Morbidity Changes between HIV Unexposed Uninfected and HIV Exposed Uninfected Children in Harare –A Secondary Data Analysis

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR

THE DEGREE OF MASTERS OF SCIENCE IN BIOSTATISTICS

IN THE FACULTY OF MEDICINE

BY

ZVIFADZO MATSENA

Supervisor: Professor S. Rusakaniko

Co-supervisor: Mr W. Tinago

DEPARTMENT OF COMMUNITY MEDICINE

FACULTY OF MEDICINE

31 August 2012

UNIVERSITY OF ZIMBABWE

DECLARATION

I Zvifadzo Matsena certify that this dissertation is my original work and submitted for the degree of Masters of Science in Biostatistics program. It has not been submitted in part or in full to any University and/or any publication.

STUDENT

Signature_____ Date _____

ZVIFADZO MATSENA

I am satisfied that this is the original work of the author in whose name it is being presented**.**

I confirm that the work has been completed satisfactorily for presentation in the examination.

PRINCIPAL SUPERVISOR

Signature_____ Date_____

Professor S. Rusakaniko

CO-SUPERVISOR

Signature_____ Date_____

Mr W. Tinago

CHAIRMAN

Signature_____ Date_____

Professor S. Rusakaniko

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

Background: Optimising the survival of HIV exposed uninfected (HEU) infants is a major challenge. There is a significant swift increase in the HEU population due to the introduction of the highly active antiretroviral therapy (HAART). Infections may be more severe in the HEU children as compared to their HIV unexposed uninfected (HUU) counterparts. Longitudinal studies give an understanding of the morbidity patterns in HEU children and possible factors associated with the observed morbidity differences between HEU and HUU can be explained through a longitudinal model. Broadly, this study aims to assess morbidity trends and factors associated with change in morbidity between HEU and HUU children in a nine months follow-up period.

Materials and Methods: A cohort of index babies was followed up from delivery for nine months. The maternal HIV status during pregnancy set as the exposure status for this cohort. Morbidity outcomes, illnesses and admissions, were observed within the follow-up period between the HEU and HUU children. HIV exposed infected (HEI) index babies were excluded from the analysis. The follow-up time points were at six weeks, four months and nine months. Mixed effects logistic regression analysis was used to determine factors associated with change in morbidity between the HEU and HUU.

Results: The average child-specific intercept for the log odds of morbidity was 1.04. There was a 1.12 heterogeneity difference at baseline. A negative exposure change of 0.06 in the first sixteen weeks and a positive exposure change of 0.04 after sixteen weeks were observed. Being HEU had a protective effect with an odds ratio of 0.77 and a confidence interval of (0.38; 1.26) which is not statistically significant. .

Conclusion: Being exposed to HIV is protective with an odds ratio of 0.77 (0.38; 1.26). There is a significance difference in the heterogeneity of the groups at baseline. The unexposed group has a significant negative trend during the first sixteen weeks and a positive trend after sixteen weeks. The exposed group has a less negative and positive trend across time. The family size has a protective effect towards morbidity in children.

DEDICATION

I dedicate this whole work to my late daddy, Alex Matsena. You really inspired me so much daddy. You saw things which were so invisible to me. I came this far because of you. I will always cherish the moments we shared together. You will forever be missed.

ACKNOWLEDGEMENTS

My heartfelt gratitude goes to my Principal Supervisor Professor S. Rusakaniko and my co-supervisor Mr W. Tinago for the proficient supervision they gave me during the development of my project. I also thank Mr V. Chikwasha for keeping on encouraging me even when things were tough. To my lecturer, Mr G. Mandozana, thank you so much for the academic mentorship you gave me throughout the whole programme.

My appreciation goes to Letten Foundation Research Centre for the scholarship package I got during the two academic years. I thank Dr E.N. Kurewa for the provision of the dataset used for this project. Thanks so much for the initial guidance of this work and support. Your guidance really was great. I also thank the Letten Foundation colleagues for the reviewing my work during proposal development. Your comments and suggestions really paid off.

I would like to thank my attachment supervisor, Dr Sue Laver for providing an enabling working and reading environment during the development of this project. Not forgetting the CCORE team for motivating me, especially, Felicia Takavarasha.

To my biostatistics classmates, Rutendo Birri, Joice Tome, Tawanda Ndapwadza, Jaspar Maguma and Fastel Chipepa, thanks guys. You really supported me to make this work a success. You had so much words of encouragement and you were ever there for me.

My appreciation also goes to my dearest family, mum, Christine, Dzikamai and my daughter Thelma for giving me time to pursue my dream. Your love is great. To my beloved fiancé, Nathan Zingoni, thanks so much for being the pillar of my strength. You always gave me a shoulder to lean on. My appreciation goes to Mr W. Soko for relighting my wish to do this programme. You really did great. Thank You Lord!

DEFINITION OF TERMS

Morbidity is the incidence of ill health or prevalence of a disease within a defined population. In other studies it has been defined as any illness and /or hospital admission and malnutrition (weight-for-age Z-score, <=3).

Index infant was defined as the child the mother gave birth to from the pregnancy she was enrolled with**.**

LIST OF ABBREVIATIONS

HIV	Human immunodeficiency virus
AIDS	Acquired immune deficiency syndrome
PMTCT	Preventing mother to child transmission
HUU	HIV unexposed uninfected
HEU	HIV exposed uninfected
HEI	HIV exposed infected
HIC	High income countries
HAART	Highly active antiretroviral drugs
BHAMC	Better Health for African Mothers and Children
ARV	Antiretroviral drugs

CHAPTER 1

1. INTRODUCTION

Human immunodeficiency virus (HIV) is a virus that causes acquired immunodeficiency syndrome (AIDS). The three major routes of transmission are unsafe heterosexual transmission (92%), vertical transmission (7%) and other blood (1%) [1]. Globally, about 33.4 million people are HIV infected [1]and Sub-Saharan Africa (SSA) is one of the most heavily affected regions. The region is home to 10% of the world's population, yet it accounts for 70% of HIV infected people [2].

Zimbabwe remains one of the hardest hit countries by the epidemic with a national prevalence rate of 15 % in the adult population of 15-49 years[3]. Of the total number of the infected people, close to fifty-two thousand are pregnant women and in 2011, 78% of pregnant women with HIV received ARVs for PMTCT. The PMTCT programme in Zimbabwe is a national priority in the fight against HIV/AIDS in children. The HIV prevalence among pregnant women (aged 15-49) is 16.1% [4]. The high HIV prevalence in pregnant women still leaves vertical transmission of HIV as a major challenge in the country. Between 2009 and 2011, Zimbabwe has seen a 45% decline in the number of new paediatric HIV infections (HEI) from seventeen thousand and seven hundred to nine thousand and seven hundred. There has been a significant rapid increase in the HEU children population [5].

HIV-exposed uninfected (HEU) children are a rapidly growing population in the world. Prevention of mother to child transmission (PMTCT) programs, have reduced the transmission rate of perinatal HIV infection to approximately between 2% to 5% [6] and as low as 1% in High Income Countries (HIC). Highly active antiretroviral therapy (HAART) has improved the health of HIV infected patients. PMTCT programs have therefore effectively

reduced the number of HIV exposed infected (HEI) children but resulted in an increase in the HEU children [7].

HEU children have been overlooked as a group of children who may be at an increased risk of illness compared to HIV unexposed uninfected (HUU) children. Recently, increased morbidity in HEU children compared to HUU children has been reported to be different in HIV-endemic areas of Sub-Saharan Africa. In Zimbabwe, it was found that sick clinic visits in HEU children were 1.2 times more common as compared to HUU children [8].

Infections may be more common and severe in HEU children than among HUU. HIV infection is one of the leading causes of morbidity and mortality in different age groups and in Zimbabwe, it is the second highest. Mostly, 21% of the causes of mortality in the under-five year olds are indirectly linked to HIV [2, 9]. There are a number of other factors which also may contribute to the increased morbidity in the HEU children which are feeding practice, non-breastfeeding, innate deficiency, exposure to HIV drugs and poor protection of maternal antibodies [10, 11].

Non-breastfeeding is one of the major causes of morbidity since it results in malnutrition of HEU children. Innate deficiency in immunity and poor protection from maternal antibodies results in the mother being immune-compromised and has other infections which the child is most likely to be exposed to. Exposure to antiretroviral drugs of the child if the mother is on antiretroviral therapy, parental illness or death resulting in reduced care to the child can also influence the morbidity of the HEU children[12].

Optimising the survival of HEU infants is a major challenge in Sub-Saharan Africa where prevalence of HIV infection remains high among women in the reproductive age group[13]. In Zimbabwe, there is more focus on the HIV-exposed infected (HEI) children compared to

HEU. Since there is a rapid increase of this population, their health problems are of enormous public health importance.

Most studies have looked at the effect of maternal HIV exposure on the mortality of HEU children and have described less on their morbidity. It is of importance that the morbidity characterisation of HEU children is known in terms of the illness they present with at hospitals. Moreover, an understanding of some of the factors contributing to HEU morbidity patterns plays a significant role in targeting, allocating and mobilising resources especially in the fourth prong of PMTCT.

Morbidity outcomes have been defined as illnesses and or hospital admission and malnutrition (weight-for-age Z-score, <= 3) in children studies as shown by a study in Malawi on the effect of breastfeeding cessation in HEU infants[14] and the common illnesses looked at include diarrhoea, fever, vomiting, cough, oral thrush, ear infection and conjunctivitis[14] .

To have a clear understanding of the morbidity patterns in HEU children, longitudinal studies have been used with HIV-unexposed uninfected (HUU) counterparts as comparison group. Comparison is made mainly on the different specific types of illnesses they present with like diarrhoea or fever and malnutrition. Morbidity patterns can be drawn from the follow-up period and the possible factors that are associated with the observed morbidity differences can be identified. The effect of each possible factor can be explained through a longitudinal model.

The Better Health for African Mothers and Children (BHAMC) cohort in Zimbabwe is one of the longitudinal studies that are still ongoing. This study focuses mainly on the PMTCT transmission rate of HIV during pregnancy, birth and breastfeeding. Since it is a follow up

study, a number of outcomes such as morbidity outcomes can be studied on index children cohort.

This project aims to use BHAMC study data to compare the morbidity of HEU children with HUU children born under a PMTCT program in Zimbabwe in a nine months follow-up period and identify possible factors that are associated with change in the morbidity between these children. A generalised linear mixed effects model was used in this study with a key binary variable morbidity outcomes (illness or admissions/ not ill or not admitted).

1.2 DESCRIPTION OF THE BHAMC STUDY

1.2.1 Background of the BHAMC study

HIV prevalence among antenatal care attendees (ANC) was estimated to be approximately 26% in 2002. Zimbabwe embarked on a national Prevention of mother-to-child transmission (PMTCT) of HIV and voluntary counselling and testing (VCT) for HIV was offered to all women attending antenatal care. Those who would have tested HIV positive were given single dose Nevirapine (sdNVP) to self administer at the onset of labour and new born infants were given Nevirapine within three days of birth. Mothers were asked on delivery if they had taken the NVP dose on the onset of labour just to be sure the medication was taken. As a result the Better Health for the African Mothers and Children (BHAMC) study was designed and subjects recruited based on their HIV status. The objectives of the cohort study were:- To assess the role of sexually transmitted infections on mother-to-child transmission of HIV and the impact of single dose Nevirapine given to babies born to HIV positive mothers on their neurological development compared to children born to HIV negative mothers; To describe and compare the growth pattern and neurological development of children exposed to omega-3 tablets and those not exposed: To describe the incidence of sexually transmitted infections and HIV among women enrolled into PMTCT.

1.2.2 BHAMC Research Problem

HIV prevalence in Zimbabwe was estimated to be 25.7% among antenatal clinic (ANC) attendees [15]. A PMTCT programme of HIV initiative had been adapted by the government of Zimbabwe in 2002, but its contribution to the reduction of HIV vertical transmission among African populations was not clear, that is, has the programme reduced the number of HIV infected children or number of children dying from it. Evidence of the effectiveness of PMTCT and voluntary counselling and testing (VCT) was relatively weak.

1.2.3 BHAMC Research Questions

- What are the realities and challenges of following up HIV positive and negative mothers and child pairs enrolled in a PMTCT program?
- To what extend has PMTCT influenced health status and the survival of HIV positive and negative mother and child pairs?

1.2.4 BHAMC Justification

The goal of the UGASS 2001 commitment was to reduce global MTCT of HIV by 20% by 2005 and 80% of pregnant women with access to antenatal care were provided with preventive services including VCT and ARVs (UNAIDS, 2000). PMTCT intervention efficacy has been blemished by conflicting results due to its implementation in low resource settings. The effectiveness of ART regimens in developing countries at a population level was unknown and the extent to which the PMTCT program reduce vertical transmission rate in the African population is still not clear. The conflicting outcomes are mostly attributed to; the setting, the operational activities and programmatic issues. PMTCT had been described as a poor quality intervention more focused on ARV prophylaxis without provision of continued

HAART and follow up care which complicates the interpretation of its effectiveness and impact in achieving intended goals.

There is not much documented information on the impact of PMTCT on the health status and mortality of HEU children beyond the PMTCT stipulated follow up of two years as they are overshadowed by the treatment and care of HEI infants. A follow-up study was ideal to so as to evaluate the impact of HIV on child survival comparing the difference between HIV exposed and unexposed infants. Outcomes from the study covered PMTCT compliance with stipulated visits and documentation of all observed parameters at each visit, Anthropometrical measurements and morbidity and mortality during the follow up period. This is valuable information regarding trends in compliance, defaulting and health status of the children born under PMTCT initiatives.

1.2.5 BHAMC Broad Objective

To describe five years follow-up of mother and child pairs on a PMTCT program highlighting compliance, loss to follow-up, morbidity and mortality (attrition).

1.2.6 BHAMC Methodology

1.2.6.1 Study Design

A prospective cohort of HIV positive and negative pregnant women enrolled at 36 weeks of gestation and followed up for five years together with their index infant.

1.2.6.2 Study Sites

Women were enrolled from three peri-urban clinics, Epworth, St Marys and Seke North, offering maternal and child health services in Harare. These were the sites where PMTCT interventions were piloted in Zimbabwe in 1999, to assess its feasibility and acceptability.

1.2.6.3 Study Population

Pregnant women at 36 weeks of gestation booked at ANC at the respective study site having gone through VCT under the national PMTCT program.

1.2.6.4 Sample Size

The sample size was calculated in EPISTAT program using the estimated 25.7% HIV prevalence among pregnant women in 2002. A statistical power of 90% was considered to detect a risk difference of 1.6 in the HIV infection groups using a two tailed test with a level of significance observed at 0.005 and allowing an attrition rate of 25% for loss to follow-up. A minimum sample size of 300 positive pregnant women and 600 negative women was required, but the recruitment ended up with a total of 1050 participants. At the end of the study, the final sample size had 466 participants, 227 being positive mothers and 239 negative mothers, after the five year follow-up period.

1.2.6.5 Enrolment Procedure

Pregnant women underwent VCT and routine health education discussions. Study objectives were explained to these women and those willing to participate went through the enrolment process.

1.2.6.6 Inclusion Criteria

Pregnant women who had been post counselled for HIV, received their HIV test results, had consented for both themselves and the index infants to be followed up, and not recruited in any ongoing study and with no bleeding disorders.

1.2.6.7 Exclusion Criteria

Participants were excluded from the study if they were participating in other ongoing studies, had proven sickle cell disease or bleeding disorders, were allergic to benzodiazepine and were on current TB treatment and reported abnormal blood chemistry.

1.2.6.8 Intervention

All HIV positive mothers who would have consented to an HIV test through the national PMTCT program received a single 200mg Nevirapine dose to be taken at the onset of labor, whilst their infants received a single1-2 mg Nevirapine dose within 72 hours of delivery.

1.2.6.9 Data Collection

An interviewer administered questionnaire was used. The tool was pre-tested to the study team and adjustments made for it to give unbiased responses. The questionnaire collected demographic information; knowledge about HIV issues, past and current medical history, obstetric and reproductive health issues.

1.2.6.10 HIV study confirmatory test

A confirmatory HIV test was done using rapid tests on all women regardless of their national HIV test result. Discrepant and false negative and positive results were retested using an ELISA. Women with discrepant HIV test results were re-counseled reassured and were given an option to seek an HIV test from another service provider if they doubted the study result.

1.2.6.11 Follow-up

A locator form was used to document the physical and postal address of the mother; caregiver and next of kin details where home visits were consented to. Where available contact telephone numbers were obtained for follow up purposes.

1.2.6.12 Mothers' Follow-Up

Mothers were followed up according to their expected date of delivery (EDD) to ascertain site where they intended to give birth more so for the HIV positives to establish if they received sdNVP. No NVP syrup was provided to be given to the neonate at home. If the HIV positive women happened to deliver elsewhere they were encouraged to report at the study sites within 72 hours of birth for the child to get NVP. All women were encouraged to breastfeed exclusively for 4 to 6 months before introducing mixed feeding. The HIV positive mothers were encouraged to cease breastfeeding abruptly and introduce formula milk and other replacement feeds.. Follow up continued at 6 weeks where abdominal palpation was done, physical examination, gynecological speculum examination with collection of a Pap smear, collection of venous blood and questionnaire administration. The same was repeated at 4 and 9 months except for the Pap smear. After one year follow up was every 6 months for five years.

1.2.6.13 Children Follow-Up

A birth form was filled in for the neonate recording state at birth; alive/stillborn, Apgar score and anthropometrical measurements. For infants born to HIV positive mothers, time between delivery and NVP ingestion was documented. Cord blood was collected for HIV- DNA PCR analysis. Capillary blood was collected for all children regardless of maternal HIV status for FBC, urea and electrolytes (U&Es). Cotrimoxazole prophylaxis was initiated at 6 weeks to

all HIV exposed infants until their HIV status was established and those found infected continued on it. Follow up visits for children were scheduled at the same time intervals like that of their mothers. At each visit, children had anthropometrical measurements taken, information on the children's health status and feeding practices was sourced from the mothers through a questionnaire. HIV exposed children were screened for HIV using DNA-PCR up to 9 months of age. CD4 count was used as a marker to determine the child's eligibility for HAART.

1.2.6.14 Statistical Methods

Descriptive statistics were used for descriptive analysis namely mean and standard deviation for continuous variables and proportions for the categorical variables. For the infants' mortality rates of children born to HIV positive and HIV negative mothers, survival analysis was used. Cox proportional hazards were calculated and Kaplan-Meier survival curves plotted for both the exposed babies and unexposed babies. For the realities and challenges of PMTCT follow-up, categorical data was analyzed using Pearson's chi-squared test to determine if any association existed between the predictor variable and the outcome. Fisher exact test was used for categorical data and independent student t-test for the continuous data. Multiple logistic regression was used to model the predictor variables with the outcome variable using a p-value less than 0.2 from the univariate analysis.

1.3 CRITICAL APPRAISAL OF THE BHAMC STUDY

The BHAMC study objectives were closely related to the statement of the problem and the specific objectives addressed systematically the various aspects of the problem as defined in the problem statement. The objectives are expected to be specific, measurable, achievable, and realistic and have a time frame, which was observed in the BHAMC cohort. The good objectives set helped the BHAMC researchers to be focused avoid collection of unnecessary data.

A prospective cohort was appropriate to address the research objectives. For example, one of its objectives was to determine the rate of MTCT and risk factors of HIV among babied born to HIV positive mothers really required a follow-up period for the rate to be calculated and the outcome (MTCT of HIV) to be ascertained. The study could have been done retrospectively if exposure and outcome had already occurred but re-call bias would have been a major threat to the validity of the results.

The choice of their study design was based on their research question, available knowledge about the problem and the resources available. A single general cohort was good since it categorized the members into different exposure groups, one being an internal comparison group. Cross sectional studies might have been opted for but only the prevalence of HIV infected babies could have been measured not rate ratio.

Despite the major strengths of a prospective cohort, loss to follow-up of study participants is a major constraint. Study participants are lost due to drop-out, migration, deaths or loss to follow-up. Non-response or non-participation is usually observed in prospective studies and this distorts the validity (both internal and external) and reliability of results. Participants are lost due to quitting, migration, deaths or loose tracking. These constraints were controlled by

using a large sample size and incorporating the non-response rate (attrition rate) during sample size determination.

Three study sites namely Seke, Epworth and St Mary's study settings in Harare were used because they were the ones which were pilot tested on PMTCT interventions in Zimbabwe when they were launched in 1999, to assess its feasibility and acceptability. The main challenge which was most likely was of getting HIV positive and negative mothers concurrently recruiting them into the study. Selection of women who were under the PMTCT program only led to selection bias. Other group of women who had the same type of health seeking behaviour were not enrolled into the study. This limited the representation to the general population.

The sampling technique used was convenience sampling. This had the advantage of obtaining study participants especially the HIV positive ones but led to selection bias since participants did not have an equal chance of being selected. They recruited the willing ones only who had come for their ANC visits and no sampling frame was used. Simple random sampling could have been done were each study participant has an equal chance of being selected into the study. It is simple since it uses a sampling frame, for example ANC attendance register, and reduced selection bias, sampling error (standard deviation/root of sample size) can be measured and the design effect is one.

The study exposure, HIV status was ascertained at baseline and the inclusion and exclusion criteria were rigid. Since it was a prospective study, it was less susceptible to selection bias because the outcome was not known. Ascertaining of the exposure status was confirmed using laboratory tests not verbal or use of records since there were some women who had gone for VCT under the national PMTCT program. This was important so that misclassification bias of study participants would be minimized. The sample drawn was to be

a representative of the study population. Representativeness of the sample would result in the results being inferred or extrapolated to the target population (pregnant mothers).

The sample size was calculated using EPISTAT program. Sample size calculation depends on a number of factors like variability in the target population, desired precision and confidence of the estimate and feasibility. The factors which were used in the BHAMC study are prevalence of the HIV positive pregnant women of 25.7% which was available from literature; a power (1-beta) of 90% was used which was high so as to lower the probability of rejecting a false null hypothesis (beta) and this power is practically considered sufficient in research studies or 80% power.

A high power results in calculation of a large sample size. A two-tailed level of significance (alpha) of 0.005 as the probability of making a type I error was used which results in a larger sample size being obtained as compared to using a one-tailed sample size. A 99% confidence level was used as a precision though often researchers use a 95% confidence level. Power and precision are set at the design stage by the researcher. A risk difference (risk in exposed- risk in unexposed) of 1.6 was used for the measure of association and allowed an attrition rate of 25% to control for loss to follow-up or non-response since this was a follow-up study. The ratio of exposed to unexposed in the planned study was two.

A representative sample was most ideal to be more informative and able to reach the set objectives allowing internal and external validity to be met. Their calculation gave a minimum sample size of 900 participants but 1050 were recruited at the end. This was an appropriate method of calculating the sample size.

The data collection tool used was an interviewer administered questionnaire to collect the demographics and other study variables from the participants. The questionnaire was translated from English to Shona then back to English again so that the tool becomes standard

in both languages. This was important to aid communication between the interviewer and the study participants and making it possible to get same required information from both literate and illiterate study participants. The tool was pretested to team mates before the study starts to see if it was going to collect the information it was supposed to collect. Any observed potential problems were corrected.

An interviewer administered questionnaire also permits clarification of questions hence appropriate answers consistent with the question are collected. It has a higher response rate than self-administered questionnaire though the presence of the interviewer can influences the responses from the participants. Another limitation is that, reports of events may be less complete than information gained through observations. Some information came from medical professionals (nurses/midwives, gynecologists, pediatricians) through physical examinations of participants.

Index child measurements were taken during the study period. Their weight, height and head circumference were recorded at birth and on every subsequent visit they made in the follow-up period. Baby's underweight and stunting variables were collected. Standard scale units were used for the measurement variables. In research, bias cannot be avoided but can be minimized. Observer bias is most likely to result in taking measurements. This results in a systematic difference in which information is sort from participants. Standardized measurement instruments were used in this study to minimize bias and qualified personnel took the measurements. The measurement instrument was administered equally to the whole cohort.

Appropriate statistical analysis procedures were done. Categorized data were analysed using Chi-square test and Fisher exact test. Independent t-test was used on continuous data and logistic regression model was used to get the measure of effect of the study. Missing data handling methods are silent in this study.

Limitation of the study was mainly the drop-out rate during the follow-up period though they had included attrition rate is their sample size calculation. There was need to ensure cooperation at each time of those who participated at baseline since it is difficult to replace dropouts or are dead with others who did not participate at the previous measurement (attrition). Generally, if loss to follow up is large, 30-40%, the validity of results is violated. Loss to follow-up may be differential between the cohort groups, that is, loss to follow up might be high in the exposed group than the unexposed group. The effect of differential loss to follow-up is that it results in biased results. It can over-estimate or under-estimate an existing association between an exposure and outcome.

Inconsistence in follow-up of participants limited generalizability of the study findings and HIV test for exposed children were not done as per standard methods. During follow-up lay counselors were used instead of the professional social workers who had an extra knowledge of the study subject. The same counselor was used throughout so as to maintain the built in bond with the study participants though it had a disadvantage of these counselors developing lazy attitudes.

Ethical considerations were observed before the study commenced. The study received approval from the recommended boards.

1.3.1 Data Quality

Despite the challenges of dropouts and attrition faced in longitudinal studies, efforts were made to collect complete data in the BHAMC study. The BHAMC dataset has some missing variables and there is no consistence in repeatedly collecting some variables. Abiding to the research protocol was a challenge since the six months follow-up was not done for five years as was suggested.

Variables on child morbidity were not well collected. Only three time points have the variables of which a well described trend would result if the information was collected for five years. Children's CD4 counts were not done and HIV tests were done at a later stage in the cohort. Despite all these limitations, the BHAMC study gives a platform to study morbidity outcomes in exposed and unexposed index babies.

The presents of specific illnesses which the child might have suffered from as reported by the mother can be used to give a proxy morbidity outcome. The time frame in which the data was collected was up to nine months so results can be based on time specific follow-up points. Some of the index children died during birth. This results in a smaller sample size being used in analysis.

Based on these gaps in the BHAMC methodology, this current study proposes to look at the effect of maternal HIV exposure on children morbidity, particularly HEU and HUU using generalised linear mixed effects model technique. The main reason of using this technique is that, when data is collected longitudinally, missing data may result and correlation of responses needs to be controlled for.

1.4 RESEARCH PROBLEM

The main goal of HAART is to have an HIV free generation and prolong life to those who are HIV infected. Children exposed to HIV (HEU) are of less public health concern once deemed HIV negative as compared to those who would have tested HIV positive (HEI). Due to the advent of HAART in PMTCT programmes which has reduced the vertical transmission rate to as low as 1%, there is a swift increase in the HEU population but their care after delivery is limited to increase their survival. Studies done in Zimbabwe reported a high mortality rate of 9.2% (n=3135) among the HEU and mortality rates in HEU infants were also at least twice the mortality risk of HUU infants[8]. The HEU children tend to report more to hospitals as compared to their HUU counterparts. A description of the morbidity in HEU children in terms of specific illnesses and / or hospital admissions and identification of possible factors which are associated to morbidity in HEU and HUU will help in declining the currently reported child mortalities of 84 deaths per 1000 live births. Modeling of these factors using generalized linear mixed effects model can help in the estimation of risk contributed toward HEU morbidity by interpreting subject- specific regression coefficients.

CHAPTER 2

2. LITERATURE REVIEW

2.1 Other Studies

The issue of HEU babies is a matter of concern since not much has been documented particularly in Zimbabwe and there is a rapid increase of this group in the country. A retrospective study in Belgium found that, 77% of HEU babies were hospitalized during the first year of life and 48% of these children were admitted in hospital for an infectious disease with 54.82% of them suffering from serious infections. Furthermore, the study also observed that HEU babies were almost twenty times more at risk of developing group B streptococcal disease compared to those born to uninfected mothers (HUU) [16].

In Sub-Saharan Africa, a number of studies have been conducted to explain the high morbidity rate in the HEU infants. A prospective study was performed in Cape Town, Western Cape in South Africa from 2004 to 2008 at a surgical centre. Broadly, the study looked at the morbidity outcomes in children undergoing a surgery. This study concluded that HEU have a higher risk of developing complications and mortality (5.2%) after surgery as compared to HUU children (0%). However, their risk was lower than that of the HEI children[17].

A study in Malawi on cessation of breastfeeding found out that in HEU children, breastfeeding cessation was associated with acute morbidity events. The adjusted rate ratio at 9-12months for illness and / or hospitalisation was 1.66 for non-breastfeeding and breastfeeding infants. The Poisson regression model was used to assess the association between non-breastfeeding and morbidity at each mutually exclusive interval controlling for other factors.

Other risk factors that have been found to be associated to the differences in morbidity in HEU and HUU children include maternal death and maternal low CD4 counts. Maternal death was found to be a risk factor of persistent diarrhoea in HEU children but birth weight, gestational age at birth and age at weaning were not[14].

In Zimbabwe, much work has been done on the mortality in HEU children. The Colloquium Aboard conference reported a death rate in HEU as three-folds higher than in the HUU children[13]. This is similar to the study done by Marinda et al (2007) which showed that mortality rates in HEU infants were at least twice the mortality risk of HUU infants [7].

From the Better Health for African Mother and Child cohort, a study on the Effect of Maternal HIV exposure on Infant Mortality observed that at five years, the HIV exposed mortality rate was 53 per 1000 person years and HIV exposed uninfected infants had a mortality rate of 15 per 1000 person years. The mortality rate for the HIV infected children was 112 per 1000 person years compared to 21 per 1000 person years for the exposed uninfected infants[18] .

The Zvitambo study group found that morbidity was high among HEI infants. The HEU infants had a higher morbidity as compared to HUU. Sick clinic visits were 1.2 times more common among HEU infants as compared to HUU, and were significantly higher for all mothers with CD4 count less than 800 cells per micro litre[19]. This study recruited its participants between November 1996 and 2000, and this was before the availability of HAART. No description of the common illnesses was stated which the children presented with in the rural settings.

2.2 Review of Longitudinal Data Analysis Techniques

The BHAMC database was collected longitudinally, and consists of repeated measurements over a variable for the number of follow-up years.

Longitudinal data have important characteristics. They are repeated measurements obtained from a single individual at different points in time. Observations made on one individual over time are positively correlated. Failure to take this correlation into account in the statistical analysis will lead to incorrect estimates of the sampling variability and incorrect inferences [20, 33, 37]. Longitudinal data have also a temporal order, the measurements being taken in an ordered time sequence.

Longitudinal studies have the outcome variable measured repeatedly over time and balanced or unbalanced designs results. Advantages of longitudinal data are that, they allow investigation of events that occur in time; essential to the study of temporal patterns of response to treatments, permit more complete ascertainment of exposure histories in epidemiological studies and reduce unexplained variability in the response by using subject as his or her own control. A number of models have been developed to give reliable results from longitudinal data[21].

2.2.1 Alternating Logistic Regression (ALR) and Probit Model

Marginal models for multivariate binary data permit separate modelling of the relationship of the response with explanatory variables and the association between pairs of responses. ALR is an analytic approach used for simultaneously regress the response on explanatory variables as well as modelling the association among responses in terms of odds ratios. The model overcomes the limitation about the longitudinal associations within the repeated outcomes. ALR models the association between the outcomes at various time points. The response

variable is binary (dichotomous response) hence considers the association between pairs of responses with the log odds ratios instead of correlations [22].

Probit model is a regression model where the independent variable has a binary outcome. The model estimates the probability that an observation with particular characteristic will fall into a specific one of the categories. It estimates probabilities that greater than half of the observations are treated as classifying an observation into a predicted category. This model is considered as a binary classification model. It assumes the error terms to be independently distributed according to the standard normal distribution [20].

2.2.2 Generalised Estimating Equations (GEE)

Marginal models or population-average models are an extension of the general linear models using quasi-likelihood estimation [23]. A known transformation of the marginal expectations of the outcome is assumed to be a linear function of the covariates. They are relevant when the main focus of a study is investigating the effect of covariates on the population mean and not necessarily at individual level [24]. Marginal models are considered more flexible than classic generalized linear models since they can handle unbalanced longitudinal data with repeated measurements and therefore they can handle as well some patterns of missing data[25]. Marginal models do not require precise specification of the outcome distribution and accommodate time-dependent covariates [26]. Different link functions can be used in these models which converts the expected value to be unrestricted linear predictor form. The link functions are *identity* for continuous data; *log link* for count data and binary data; and *logit* link for binary data[27].

In marginal models it is useful to specify the distribution of the outcome variable so that the variance can be calculated as a function of the mean. GEE treat correlation structures as a "nuisance" hence not modelled. The correlation structures can be independent, exchangeable,

autoregressive or unstructured among others. An important step in choosing a specific correlation structure is to find the simplest structure which fits the observed data well [20]. A useful feature of the GEE model is that the estimators are robust to departures from the true correlation patterns. A loss in estimator efficiency can occur but this loss decreases as the sample becomes larger [28].

2.2.3 Generalised Linear Mixed Effects Models (GLMM)

These models extend the GLMs by the inclusion of the random effect in the predictor. The random effects are used as an approach to account for within and between subject associations. These conditional models allow a subject of the regression coefficient to vary from one individual to another. The introduction of the random effect produces a greater degree of conceptual and analytic complexity relative to marginal models or to random effects in linear models[29].

GLMM is a regression model with randomly varying intercept but can also include poison and other distributions. The model posits that there is natural heterogeneity in individuals' propensity to respond positively that persist throughout all binary response obtained on any individual. GLMMs are most useful when the main scientific objective is to make inferences about individuals rather than the population averages and for modelling the dependence among the response variables inherited from longitudinal or repeated studies for accommodating over-dispersion among binomial or Poisson responses. These models address questions that are concerned with mean changes in the mean response for any individual and the impact of covariates on these changes. Model inferences are based on the maximum likelihood function[30, 37].

Model diagnostics and goodness of fit test are more limited in GLMM as compared to linear mixed models. Since GLMM are likelihood based model selection is done using are

likelihood ratio tests, the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) to compare different models. GLMM require maximum likelihood solutions so all tests and comparisons are under maximum likelihood [31, 32].

Model diagnostics which is an important part in model building process involves residual analysis, outlier analysis, checking for normality in the distribution and model validity. Residual analysis is used to assess model fitting. Model validity completes the fitting process.

The literature review of statistical methods suggests a number of aspects that inform the methodological choice of this study on the morbidity trends in HEU and HUU children at nine months follow-up period. The choice of the model is based on the type of dataset available and the type of question to be answered. Since this current study has a longitudinal dataset and looking at binary outcome, the generalised mixed effects model will be fitted. Possible risk factors that are associated with morbidity in HEU children can be modelled through a generalised linear mixed effects model so give subject specific inference.

2.3 JUSTIFICATION

Among the mortalities in under-fives, 21% of them are indirectly related to HIV and HIV is the second highest contributing factor. The country is still far from achieving the MDG 4 target of 24 deaths per 1000 live births for under-five mortality rate since it is as high as 84 deaths per live births [2]. With the increase in the efficiency of PMTCT program, there is a rapid increase in the HEU population. There is need to describe HEU children morbidity in relation to HIV exposure so target interventions that reduce the mortality in under-five year old children and improve child survival. A comparison of the morbidity in HEU children with the HUU children will describe their health status and provide clear pictures of the possible determinants in which these two groups differ.

Fewer studies have explained the differences in morbidity between HEU and HUU children. So an understanding of factors contributing to HEU morbidity patterns plays a significant role in targeting, allocating and mobilising resources especially in the PMTCT prongs. The HEU group needs a quality of care so reduce any morbidity incidences in their group. Moreover, knowledge of these morbidity conditions may assist in providing appropriate clinical care and designing potential interventions. The use of a GLMM technique can help in providing meaningful results since the technique handles missing data.

2.4 RESEARCH QUESTION

What is the effect of maternal HIV exposure (during pregnancy, birth) on the morbidity in HEU and HUU?

2.5 OBJECTIVES

2.5.1 Broad Objectives

- To assess morbidity trends and factors associated with change in morbidity between HEU and HUU children in a nine months follow-up period.

2.5.2 Specific Objectives

- To describe the socio-demographic characteristics of HEU and HUU children.
- To compare the differences in the reported morbidity conditions and morbidity rates among HEU and HUU children.
- To determine the association between maternal HIV exposure with morbidity in HEU and HUU children.

2.6 HYPOTHESIS

There is no difference in morbidity between HIV- exposed uninfected (HEU) children and HIV- unexposed uninfected (HUU) children in the BHAMC cohort.

CHAPTER 3

3. METHODOLOGY

3.1 Data Sources

This is a secondary data analysis on a subset of the BHAMC cohort which includes only HEU and HUU children.

3.2 Sample Size

The BHAMC study enrolled a total number of 1050 participants. Of these, 479 were HIV positive and 571 were HIV negative pregnant mothers. From these mothers 470 HIV exposed children and 571 unexposed children were born. From the exposed children, there were 5 multiple births hence 469 were considered. From the 469 children 70 were infected (HEI) while 399 were uninfected (HEU). For this analysis, a total of 970 participants were considered.

3.3 Definition of Study Variables

The main study outcome variable is morbidity. Morbidity was defined as any illness and /or admission in this study. HIV related illnesses reported by the mother included fever, cough, diarrhoea and ear discharge were extracted from the master database to generate the binary morbidity outcome variable. These variables were measured longitudinally.

Explanatory covariates are maternal HIV status, breastfeeding and mother on ART. The study is designed to control for age of the child, weight, sex and HIV exposure of the child (HEU or HUU). Table 1 presents a description of the variables extracted for analysis and how they are coded.

Table 1: The Data Dictionary of the Database

Variable Type	Variable Name	Variable Description	Variable format	Variable code
Outcome	Child Morbidity	Any illness and/or admission	Coded	1-child sick 0-child not sick
Study	Hivlab	Mother's HIV status	String	positive negative
	Mother.arv	ARV mother took during pregnancy	Coded	0-No 1-NVP 2-AZT
	Hivbabydefi	HIV status of the child throughout the follow-up period	String	positive negative
	Breastfe	Child breastfeeding	Coded	0-no 1-yes
	b.wgt	Baby weight	Continuous	
	b.leng	Baby length	Continuous	
	b.underw	Baby underweight (wt <=3rd percentile)	Coded	0-no 1-yes
	Parity	The number of children the mother has	Coded	
	Mother CD4 counts	Severity of disease	Discrete	

3.4 Data Extraction

A subset of the variables collected in the BHAMC study was used. The main study had about 250 variables but for our analysis fewer variables were used (Table 1). Each study participant was identified by a unique identification number in all the data files. The main data file for this study was generated after merging data from single files using a unique identification code.

3.5 DATA MANAGEMENT

The BHAMC database serves as the source for the data subset on which the current research is based on. The database consists of a series of electronic data sets for each follow-up point. Data has been collected already through a questionnaire and laboratory tests, and was entered and store electronically using SPSS software.

Data management is the process which involves data collection, coding, entry, cleaning, data analysis, and storage. The process of data management is done to ensure maximum quality control of data which results in better quality of data and results. Data has been collected already through a questionnaire and laboratory tests, and was entered and store electronically using SPSS software.

3.5.1 Data Cleaning

When multiple data sources are integrated like the BHAMC data files, data cleaning increases since redundant data is contained in different representations. In order to have access to consistent data, elimination of duplicate information was necessary. Data screening was done were irrelevant variables were dropped from the dataset. This was done to remove excess data like mothers' information and check for outliers on continuous variables like birth head circumference, inconsistence and any observed strange patterns. Data diagnostics were done to check for errors, true extreme and true normal on the variables.

3.5.2 Data Coding

Variables without codes were coded in Stata 11 and some used the codes from the main cohort study database. The main outcome (morbidity) was coded as 0=healthy child and 1=ill child since it was a dichotomous outcome. Some variable were changed their format from string to integers like maternal HIV status was changed from positive or negative to 1 or 0

respectively. Generated variables were labelled for easy identification and interpretation of results. Follow-up variables were renamed for easy transforming from wide to long. Coding was done when the dataset was both in the wide and long format. The long format was used since this is recommended for longitudinal studies.

3.6 Statistical Analysis Methods

Exploratory Data Analysis (EDA) is an approach for data analysis that employs a number of techniques to maximise insight into a dataset, extract important variables, detect outliers and anomalies, test underlying assumption and develop a parsimonious model. This approach provides summary statistics; graphs for example scatter plots, histograms to visualize data patterns. Insight gained leads to the most appropriate analysis technique to be used.

Descriptive statistics were used to quantitatively describe the main features of a dataset mainly the measure of central tendency (mean) and the measure of variability (standard deviation) for weight, length and head circumference at delivery. If the data is skewed, median and quartiles are reported.

Bar charts were plotted for categorical variables like child HIV exposure, that is, HUU or HEU. Line graphs were generated to explore and visualise patterns of morbidity change over time. Loss to follow up chart was done for the nine months follow-up period. The *Z* test for difference in two sample proportions was used to compare the difference in morbidity outcomes between the HUU and HEU. Significance tests were set at 0.05.

To model change in morbidity between the HUU and HEU, the mixed effects logistic regression model was used. These models have a fixed (non-random) effect and a random effects which, accounts for the within subjects association via the introduction of a random effect term in the model. The model allows a subset of the regression coefficient to vary

randomly from one individual to another. The random effects reflect natural heterogeneity due to many unmeasured factors. Conditional to the random effect, the responses for any individual are independent observations belonging to the Bernoulli distribution since the response variable is binary hence this is the "conditional independence" assumption. The conditional mean of the response variable depends upon fixed and random effects via a linear predictor by a logit link function. The single random effect is assumed to have a univariate normal distribution with a mean of zero and some variance which depends on the exposure group.

The following model building steps done where:

- Fixed (covariates and exposure variable) and random effects (individual) were specified.
- The random error was assumed to have a multilevel normal distribution and a link function (logit link was specified).
- Variances of the data (transformed by the logit link function) were checked and it is supposed to be homogeneous across categories[36].
- A full model containing all explanatory variables (fixed effects), main effects and the random effects was fitted.

Backward selection by statistical significance testing of regression coefficients, with p-value at 0.05 was done. All variables significant were left to give the final model. The comparison techniques to select the final model used were the likelihood ratio test. The regression coefficients were interpreted as the difference in log odds associated with a unit change in the corresponding covariates and the exponential regression coefficient as an odds ratio. This was due to the fact that the coefficients of this model are conditional on a child. Wald tests and Z tests were used to make conclusions on other regression coefficients. Wald test for linear

combinations of regression coefficients can be used to test corresponding multiplicative relationships among odds of different covariates values.

3.7 Model Diagnostics

Model diagnostics and goodness of fit test are more limited in GLMM. Since GLMM are likelihood based model selection is done using are likelihood ratio tests, the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) to compare models. GLMM require maximum likelihood solutions so all tests and comparisons are under maximum likelihood.

The similar procedure was done to single morbidity outcomes, that is, illness and admissions using GLMM before combing the overall outcome.

3.8 Ethical Consideration

Ethical consideration was addressed so as to observe and maintain confidentiality of study participants. Ethical approval was sorted from Joint Research Ethics Committe (JREC) and permission to use the dataset was sorted and granted by the owners. Both copies of ethical clearance and dataset approval have been attached to this document.

CHAPTER 4

4. RESULTS

4.1 Exploratory and Descriptive Results

The BHAMC study enrolled a total number of 1050 pregnant mothers. Of these mothers, 479 (45.6%) were HIV positive and 571 (54.4%) were HIV negative. From the HIV positive mothers 470 (98%) HIV exposed children and from the HIV negative mothers 571 (100%) unexposed children were born. From the exposed children, 70/470 (14.6%) were HIV infected (HEI). For this analysis we considered a total of 970 children, 409 (41.7%) HIV exposed uninfected (HEU) and 571 (58.3%) HIV unexposed uninfected children (HUU).

The cohort was followed from birth to 9 months. The number of participants decreased for each group during the follow-up period of 9 months as some of the participants were lost to follow-up. At the end of 9 months 363 (63.5%) of the HUU and 220 (53.8%) were lost to follow-up (Figure 1).

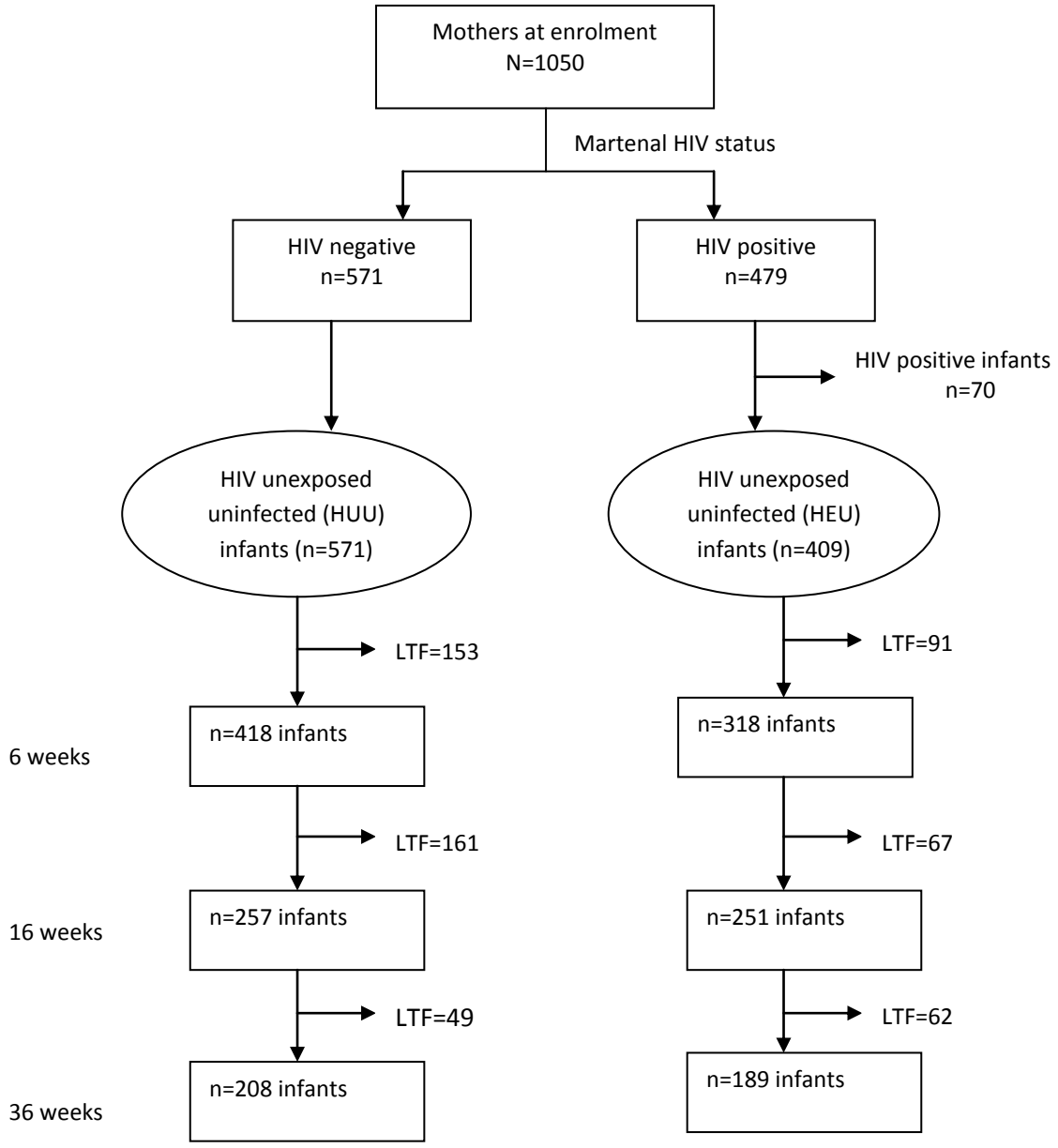

Figure 1: Study flow chart of participants in the nine months follow up period

Table 2: Description of Anthropometric measures and Demographics at Delivery by group

Variables	HUU Mean (s.d)	HEU Mean (s.d)	P-value
Weight at Delivery(mg)	3081.0 (409.7)	2997.9 (423.8)	0.001*
Length at delivery (cm)	49.1 (2.2)	48.8 (2.6)	1.000
Head circumference at delivery (cm)	34.3 (2.0)	34.3 (3.4)	1.000
Apgar	9.4 (0.8)	9.3 (0.8)	1.000
Enrolment Age(days)	12 (4.6)	10 (3.7)	0.018*
Sex Males {(n (%)}	207 (55.8)	236 (58.1)	0.512
Females {n (%)}	164 (44.2)	170 (41.9)	0.512

Significant difference=*

Table 2 shows anthropometric measurements at delivery. Generally, HUU have high mean values for most of the measurements and less variability as compared to the HEU. There was no significant difference in child length; head circumference, apgar and sex between the HEU and HUU. A significant difference was observed in age and weight at delivery between the groups.

Table 3: Overall relative and absolute Distribution of illnesses for nine months by group

Illness type	HUU n (%)	HEU n(%)	P-value
Fever	282(32.3)	206 (27.8)	0.052
Skin rash	193 (22.1)	129 (17.5)	0.021*
Cough	353 (40.4)	245 (33.2)	0.003*
Oral thrush	8(1.8)	13(3.9)	0.196
Diarrhoea	184 (21.3)	15 (8.1)	0.050
Convulsions	21 (2.4)	11 (1.5)	0.190

Significant difference=*

From Table 3, a higher proportion of illnesses were reported from HUU group as compared to the HEU group. Oral thrush was high in HEU than HUU, but not statistically different. Fever and cough were the most illnesses experienced in the two groups but significantly higher in HUU. Convulsions were the least condition to be experienced in the cohort with proportions as low as 2%. A significance difference was observed between cough and skin rash with p-values less than 0.05. Fever and diarrhoea were at the margin with a p-value of 0.05.

Table 4: Overall relative and absolute Distribution of Children by Possible risk factors to morbidity

Possible factors	HUU n (%)	HEU n (%)	P-value
Breastfeeding	415 (95.4)	295 (87.8)	<0.001*
Deceased mother	9 (0.5)	36 (2.9)	0.002*
At least one child	531 (93.0)	356 (87,0)	0.002*

Significance difference=*

Looking at other factors that have a positive effect toward a child getting sick (Table 4), higher proportions (above 80%) have been noticed in the HUU group as compared to the HEU though the HUU have a higher proportion in breastfeeding. Both groups were breastfed during the follow-up and their families had more than one child. Deceased mothers are higher in the HEU group as compared to the HUU who had a proportion as low as less than 1%. There is significance difference in all factors between the two groups.

Table 5: Descriptive Analysis results of Morbidity Outcomes

Time	Morbidity Outcomes	HUU n(%)	HEU n (%)	P-value
Six weeks	Illness	202 (54.0)	124 (43.7)	0.007*
	Admissions	8 (2.1)	4 (1.4)	0.471
	Overall	204(54.6)	125(43.7)	0.006*
Four months	Illness	141 (61.6)	132(55.3)	0.219
	Admissions	140 (61.1)	131 (55.5)	0.678
	Overall	8(3.5)	10(4.2)	0.386
Nine months	Illness	140 (61.1)	131 (55.5)	0.039*
	Admissions	8 (3.5)	10 (4.2)	0.003*
	Overall	142(71.0)	112 (61.9)	0.059

Significant difference=*

Within this study, morbidity was defined as any illness or hospital admission that the child had experience during the follow-up period (Table 5). The HUU children experienced more morbidity outcome as compared to HEU during the first and last time points. Notably is a sharp decrease in the morbidity outcomes, proportions less than 5%, in both groups at sixteen weeks. Significance difference where observed at admissions at nine months, illnesses and overall illnesses in six weeks.

Table 6: Absolute and Relative Distribution of the Morbidity outcomes by group

Morbidity outcomes	**HUU** **n (%)**	**HEU** **n (%)**	**P-value**
Illness	532 (61.2)	386 (52.7)	0.001*
Admissions	24 (2.8)	31 (4.2)	0.11
Illnesses and/or Admissions	536 (61.7)	389 (53.2)	0.001*

Significant difference=*

Generally, the HEU children had less morbidity outcomes as compare to the HUU children (Table 6). There is a sharp decrease in morbidity outcomes of less than 10% during the second follow-up visit for both groups. There is a significant difference between the HUU and HEU groups in illnesses and the overall morbidity for the nine months. There were no significant differences in admissions within the two groups.

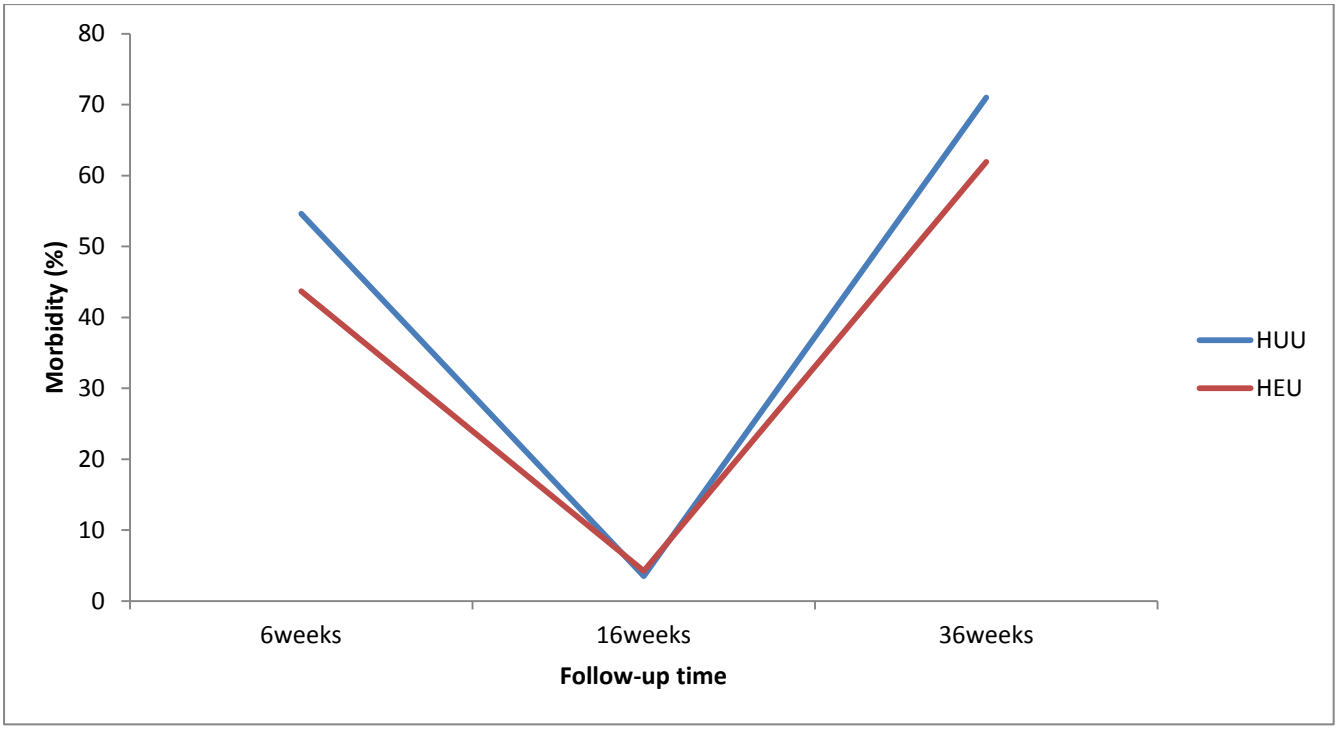

Figure 2: An Overall Morbidity Trends for nine months by group

The morbidity trends between the two groups follow a similar close pattern (Figure 2). From six weeks the overall morbidity outcome deceases to below 10% at sixteen weeks and peaks up again at thirty six weeks. The HUU group has a higher proportion of the outcomes as compared to the HEU group. At 5% level of significance, there is a significance difference between the two groups during the whole nine months follow-up period.

4.2 Longitudinal Model Analysis

We fitted a random effects logistic regression model with a time linear spline with a knot at 16 weeks to account for the difference in change in morbidity before and after 16 weeks as indicated in Figure 2. Since the objective was to assess the effect of HIV exposure (HIV positive mother and HIV negative mother) on morbidity the following base model was considered.

Table 7: Random Intercept Logistic Regression Results for Morbidity outcomes over time

	Estimates	SE	Z	95% CI
Intercept (β_o)	0.55	0.167	3.25	0.21 - 0.87
Exposure(HEU)	-0.22	0.23	-0.93	-0.68 - 0.24
Time1 <=16 weeks	-0.052	0.033	1.61	-0.12 -0.12
Time2 > 16weeks	0.067	0.023	2.95	0.022 – 0.11
Time1 by Exposure	0.031	0.033	0.96	-0.033 - 0.097
Time2 by exposure	-0.053	0.048	-1.11	-0.15 -0.87
-2 log L= 1979.29				

SE=standard error

From the baseline model (Table 7), there is a 0.55 average child-specific intercept for the log odds of having a morbidity outcome as a function of maternal exposure to HIV in the thirty-six weeks follow-up period for the unexposed (HUU) group. In the HUU group, the estimated log odds of outcome decreases by 5.2% per week during the first sixteen weeks and increases by 6.7% per week after sixteen weeks.

In the exposed group (HEU), the log odds of morbidity outcome changes positively by 3.1% per week during the first 16 weeks giving a downward slope of 0.02 and negatively changes by 5.3% per week after 16 weeks giving an upward slope of 0.014. The estimated difference in the slopes between the two groups is the effect change in the nine months follow-up period. The effect change or exposure effect during the first sixteen weeks is 0.03 and -0.05 after sixteen weeks. Both exposure effects for the two time points are significant at 5%.

The estimated odds for a subject in the unexposed group are multiplied by 0.95 every week during the first sixteen weeks and 1.07 after sixteen weeks. The odds for a child in the exposed group are multiplied by 0.98 every week during the first sixteen weeks and 1.01 after sixteen weeks. The odds decrease during the first sixteen weeks is estimated to decrease by 5% in the unexposed group and 2% in the exposed group. The log likelihood for the baseline model is 1979.3.

A model with the other covariates, considered potential risk factors for morbidity outcomes was run and is shown below:

Table 8: Logistic Regression Results with covariates for morbidity outcomes over nine months

	Estimates	SE	Z	95% CI	
Intercept	1.04	0.470	2.2	0.120	1.964
Exposure(HEU)	-0.255	0.237	-1.08	-0.721	-0.210
Time1 <=16 weeks	-0.637	0.035	-1.84	-0.131	0.004
Time2 >=16weeks	0.062	0.023	2.73	0.018	0.107
Time1 by Exposure	-0.057	0.048	-1.20	-0.150	0.036
Time2 by exposure	0.036	0.033	1.08	-0.029	0.101
Family	-0.333	0.238	-1.40	-0.801	0.133
Breastfeeding	-0.376	0.311	-1.21	-0.986	0.234
Age	0.041	0.040	1.02	-0.038	0.120
SD(constant)	1.124	0.14	--------	-------	-------
-2 log L= 1973					

From the full model with demographic covariates, there is a 1.04 average child-specific intercept for the log odds of having a morbidity outcome as a function of maternal exposure to HIV in the thirty-six weeks follow-up period for the unexposed (HUU) group. In the HUU group, the estimated log odds of outcome decreases by 6.4 % per week during the first sixteen weeks and increases by 6.2% per week after sixteen weeks.

In the exposed group (HEU), the log odds of morbidity outcome further decrease by 5.7 % per week during the first 16 weeks giving a downward slope of 0.12 and positively changes by 3.7 % per week after 16 weeks giving an upward slope of 0.098. The estimated difference

in the slopes between the two groups is the effect change in the nine months follow-up period. The effect change during the first sixteen weeks is -0.06 and 0.04 after sixteen weeks. Both exposure effects for the two time points are not significant at 5%.

The estimated odds for a subject in the unexposed group are multiplied by 0.94 every week during the first sixteen weeks and 1.06 after sixteen weeks. The odds for a child in the exposed group are multiplied by 0.89 every week during the first sixteen weeks and 1.1 after sixteen weeks. The odds decrease during the first sixteen weeks is estimated to decrease by 6% in the unexposed group and 11% in the exposed group.

The estimated difference in the slopes of time between the two groups is the effect change with time. During the first 16 weeks, the estimated difference in the slopes was -0.057. This is the effect change within the first 16 weeks of follow-up. There was a positive effect change after the first 16weeks of 0.036 between the two groups. The effect changes are not statistically significant at 5%.

For the child-specific regression coefficients of covariates, the child's log odds of morbidity increases by 0.04 for a day increase in the age of a child whilst holding other child's covariates fixed. The 95% confidence interval of child's age lies between -0.038 and 0.12; this is statistically insignificant. The child's log odds of morbidity decrease by 0.38 for a breastfed child as compared to a non-breastfed child whilst holding other child's covariates constant. A child being breastfed has a 95% confidence interval which lies between -0.99 and 0.23 and is not statistically significant. The child's log odds of morbidity decreases by 0.33 if a child belongs to a family with more than one child as compared to a child belonging to a family with only one child holding other child's covariates constant. The 95% confidence interval for the family size lies between -0.8 and 0.13; this is statistically insignificant.

The estimated standard deviation of the random intercept was 1.12 with a confidence interval of between 0.874 and 1.447. This in between subject intercepts reflects the variation in the propensity of exposure for morbidity outcomes in the children.

4.3 Model Diagnostics

Testing whether the random effect is significant or not, a likelihood ratio test had a chi-squared value of 35.5 and a very low p-value (<0.001). We reject the hypothesis that there is no cross subject variation in the intercepts in this model. We conclude that there is significant variability between intercepts. The fitted model has a log likelihood of 1972. There is an appreciable heterogeneity in the individual trend between the two groups.

CHAPTER 5

5. DISCUSSION

The cohort had 980 participants of which, 409 (41.7%) were HEU and 571 (58.3%) were HUU children. The demographics information was similar in both groups. This gave a proper baseline for comparison between the HUU and the HEU. Significant difference was observed in weight at delivery and age at enrolment whilst the other variables were the same across the two groups.

The overall marginal distribution of morbidity outcomes behaviour over time is that, 41.7 % of the children population had no outcome of morbidity during the entire follow up period. 71.2% of the specific subjects had at least one morbidity outcome during the three follow-up visits. 76.6% responses on subjects reporting illness or admissions at least once had a morbidity outcome.

The transition proportion from one time point to the next shows that it was likely that once a child had never had any morbidity outcome, that child would experience an outcome during the follow up. There was an increase from 28% to 72% of morbidity outcomes in those who had once reported an illness or been admitted.

A number of illnesses were considered in this study namely fever, diarrhoea, cough, skin rash, convulsions and oral thrush. These conditions are HIV related conditions. Fever and cough were reported more because are common signs for a number of other conditions like pneumonia in the early infant life. The HEU reported more on oral thrush comparing with HUU among all the reported illnesses. Oral thrash usually is an illness that is seen in immune compromised individuals mostly. Since the HEU had not developed much their immunity, they were most likely to suffer from such conditions due to their exposure to HIV. For all the

illnesses, skin disease and cough gave a statistically significance difference between the two groups.

A number of possible risk factors can explain morbidity outcomes in children. Some of them include breastfeeding, polygamy and whether the mother is alive or not. Non-exclusive breastfeeding can result in an increase in child morbidity outcomes like diarrhoea. Non-hygienic feeding practices are also associated with child morbidity outcomes [11, 13]. There was a statistical significant difference in these factors among the HUU and HEU.

The HUU had a higher percentage of breastfeeding compared to HEU hence they also reported more diarrheal cases than HEU. Breastfeeding was less in the HEU maybe because of the maternal HIV status. HIV positive mothers were encouraged to exclusively breastfeed during the follow-up period or give formula milk, so they might have done it more consistently as compared to the HIV negative mothers who might have ended up doing mixed feeding practices. This can explain the observed morbidity outcome in the HUU than the HEU.

Morbidity trends have been observed to be high in the first six weeks follow-up period. The reasons might be due to unsafe feeding practices when giving a child formula milk like proper handling of the milk, taking correct measures to prevent food-borne illnesses in their infants or mothers started to give complimentary food. High proportions in morbidity outcomes were in HUU babies. These infants might not have been exclusively breastfed hence disease risk results through exposure to foreign antigens in formula milk. The difference in morbidity outcomes between HEU and HUU children was found to be statistically significant in this study. This finding contradicts what other studies have found [7, 18]. The morbidity rate fell at four months and peaks up at nine months.

The other contributing factor to these discrepancies is that, HUU mothers might not have been taking much care of their children as compared to HEU mothers. HEU mother because of their HIV status, were most likely to get knowledge on how best they can take care of their babies. During the same period, exclusive breastfeeding practice was being encouraged especially to HIV positive mothers hence the HEU mother's could have got much knowledge as compared to HUU babies.

Due to the observed trend of the morbidity outcomes during the follow up period, a spline was done to fit a model two spline points. From the model results, breastfeeding and having a big family results in a protective effect in this study. Less morbidity outcomes were seen if a child is not breastfed. The HEU were breastfed less than the HUU hence the morbidity outcome differences between the two groups. HEU infants might have been exclusively breastfed or not breastfed at all since their mothers were HIV positive.

The model gave a similar structure between the two groups as the graph (Figure 2). During the first 16 weeks, there is a negative gradient slope of 0.02 and a positive slope of 0.014 after 16 weeks. This means that during the first sixteen weeks, less children reported with illness in both groups and gradually increase after sixteen weeks. This is supported by the effect change of exposure with time was negative, which decreases at 5.7% in the first sixteen weeks and increase by 3.6% after sixteen weeks. Being exposed had a protective effect of being ill or admitted since the log odds of morbidity outcome decrease by 0.26 for an HEU child as compared to HUU child.

For the child-specific regression coefficients of covariates, the child's log odds of morbidity increases by 0.04 for a day increase in the age of a child whilst holding other child's covariates fixed. Since the log odds are close to zero, the age factor had an odds ratio of 1.04. This means that the age factor has no effect on morbidity. The child's log odds of morbidity

decrease by 0.38 for a breastfed child as compared to a non-breastfed child whilst holding other child's covariates constant. Breast feeding had an effect in the HEU group who were breast fed less than HUU. Breastfeeding had an odds ratio of 0.98 which was less protective since it is close to one. Much protective effect was observed on the family size. The bigger the family the less likely one is to get sick. Family size had an odds ratio of 0.71. The log odds of morbidity outcome were 29% less likely in children from more than one child family as compared to children with only one child. None of these factors was statistically significant.

Incorporating some covariates in the base model resulted in a change in some of the estimates. The average child-specific regression coefficient increased from 0.55 to 1.04. Though the added covariates were statistically insignificant, they can be biologically plausible. A heterogeneity existed between the two groups hence the random effect part of the model was necessary.

Unbalance designs are most likely in longitudinal studies. These are due to missing data drop-outs, late entries and gaps in information. Drop outs occur when subjects leave permanently and do not return. In this study an overall loss to follow up of 53% was experienced within cohort. Above 50% of the study participants in both the HUU and HEU did not complete the study. The main type of missing data was mainly drop outs and less of information gaps. Late entries were not experienced in this study since no recruitment transpired after the study had began. The reason of loosing participants might be the death of the mother especially for the HEU children or relocation of participants during the study.

Though the study produced a different track of results from what has been proposed in other research, a number of limitations might have resulted from this project. Loss to follow up of participant and recall bias might have affected the results to deviate from what is mostly

being reported from some of the studies elsewhere. For this study, mother where recalling from past so this might have affected the mothers of HEU children since they might not have been there with the child or remember due to their compromised health.

In longitudinal studies, drop-out is one of the challenges that are faced by researcher hence resulting in missing data. Different mechanisms are associated with missing data including missing at random and missing completely at random. Regression based methods can be used in analysis to address outcome related dropout. Some of the ways which can be used are pattern mixture models and semi-parametric selection models.

Conclusion

Exposed to HIV had shown to be protective with an odds ratio of 0.77. There is a significance difference in the heterogeneity of the groups at baseline. The unexposed group have a significant negative trend during the first sixteen weeks and a positive trend after sixteen weeks. The exposed group has a less negative and positive trend across time. The family size has a protective effect towards morbidity in children.

Recommendations

Further research is required to determine the morbidity outcomes in HUU and HEU groups using clinical proxy variables like CD4 counts and cytokines levels of these infants to determine whether they are ill or not ill (morbidity outcome). Nutritional status can also be another variable used in infants illness diagnosis which showed whether the child in malnourished or not. Incorporation of these variables can give a more information on morbidity outcomes among HUU and HEU.

REFERENCES

1. UNAIDS. (2009).Report on Global AIDS Epidemic. Table of Country specific HIV/AIDS estimates and data end 2009.
2. Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International (2012). Zimbabwe Demographics Health Survey 2010-2011. Calverton, Maryland: ZIMSTAT and International Inc.
3. United Nations General Assembly Special Session (UGASS), (2010). Report on HIV and AIDS. Follow-up to the Declaration of commitment on HIV and AIDS. Zimbabwe Country Report. Reporting period: January 2008- December 2009.
4. Joint United Nations Programme on HIV/AIDS, *Together We Will End AIDS*, 2012.
5. Geddes, R., Knight, S., Reid, S., Giddy, J., Esterhuizen, T. & Roberts, C. (2008). Prevention of Mother-to-child Transmission of HIV Programme: low vertical transmission in KwaZulu-Natal, South Africa. *South African Medical Journal* **98(6)**:458–462.
6. Slogrove, A.L., Cotton, M.F. & Esser, I.M.M. (2010). Severe infections in HIV-exposed uninfected infants: Clinical Evidence of immunodeficiency. *Journal of Tropical Pediatrics* **56**:75-81.
7. Marinda, E., Humphrey, J.H., Ilif, P.J., Mutasa, K., Nathoo, K.J., Piwoz, E.J., Mouton, L.H., Salama, P. & Ward, B.J. Zvitambo Group. (2007). Child mortality according to maternal and infant HIV status in Zimbabwe. *Pediatric Infection Disease Journal*. **26(6):**519-526.
8. Ministry of Health and Child Welfare Zimbabwe (2010). Zimbabwe National AIDS Estimate. Zimbabwe.

9. Patel, D., Bland, R., Coovadia, H., Rollins, N., Coutsoudis, A. & Newell, M.L. (2009). Breastfeeding, HIV status and weight in South Africa children: a comparison of HIV-exposed and unexposed children. *Europe PupMed Central*. **24(3)**: 437-445.
10. Embree, J.E., Njenga, S., Datta, P., Nagelke, N.J., Ndinya- Achola, J.O., Mohammed, Z., Ramdalin, S., Bwayo, J.J. & Plummer, F.A. (2000): Risk factors for postnatal mother-child transmission of HIV-1. *PupMed AIDS.* **14**: 2535-2541.
11. Filteu, S. (2009). The HIV-1 exposed uninfected African child. *Tropical Medicine and International Health* **14**:276-287.
12. UNAIDS (2008). AIDS epidemic update: December 2008, UNAIDS and WHO. Geneva. Switzerland.
13. Proceedings of Peter Wall Institute for Advanced Studies Colloquium Abroad at the Stellenbosch Institute for Advanced Study (STIAS), Nov4-5. (2009).
14. Taha, E.T., Hoover, D.R., Chen, S., Kumwenda, N.I., Mipando, L., Nkanaunema, K., Thigpen, M.C., Taylor, A., Fowler, M.G. & Mofenson, L.M. (2011). Effects of cessation of Breastfeeding in HIV-1-Exposed, Ininfected Children In Malawi. *Clinical infectious Disease* **53(4):** 388=395.
15. Ministry of Health and Child Welfare AIDS and TB Unit, Zimbabwe. (2002). Report on the evaluation of the national prevention of mother to child transmission of HIV Pilot project of the Ministry of Health and Child Welfare Zimbabwe 2002.
16. Epalza, C., Goetghebuer, T., Hainaut, M., Prayez, F., Barlow, P., Dediste, A., Marchant, A. & levy, J. (2010). High incidence of Invasive group B Sreptococcal Infections in HIV-Exposed uninfected Infants. *Pediattrics* **126(3):** e631-e638.
17. Karpelosky, J.S., Millar, A.J.W., Graff, N., Bojerigen, G. & Heather, J. (2011). Outcomes of HEU Children undergoing a surgery. *BMC Paediatrics* **11**: 69.

18. Kurewa, E.N., Gumbo, F.Z., Munjoma, M.W., Mapingure, M.P., Chirenje, M.Z., Rusakaniko, S. & Stray-Perdersen, B. (2009). Effect of maternal HIV status of infant mortality: Evidence from a nine months follow-up of mothers and their infants in Zimbabwe. *Perinatology Journal* **30(2):** 88-92.
19. Kayanagi, A., Humphrey, J.H., Ntozini, R., Natoo, K., Mouton, L.H., Mutasa, K., Iliff, P., Ruff, A. & Ward, B. (2011). Morbidity among HIV exposed but uninfected, HIV infected and HIV uninfected infants in Zimbabwe before the availability of highly active ARV therapy. *Paediatric Infectious Disease* **30(1):** 45-51.
20. Fitzmaurine, G.M., Laird, N.M. & Ware, J.H. (2004). Applied Longitudinal Analysis (Wiley series in Probability and Statistics), Second edition, Hoboken NJ: *Wiley-Interscience* 189-209 & 345-350.
21. Diggle, P.J. (1991). Testing for random Dropouts in repeated measurements data. *Biometrics* **45(4):**1255-1258.
22. Carey, V., Zeger, S. & Diggle, P. Modeling multivariate binary data with Alternating Logistic Regressions. *Oxford Journals*. *Life Sciences and Mathematics and Physics. Biometrika.* **80(3):** 517-526.
23. Burton, P., Gurrin, L., Sly, P. & D'Agotino, R. (2005). Clustered Data: Extending the Simple Linear Regression Model to Account for Correlated Responses: An introduction to GEE and Multi-level Mixed Modeling. *Tutorial in Biostatistics Statistical Modeling of Complex Medical Data* **17(11):**1261-1291.
24. Ballinger, G.A. (2004). Using Generalised Estimating Equations for Longitudinal Analysis. *Organizing Research methods* **7(2):** 127-150.
25. Pregibson, D. (1979). Goodness of link test for Generalised Linear Models. *Applied Statistics* **29(1):**15-24.

26. Edwards, L.J. (2000). Modern Statistical Techniques for the Analysis of Longitudinal Data in Biomedical Research. *Pediatric Pulmonary* **30(4):** 330-344.
27. Rochon, J. (1996). Analyzing Bivariate Repeated Measures for discrete and continuous outcome variable. *Biometrics* **52**:740-750.
28. Nakai, M. & Ke, W. (2009). Statistical Models for Longitudinal Data Analysis. *Applied Mathematical Sciences* **3(40):** 1979-1989.
29. Pan, W. (2004). Model selection in Estimating Equations. *Journals of International Biometrics Society* **57(1):** 529-534.
30. Pan, W. (2001). Akaikes Information criterion in Generalized Estimation Equations. *Biometrics* **57(1):** 120-125.
31. Eley, B. (2006). Addressing the Paediatric HIV epidemic: a perspective from the Western Cape Region of South Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100 (1):**19–23.
32. Lallemant, M., Jourdain, G., Le Coeur, S., Mary, J.Y., Ngo-Giang-Huong, N., Koetsawang, S., Kanshana, S., McIntosh. K. & Thaineua, V. (2004). Single-dose perinatal nevirapine plus standard zidovudine to prevent mother-to-child transmission of HIV-1 in Thailand. *The New England Journal of Medicine* **351(3):**217–228.
33. Hedeker, D. & Gibbons, R.D. (2006). Longitudinal Data Analysis (Wiley-Series in Probability Statistics), *Wiley-Interscience*, New Jersey.
34. Good, M.F. & Doolan, D.L. (1999). Immune effector mechanisms in malaria. *Current Opinion in Immunology* **11**:412-419.
35. McCulloch, C.E. (2001). Generalised Linear and Mixed Models, Wiley New York. *Institute of mathematical Statistics and the American Statistical Association*

36. Agresti, A., Booth, J.G., Hobert, J.P. & Caffo, B. (2000). Random Effects Modeling of Categorical Response Data. *Sociological Methodology* **30**:27-28.

37. Froebel, K., Howard, W., Schafer, J.R., Howie, F., Whitworth, J., Kaleebu, P., Brown, A.L. & Riley, E. (2004). Activation by Malaria Antigens renders mononuclear cells susceptible to HIV infection and re-activates replication of endogenous HIV in cells from HIV-infected adults. *Parasite Immunology* **26(5)**:213- 217.

38. Walter, E.A., Greenberg, P.D., Gilbert, M.J., Finch, R.J., Watanabe, K.S., Thomas, E.D. & Ridell, S.R. (1995). Reconstitution of cellular immunity against cytomegalovirus in recipients of allergenic bone marrow by transfer of T cell clones from the donor. *The New England Journal of Medicine* **333:**1038-1044.

Annex I

Generalised Linear Mixed Effects Models

Generalized linear models (GLMs) represent a class of fixed effects regression models for several types of dependent variables (i.e., continuous, dichotomous, counts). Common Generalized linear models (GLMs) include linear regression, logistic regression, and Poisson regression. Generalised linear mixed models (GLMM) are an extension of generalised mixed models by the inclusion of random effects in the predictor. GLMMs have fixed effects and random effects hence they are called mixed effects models. The fixed effects are those factors whose levels are exponentially determined or whose interest lies in the specific effects of each level, such as effects of covariates, differences among treatments and interactions. The random effects are factors whose levels are sampled from a larger population, or whose interest lies in the variation among them rather than the specified effects of each other.

The key components of a generalized linear mixed model or a generalized linear fixed effects for longitudinal response Yi are:

- Linear predictor conditional on subject-specific effects **Ui:**

$$\cup \boldsymbol{ij} = \boldsymbol{Xij\beta} + \boldsymbol{dijUi}$$

- Link function: The conditional mean is connected to the conditional linear predictor via the link function h(·):

$h(\mu ij) = \cap\, ij$ where $\mu ij = E(Yij|Ui, Xi, Di)$

- Conditional distribution of Yij given μij (i.e., given (Ui,Xi)

There are three components of a generalized linear conditional (on **U**i) model for Yij. For a given subject **i**, all μij's and ηij's share random effect Ui. Ui accounts for the natural heterogeneity between subjects due to unmeasured factors (Ui is not observed) and Ui

accounts for the observed correlation (association) among the repeated measures Yij comprising $Y_i=(Y_{i1},...,Y_{ini})$ given Ui, the Yij's are assumed independent of one another and the Ui's are independent across subjects.

- In a generalized linear fixed effects model, the Ui's are treated as fixed quantities therefore can be used to control subject-level confounders
- The fourth component of a generalized linear mixed (or random effects) model for longitudinal data is a distribution for the random effects Ui:

Inferences of the GLMM random intercept

The regression coefficients of beta can be interpreted as the difference in log odds associated with a unit change in the corresponding covariate and the exponential regression coefficient as an odds ratio.

Random Effect Model

The model uses a random effect to model the relative similarities of observations made on the same statistical unit. Assumes the response variables are independent given some realised value of a random effect that appears in the conditional distribution of Yij given the random effect model.

Model Building

- Specify the fixed (treatment and covariates) and random effects(experimental, individual, temporal blocks)
- Choose an error distribution and link function (logit link for proportional data)
- Graphical Checking of the assumptions- are the variances of data (transformed by the link function) homogeneous across categories.
- Fix the fixed effect
- Model diagnostics checking model fitnes

DO-FILE

```
// Zvifadzo Matsena
***Dissertation Do-file
cd "C:\Users\Zvifadzo\Desktop\Merge\"
use final_, clear
log using "C:\Users\Zvifadzo\Desktop\Merge\zv.log"
****Descriptive
drop if child_exp>=.
foreach var of varlist  d_wt d_ht hd_cir apgar age{
tabstat `var', by(child_exp) c(s) s(n  p25 p50 p75 mean sd ) format(%9.1f)
ranksum `var',by (child_exp)
}
foreach var of varlist  Fever_ Cough_ Diarr_ Skin_ Convul_ Oral_{
tab `var'  child_exp, col
}
gen SEX=.
tab b_sex,nol
replace SEX=0 if b_sex==1
replace SEX=1 if b_sex==2
label define SEX 0"Male" 1"Female"
label values SEX SEX
foreach var of varlist breastfedsumm mumdecea family {
tab `var' child_exp, col
}
xttab breastfedsumm if child_exp==1
xttab breastfedsumm if child_exp==0
xttab family if child_exp==1
```

```
xttab family if child_exp==0
foreach var of varlist Over_morb Illnes_ admission {
tab `var' child_exp if week==6, col
}
foreach var of varlist Over_morb Illnes_ admission {
tab `var' child_exp if week==16, col
}
foreach var of varlist Over_morb Illnes_ admission {
tab `var' child_exp if week==36, col
}
var of varlist Over_morb Illnes_ admission {
tab `va
foreachr' child_exp, col
}
prtesti 869 536 731 389, count
prtesti 869 532 732 386, count
prtesti 873 24 739 31, count
 foreach var of varlist week {
 tab `var' child_exp, col
 }
 foreach var
*****Model Analysis
xtset Mster_idno week
xtdes
xttrans Over_morb
gen wk_exp= week* child_exp
****Fitting a fixed effect model
```

```
xtmelogit Over_morb child_exp week || Mster_idno:
estimates store zvi1
*****Random intercept
xtmelogit Over_morb child_exp week  wk_exp || Mster_idno:
estimates store zvi2
xtmelogit,or
*****Random slope
xtmelogit Over_morb child_exp week  wk_exp || Mster_idno :, cov(uns)
estimates store zvi3
xtmelogit,or
*****Random slopes with time squared
gen week2=week*2
xtmelogit Over_morb child_exp week wk_exp week2|| Mster_idno:,cov(un)
estimates store zvi3
*****Covarince of re
estat recovariance
*******Lkelihood test
lrtest zvi1 zvi2
lrtest zvi1 zvi3
****Model building/ splines
mkspline w_16 16 w_17=week,dis
gen w_16exposure=w_16* child_exp
gen w_17exposure=w_17* child_exp
xtmelogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure || Mster_idno:
estimates store zv1
****fixed effects model
xtlogit  Over_morb child_exp w_16 w_17 w_16exposure w_17exposure,fe
```

*****random effects

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure, re

estimates store ran1

xtlogit,or

*****random effects models with other covariates

****with age

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure age, re

estimates store ran2

xtlogit,or

***with sex

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure b_sex, re

estimates store ran3

xtlogit,or

***with mumdeceased

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure mumdecea, re

estimates store ran4

xtlogit,or

***breastfeeding

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure breastfedsumm, re

estimates store ran5

exlogit,or

*****with family

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure family, re

estimates store ran6

****combined covariates

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure family breastfedsumm age, re

estimates store ran7

****re with family and breastf

xtlogit Over_morb child_exp w_16 w_17 w_16exposure w_17exposure family breastfedsumm, re

*****Using xtmelogit

xtmelogit Over_morb child_exp w_16 w_17 w_16exp w_17expos breastfedsumm age SEX family mumdeceas || Mster_idno:,cov(un)

estimates store xt1

***drop mumdecease

xtmelogit Over_morb child_exp w_16 w_17 w_16exp w_17expos breastfedsumm age SEX family|| Mster_idno:,cov(un)

estimates store xt2

****Likelihood ratio test

lrtest xt1 xt2

****drop SEX

xtmelogit Over_morb child_exp w_16 w_17 w_16exp w_17expos breastfedsumm age family|| Mster_idno:,cov(un)

estimates store xt2

****likelihood ratio test

lrtes xt2 xt3

****drop age

xtmelogit Over_morb child_exp w_16 w_17 w_16exp w_17expos breastfedsumm family|| Mster_idno:,cov(un)

estimates store xt4

//a number of estimates changes by 10% hence I returned it in the model

*****Final model

xtmelogit Over_morb child_exp w_16 w_17 w_16exp w_17expos breastfedsumm age family|| Mster_idno:,cov(un)

estimates store xt1