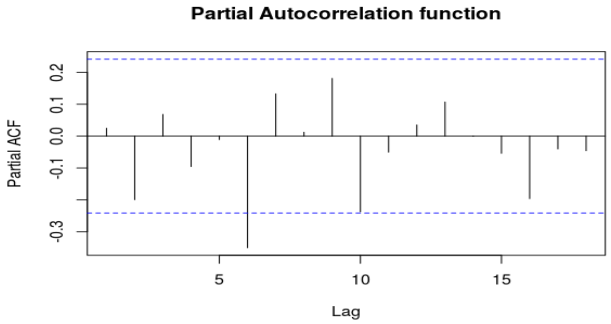


*Figure 4.4 a***.** ACF function

*Figure 4.4 b***.** PACF function

(Figure 4.5) indicates the relationship in the scatter between monthly DAD admissions and time is random and that there no obvious autocorrelation implying that the model was well fitted.

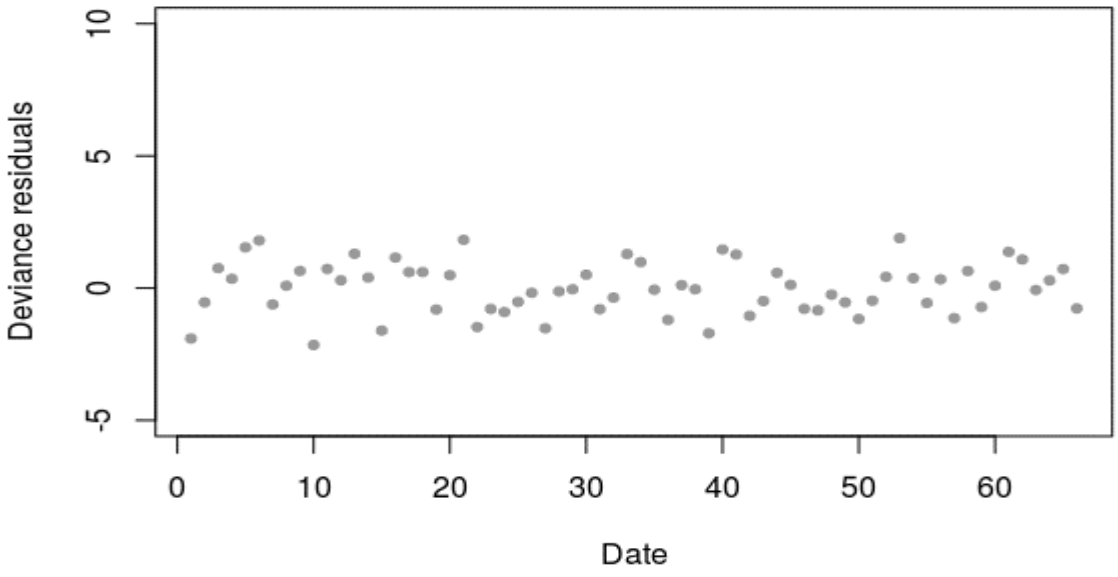


Figure 4.5: Plot of residuals over time

4.2.5. Results from Sensitivity Analysis

Interrupted time series of hospital admissions due to surgical burns indicated no significant change in trend post vaccination as summarized by Figure 4.6 and regression coefficients shown in Table 4.2

Table 4.2 : Regression coefficients comparing change in DAD admissions and burns. ** shows significant coefficients

Parameters	Coefficients	95% confidence interval	p-values
β_4	-1.91	0.20-0.57	0.00
B_5	-0.085*	0.81-1.00	0.00**
B_6	-1.20**	0.25-0.82	0.00**
B_7	0.098**	1.01-1.19	0.00**

Note: B_4 - difference in intercept; B_5 - difference in slope between intervention and control group before vaccination; B_6 - change in level difference in association with vaccination; B_7 - difference in slope change between treatment and control groups following vaccination

Analysis with both admissions due to burns and diarrhea and dehydration showed that the differences in logs of expected admissions due to DAD following vaccination was anticipated to increase by 0.098 compared to that of burns. Moreover, difference in the logs of anticipated admissions due to DAD immediately after vaccination was also expected to reduce by1.20 as summarized in Table 4.3.

This implies a reduction in slope following vaccination for the treatment group and no significant change for the control group and a further drop in level immediately when the vaccine was introduced for the treatment group and no change in the control group.

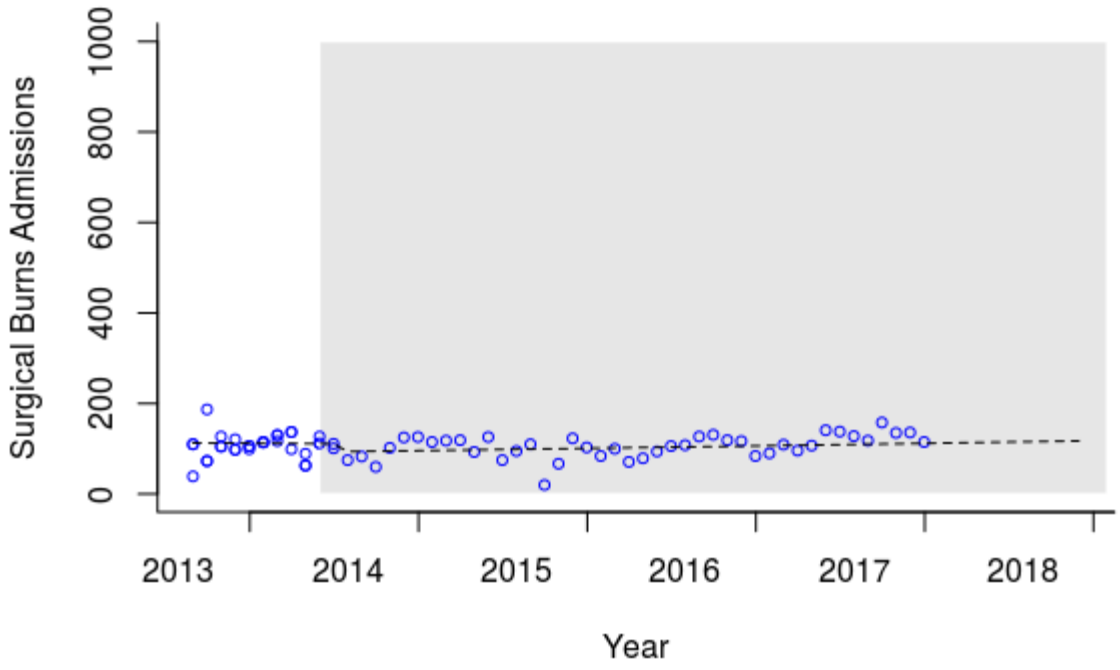


Figure 4.6: Trends in admissions due to surgical burns over time.

Missing data sensitivity analysis yielded results almost similar to the analysis from imputed datasets under the MAR assumption. This indicates that under the assumption, the imputation model gave plausible values. Regression coefficients from all the pooled pattern mixture scenarios discussed in Chapter Three are summarized in Table 4.3.

Table 4.3: Results from pattern mixture analysis

Parameters	Coefficients	Exp (coefficients)	(Exp-1)*100	95% confidence Interval	P-Value
β_1 (change in slope of DAD admissions before July 2014))					
	-0.0006	0.9994	-0.100%	0.984 to 1.116	0.2399¨
β_2 (change in admissions immediately after July 2014)					
	-0.333	0.7168	-28.32	0.945 to 2.185	0.088*
β_3 (change in slope of admissions after July 2014)					
	-0.030	0.970	-3.000%	0.786 to 0.950	0.099*

*significant at p-value less than 0.1. ¨ P value not significant. Case 1: case when any one variable is missing; Case 2: Case when any two variables are missing; Case 3: Case when any three variables are missing; Case 4: Case when any five variables are missing.

4.3. Discussion of Results

Exploratory analysis of the data from the 13 CIN hospitals showed diarrhea and dehydration to be among the top causes of hospitalization in children aged between two and thirty-six months. This is in line with a report from the Kenya Demographic Health Survey (KHDS) which also categorized diarrhea and dehydration as the second leading cause of hospitalization and death in Kenya and Sub-Saharan African countries in general (Heaton & Ciarlet, 2007).

Routinely collected data from 13 county level hospitals in Kenya were used. To decide whether a patient had both diarrhea and dehydration, I used the Ministry of health protocol for diarrhea and dehydration.

Patients with burns and surgical cases were excluded together with those who did not meet the threshold of having both diarrhea and dehydration. Patient characteristics as determined by the clinician were used.

Natural immunity for rotavirus is attained at age above 36 months therefore the patients in this age category were excluded.

In the study, 29231 patients were admitted due to DAD, 38% of all the 79784 admissions to the 13 CIN hospitals. Mortality from diarrhea and dehydration was approximated to be 9.05% which is similar to that reported by a 2018 study assessing the risk factors for DAD deaths. The data also showed more males (55.6%) than females (44.4%) being admitted due to DAD.

Hospitals joined the clinical information network between October 2013 and March 2014 and this therefore implied a short pre-intervention period of 11 months and a longer post intervention period of 55 months. Simulations were used to determine whether the available data would yield statistically significant power to detect change in DAD hospitalization. The data met the above 80% power mark.

Approximately 5% of all the data were missing and reason for missingness was unclear with the data being routinely collected hospital admission records. I therefore performed multiple imputation with the Missing At Random (MAR) assumption. Diagnostics for multiple imputation showed that the imputation model yielded plausible values and graphically shown by Figure 4.1. The distribution of imputed versus observed datasets appears to be similar. A sensitivity analysis would later show that there were no significant departures from the MAR assumption as shown by the pattern mixture results in Table 4.2.

Pre-Post analysis of the data showed a reduction in mean DAD hospitalization after intervention. The fitted regression analysis model also showed an approximately 3% reduction in all-cause DAD hospitalization following vaccination for a unit change in time. This indicates a clear association between change in volumes of children admitted due to all-cause DAD and the period of vaccine introduction.

A sensitivity analysis using admissions due to burns showed no change in hospitalizations and therefore clears the doubts for other confounding factors that might have contributed to the reduction in all-cause DAD. A confounding factor, if there had been any, would have also affected admissions due to burns.

This study however, was limited by the availability of missing data though handled by multiple imputation. The reason for missing observations was not clear and therefore, it would be more appropriate if missingness was reduced if not avoided in future times. Moreover, there is no specific formula that would be used for the calculation of sample size and power in ITS studies. However, simulations were used to determine if the sample size yielded enough power.

As a next step, it would be necessary to develop a formula that would be used by other researchers in calculation of power and sample size in interrupted time series analysis. Furthermore, study of how the vaccine has affected admissions due to other non-febrile diseases would also be investigated. A continuous surveillance of the vaccine performance will also be carried out regularly.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This section gives the summary of the study, recommendation, the limitations of the study, and plans for future work.

5.2 Summary

In summary, the was a total of 79784 patients admitted to the 13 CIN hospitals and 38% of this were due to all-cause diarrhea and dehydration. The study group included patients who were clinically diagnosed to be having diarrhea and dehydration and no stool samples were taken. More males were affected then females and case fatality for DAD was found to be 9.05%.

Routinely collected data is categorized by missing data and this was handled in the study using multilevel multiple imputation with chained equations under the Missing At Random assumption. Diagnostics for the imputations showed similar distributions of imputed versus observed datasets and this indicates that the model was able to yield plausible values. Sensitivity analysis relative to missing data using the pattern mixture model showed that there were no much departures from the MAR assumption.

5.3 Conclusions

In conclusion, since there is no specific formula for calculating power of the sample size for ITS studies, I used simulations to determine whether the sample size was powerful enough to detect change in volumes of admissions following vaccination. A power greater than 80% is required and the study was able to achieve this. The sample size was therefore able to yield enough power to detect change in volumes of admission following vaccination.

Interrupted time series model fitted to routinely collected data with seasonality and autocorrelation checked was able to show us the trends in hospital admissions due to diarrhea and dehydration following vaccination. There was a drop in volumes of admissions as soon as the vaccine was introduced followed by a gradual drop of approximately three percent.

Sensitivity analysis using burns showed no change in volumes of its admissions over time. This would therefore indicate that there was no confounding factor that would affect all-cause DAD admissions during that time and the results are less biased

In conclusion, there is an association between hospital admissions due to diarrhea and dehydration to the 13 Kenyan hospitals and rotavirus vaccine introduction. This is in children with age less than 36 months.

5.3 Recommendation

The vaccine has so far done well in reducing volumes of children admitted to CIN hospitals due to all-cause diarrhea and dehydration in children from 2013 to 2019. It is therefore recommended that continuous monitoring be done to ensure that its performance over time is known.

In addition, Missing observations in routinely collected data has to be handled with care so as not to bias results. The imputation model chosen has to be in agreement with the hypothesized missing data pattern and a sensitivity analysis to asses departures from the assumption.

Finally, the interrupted time series analysis methodology lacks in power and sample size calculation formulas. It is therefore recommended that more research be done on this part and methodologies be made available for people using ITS.

5.4 Future Work

It is necessary to expand the sample size calculation procedure to cover a wider range of assumptions and make it available for use by ITS users.

It will also be appropriate to check the effect of the vaccine to hospital admissions due to other non-febrile diseases and a continuous surveillance of vaccine performance be done

REFERENCES

Akech, S., Ayieko, P., Gathara, D., Agweyu, A., Irimu, G., Stepniewska, K., . . . Mutai, L. (2018). Risk factors for mortality and effect of correct fluid prescription in children with diarrhoea and dehydration without severe acute malnutrition admitted to Kenyan hospitals: an observational, association study. *The Lancet Child & Adolescent Health, 2*(7), 516-524.

Ayieko, P., Ogero, M., Makone, B., Julius, T., Mbevi, G., Nyachiro, W., . . . Irimu, G. (2016). Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in Kenyan hospitals within a new Clinical Information Network. *Archives of Disease in Childhood, 101*(3), 223-229.

Banajeh, S. M., & Abu-Asba, B. A. (2015). The epidemiology of all-cause and rotavirus acute gastroenteritis and the characteristics of rotavirus circulating strains before and after rotavirus vaccine introduction in Yemen: analysis of hospital-based surveillance data. *BMC Infectious Diseases, 15*(1), 1-10.

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior, 32*(1), 83-92.

Carpenter, J., & Kenward, M. (2012). *Multiple imputation and its application*: John Wiley & Sons.

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods, 23*(2), 298.

Fischer, T. K., Valentiner-Branth, P., Steinsland, H., Perch, M., Santos, G., Aaby, P., . . . Sommerfelt, H. (2002). Protective immunity after natural rotavirus infection: a

community cohort study of newborn children in Guinea-Bissau, west Africa. *The Journal of Infectious Diseases, 186*(5), 593-597.

Folorunso, O. S., & Sebolai, O. M. (2020). Overview of the development, impacts, and challenges of live-attenuated oral rotavirus vaccines. *Vaccines, 8*(3), 341.

Gikonyo, J., Mbatia, B., Okanya, P., Obiero, G., Sang, C., & Nyangao, J. (2019). Rotavirus prevalence and seasonal distribution post vaccine introduction in Nairobi county Kenya. *The Pan African Medical Journal, 33*.

Health, M. o. (2007). Basic Paediatric Protocols.

Heaton, P. M., & Ciarlet, M. (2007). The pentavalent rotavirus vaccine: discovery to licensure and beyond. *Clinical Infectious Diseases, 45*(12), 1618-1624.

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299-314.

Irimu, G., Ogero, M., Mbevi, G., Agweyu, A., Akech, S., Julius, T., . . . English, M. (2018). Approaching quality improvement at scale: a learning health system approach in Kenya. *Archives of Disease in Childhood, 103*(11), 1013-1019.

Jonesteller, C. L., Burnett, E., Yen, C., Tate, J. E., & Parashar, U. D. (2017). Effectiveness of rotavirus vaccination: a systematic review of the first decade of global postlicensure data, 2006–2016. *Clinical Infectious Diseases, 65*(5), 840-850.

Khagayi, S., Omore, R., Otieno, G. P., Ogwel, B., Ochieng, J. B., Juma, J., . . . Ngama, M. (2020). Effectiveness of monovalent rotavirus vaccine against hospitalization with acute rotavirus gastroenteritis in Kenyan children. *Clinical Infectious Diseases, 70*(11), 2298-2305.

Kirk, M. D., Angulo, F. J., Havelaar, A. H., & Black, R. E. (2017). Diarrhoeal disease in children due to contaminated food. *Bulletin of the World Health Organization, 95*(3), 233.

Lerman, P. (1980). Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 29*(1), 77-84.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793): New Jessey, John Wiley & Sons.

López Bernal, J. (2018). *The use of interrupted time series for the evaluation of public health interventions.* London School of Hygiene & Tropical Medicine,

Mistler, S. A., & Enders, C. K. (2017). A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics, 42*(4), 432-466.

Muendo, C., Laving, A., Kumar, R., Osano, B., Egondi, T., & Njuguna, P. (2018). Prevalence of rotavirus infection among children with acute diarrhoea after rotavirus vaccine introduction in Kenya, a hospital cross-sectional study. *BMC Pediatrics, 18*(1), 1-9.

Ngabo, F., Tate, J. E., Gatera, M., Rugambwa, C., Donnen, P., Lepage, P., . . . Parashar, U. D. (2016). Effect of pentavalent rotavirus vaccine introduction on hospital admissions for diarrhoea and rotavirus in children in Rwanda: a time-series analysis. *The Lancet Global Health, 4*(2), e129-e136.

Nicholls, S. G., Langan, S. M., & Benchimol, E. I. (2017). Routinely collected data: the importance of high-quality diagnostic coding to research. *CMAJ, 189*(33), E1054-E1055.

Otieno, G. P., Bottomley, C., Khagayi, S., Adetifa, I., Ngama, M., Omore, R., . . . Ochieng, J. B. (2020). Impact of the introduction of rotavirus vaccine on hospital admissions for

diarrhea among children in Kenya: a controlled interrupted time-series analysis. *Clinical Infectious Diseases, 70*(11), 2306-2313.

Raes, M., Strens, D., Kleintjens, J., Biundo, E., Morel, T., & Vyse, A. (2016). Epidemiological trends for hospital admissions for acute rotavirus gastroenteritis in Belgium following the introduction of routine rotavirus vaccination and the subsequent switch from lyophilized to liquid formulation of Rotarix™. *Epidemiology & Infection, 144*(14), 3017-3024.

Rahajamanana, V., Raboba, J., Rakotozanany, A., Razafindraibe, N., Andriatahirintsoa, E., Razafindrakoto, A., . . . Burnett, E. (2018). Impact of rotavirus vaccine on all-cause diarrhea and rotavirus hospitalizations in Madagascar. *Vaccine, 36*(47), 7198-7204.

Rezvan, P. H., Lee, K. J., & Simpson, J. A. (2015). The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Medical Research Methodology, 15*(1), 1-14.

Rubin, D. B. (1988). *An overview of multiple imputation.* Paper presented at the Proceedings of the survey research methods section of the American statistical association.

Schwartz, L. M., Zaman, K., Yunus, M., Basunia, A.-u. H., Faruque, A. S. G., Ahmed, T., . . . Rowhani-Rahbar, A. (2019). Impact of rotavirus vaccine introduction in children less than 2 years of age presenting for medical care with diarrhea in rural Matlab, Bangladesh. *Clinical Infectious Diseases, 69*(12), 2059-2070.

Tang, L., Song, J., Belin, T. R., & Unützer, J. (2005). A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine, 24*(14), 2111-2128.

Tuti, T., Bitok, M., Paton, C., Makone, B., Malla, L., Muinga, N., . . . English, M. (2016). Innovating to enhance clinical data management using non-commercial and open source

solutions across a multi-center network supporting inpatient pediatric care and research in Kenya. *Journal of the American Medical Informatics Association, 23*(1), 184-192.

Van Buuren, S. (2011). *Multiple imputation of multilevel data*: Routledge.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1-67.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., . . . Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open, 3*(8), e002847.

Wandera, E. A., Mohammad, S., Bundi, M., Nyangao, J., Galata, A., Kathiiko, C., . . . Komoto, S. (2018). Impact of rotavirus vaccination on rotavirus hospitalisation rates among a resource-limited rural population in Mbita, Western Kenya. *Tropical Medicine & International Health, 23*(4), 425-432.

Wandera, E. A., Mohammad, S., Ouko, J. O., Yatitch, J., Taniguchi, K., & Ichinose, Y. (2017). Variation in rotavirus vaccine coverage by sub-counties in Kenya. *Tropical Medicine and Health, 45*(1), 1-5.

APPENDICES

Appendix 1: Interrupted Time Series Sample Power Simulation Codes

```
rm (list = ls())

library (gcmr)

####-----------Defining variables-----#####

set.seed(123)

#months_pre - the number of pre - intervention time points (months)

#months_post - the number of post - intervention time points (months)

#pre_probability - pre - intervention event rate

#cntrl_pre_probability - pre intervention event rate in the control group

#post_probability - post - intervention event rate

#cntrl_post_probability - intervention event rate in the control group

#n_month - sample size per month

#simul_n - number of simulated datasets

#intervention - denotes status of intervention (0/1)

#slope - time after intervention
```

```
its_power_calculation = function
(months_pre=12,months_post=12,pre_probability=0.4,cntrl_pre_probability=0.3,
post_probability=0.2,cntrl_post_probability=0.8,n_month=100,simul_n=1000)

{

  #1.----------------------function to create design matrix for interrupted time series-------------------
---############

  its_model_matrix<-
function(n_pre=months_pre,n_post=months_post,pre_prob=pre_probability,cntrl_pre_prob=cntrl
_pre_probability,
post_prob=post_probability,cntrl_post_prob=cntrl_post_probability,n=n_month){

    month<-rep(1:(n_pre+n_post),2)

    post<-if(n_post>0){

      cbind(intervention=c(rep(0,n_pre),rep(1,n_post), rep(0,(n_pre+n_post))),

slope=c(rep(0,n_pre),1:(n_post),rep(0,n_pre),1:(n_post)),
group=factor(c(rep(1,(n_pre+n_post)),rep(0,(n_pre+n_post)))) )    }else NULL

       ####The next line of code assumes that event probability would be constant across all the
pre-intervention time points

    pre_prob_time_point = rep(pre_prob,n_pre)
cntrl_pre_prob_time_point=rep(cntrl_pre_prob,n_pre)

    #####The next line assumes the rate of change would be consistent between the months

    post_change=(pre_prob-post_prob)/n_post

    cntrl_post_change=(cntrl_pre_prob-cntrl_post_prob)/n_post
```

```
    post_prob_time_point = rep(pre_prob,n_post) - (post_change*1:n_post)

cntrl_post_prob_time_point = rep(cntrl_pre_prob,n_post) - (cntrl_post_change*1:n_post)
event_probs_intervention=c(pre_prob_time_point,post_prob_time_point)
cntrl_event_probs=c(cntrl_pre_prob_time_point,cntrl_post_prob_time_point)

    ###Estimating number of events from binomial distribution

    pre_sim_prob_time_point= rbinom (n_pre,n,pre_prob_time_point)

    post_sim_prob_time_point= rbinom (n_post,n,post_prob_time_point)
sim_probs_intervention=c(pre_sim_prob_time_point,post_sim_prob_time_point)/n
#return(cbind(month=month,post,event_probs,sim_probs))

g    ### Estimating number of events for the control group using binomial distribution
cntrl_pre_sim_prob_time_point= rbinom (n_pre,n,cntrl_pre_prob_time_point)
cntrl_post_sim_prob_time_point= rbinom (n_post,n,cntrl_post_prob_time_point)
cntrl_sim_probs=c(cntrl_pre_sim_prob_time_point,cntrl_post_sim_prob_time_point)/n
event_probs=c(event_probs_intervention,cntrl_event_probs)
sim_probs=c(sim_probs_intervention,cntrl_sim_probs)
return(cbind(month=month,post,event_probs,sim_probs))  }

  #####----------2.Simulating more than one dataset----------
############################

  datasets<- function(simul_n){ lapply(1:simul_n , function(x){its_model_matrix(}) }

 ####----------Fitting beta regression AR (1) model to datasets generated above----------
-----################################

  extracting_pvalues=lapply (datasets(simul_n), function(x)  { x = as.data.frame(x)
x$sim_probs[x$sim_probs==1]=(x$sim_probs-0.0000001)[x$sim_probs==1]
x$sim_probs[x$sim_probs==0]=(x$sim_probs+0.0000001)[x$sim_probs==0]
```

```
   model=gcmr(sim_probs ~ month + intervention+slope+group, data = x, marginal = beta.marg,
cormat =arma.cormat( p=1, q=0 ))    summary(model)$coef$marginal[3:4,4] })


  ###-----------Calculating power to detect changes in the slope-----------------------
##############################


  calculating_power = function (pvalues=extracting_pvalues) {p_value_data_frame =
as.data.frame(do.call(rbind,pvalues))    power_intervention =
round((prop.Table(Table(p_value_data_frame$mean.intervention<0.05))["TRUE"])*100)


    power_slope =
round((prop.Table(Table(p_value_data_frame$mean.slope<0.05))["TRUE"])*100)
#intervention_slope_power=cbind(power_intervention,power_slope)


    return (paste0(power_intervention," %"))  }


  return (calculating_power())}


####------------------calling the power calculation function---------------------------------------


its_power_calculation (months_pre=
11,months_post=40,pre_probability=.25,cntrl_pre_probability = 0.26,
post_probability=0.12,cntrl_post_probability =0.20 , n_month=100,simul_n=200)
```

Appendix 2: Percentage of Missing Data per Hospital per Variable

Hospital	A	B	C	D	E	F	G	H	I	J	K	L	M	Avarage
History of diarrhea														
Yes	34.9	30.9	32.1	25.0	23.1	32.2	37.5	41.2	30.9	40.8	38.3	23.6	27.8	32.7
No	60.8	65.2	61.9	68.3	74.6	60.7	53.2	58.4	66.4	57.2	59.3	74.7	66.7	63.3
Missing	4.3	3.9	6.0	6.7	2.2	7.1	9.3	0.4	2.7	2.0	2.3	1.8	5.5	4.0
Sex														
Female	43.2	42.0	45.6	44.8	43.3	45.8	44.6	44.6	42.4	45.1	44.8	43.3	42.7	44.1
Male	56.7	57.3	53.4	52.3	56.5	53.9	54.7	55.3	57.1	54.3	54.8	55.1	55.9	55.2
Missing	0.2	0.7	0.9	2.9	0.2	0.3	0.7	0.2	0.6	0.6	0.4	1.6	1.4	0.7
Age group														
<= 6 months	13.3	20.9	14.5	19.3	15.9	23.6	14.0	16.1	24.4	27.3	27.3	20.4	14.5	80.6
> 6 months	86.7	79.1	85.5	80.7	84.1	76.4	86.0	83.9	75.6	72.7	72.7	79.6	85.5	19.4
Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pulse														
Normal	79.3	84.6	77.7	84.6	87.4	86.5	76.5	91.8	90.6	93.0	86.6	94.6	80.2	85.8
Weak	8.0	4.3	5.7	2.5	2.7	2.6	4.2	4.8	3.7	3.5	9.0	2.4	4.5	4.6
Missing	12.7	11.0	16.7	12.9	9.9	10.9	19.3	3.4	5.7	3.5	4.4	3.0	15.3	9.6
Minutes till cap refill														
1 second	52.0	69.1	46.9	74.0	47.2	55.7	33.8	73.7	41.3	39.6	48.1	77.2	50.9	54.7
2 seconds	18.3	13.9	21.1	12.1	9.8	16.3	23.2	15.3	37.3	49.2	30.5	10.4	18.4	21.3
3 seconds	5.8	3.2	2.9	1.0	2.2	1.7	4.7	2.7	6.4	2.9	6.8	1.5	4.8	3.6
4 seconds	1.1	1.0	0.5	0.1	0.5	0.5	1.9	0.3	1.5	0.3	1.6	0.2	0.9	0.8
Hospital	A	B	C	D	E	F	G	H	I	J	K	L	M	TOTAL
5 seconds	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.4	0.0	0.3	0.1	0.1	0.1
6 seconds	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.2	0.0	0.0	0.0
6+ seconds	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.3	0.0	0.0	0.1
Indeterminate	0.5	2.7	3.9	0.1	21.9	6.6	8.7	0.1	1.8	0.0	0.2	0.1	6.0	3.9
Missing	22.3	10.0	24.6	12.6	18.3	19.1	27.4	7.6	11.2	7.9	12.0	10.6	18.9	15.5
Skin temperature (temperature gradient)														
Elbow	1.4	3.8	2.0	0.8	2.1	1.4	1.7	1.5	2.9	1.4	5.7	2.3	2.5	2.3
Hand	70.5	79.6	72.3	64.9	59.4	72.5	72.3	91.3	83.0	92.9	74.6	88.9	77.9	77.6
shoulder	1.9	3.5	3.0	1.4	1.5	1.3	0.9	0.6	1.3	0.5	2.2	1.3	1.7	1.7

Missing	26.1	13.1	22.7	32.9	37.1	24.8	25.1	6.3	12.8	5.2	17.5	7.4	17.8	18.4
Delayed skin pinch														
1-2 seconds	19.3	14.0	14.7	13.5	10.1	5.3	19.2	10.9	19.2	7.7	20.1	10.8	11.6	13.4
Immediate	61.5	72.1	66.9	74.0	78.6	81.3	56.1	80.4	68.2	87.7	65.2	82.1	69.8	73.0
>=2 seconds	5.3	3.5	3.4	2.5	2.3	2.3	5.6	4.3	5.4	1.0	9.3	2.4	3.1	3.9
Missing	13.8	10.4	15.0	10.0	9.0	11.1	19.1	4.3	7.2	3.6	5.4	4.6	15.5	9.7
Sunken eyes														
Yes	14.2	12.1	8.7	6.1	7.3	9.5	19.3	16.6	14.8	5.1	20.0	10.1	5.6	11.7
No	76.6	79.7	77.8	54.0	89.0	75.1	64.4	80.1	74.3	90.9	73.9	87.2	80.7	78.3
Missing	9.2	8.1	13.5	39.8	3.7	15.4	16.3	3.3	11.0	3.9	6.1	2.7	13.7	10.0
Alertness														
Alert	83.8	90.8	89.5	89.9	94.1	88.0	86.1	92.5	88.1	88.7	86.8	94.6	81.2	89.0
Other scale	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.1
Pain response	4.1	2.7	3.2	0.9	1.8	2.0	2.7	3.0	3.5	4.6	5.2	1.4	4.4	3.1
Unresponsive	1.1	1.3	1.1	0.5	0.4	0.6	0.6	0.9	1.2	0.9	1.5	0.6	1.6	0.9
Verbal response	3.1	1.1	1.5	3.2	0.8	0.8	2.1	1.8	2.7	2.6	3.6	1.2	2.9	2.0
Missing	7.9	4.0	4.7	5.6	2.8	8.5	8.5	1.9	4.4	3.1	2.9	2.3	9.8	4.9
Inability to drink														
Yes	72.2	77.3	76.7	76.7	69.3	76.4	66.4	85.1	77.6	79.6	67.6	86.3	66.9	75.8
No	18.1	14.3	9.4	12.5	27.2	12.3	15.4	10.4	17.01	17.1	27.8	9.6	17.6	15.8
Missing	9.7	8.4	13.9	10.7	3.5	11.3	18.2	4.5	5.3	3.3	4.6	4.1	15.5	8.4

Appendix 3: Funding

This work was supported through the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769/Z/10/Z] and the UK government. The views expressed in this thesis, however, are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Appendix 4: Publication