



**Comparison of CD4+ T-Cell Changes In Response To Highly Active Antiviral
Therapy (HAART) In Adolescents And Children Enrolled At Parirenyatwa
Hospital Family Care Centre (2005-2010) - Secondary Data Analysis**

**Dissertation Submitted In Partial Fulfillment of The Requirements For The
Degree In Masters Of Science In Biostatistics**

By

Wellington Murenjekwa (R035006D)

University Of Zimbabwe College Of Health Sciences

Department Of Community Medicine

August 2013

Abstract

Background

In general, there is an increase in CD4 cell count after initiating on HAART. Despite an increasing trend of access to HAART there is a paucity of studies examining the changes in CD4 count over time in Zimbabwe and also no study has been done at PHFCC specifically comparing children and adolescents at the clinic. The PAP study data has not been analyzed to compare how CD4 count changes over time between different age groups and to find out the factors which predicts CD4 count response after initiating on HAART. The study aims to determine the factors associated with changes in CD4 count in adolescents and children and to compare the changes between these two age groups over time.

Methodology

Out of 2200 HIV infected children and adolescents who have been enrolled into HIV/AIDS care between January 2004 and December 2012, a total of 512 subjects who met the inclusion criteria were selected for this secondary data analysis study. Differences between groups in CD4 cell response at different time points was assessed using Wilcoxon rank-sum test. Mixed effects model was used to compare the pattern of changes in CD4 count over time between adolescents and children and to identify the factors which are associated with changes in CD4 count after HAART initiation.

Results

A total of 512 subjects were selected for the study. More (59.6 %) of the subjects were adolescents and the female gender (52.3%) was mostly represented. The change in CD4 count in response to HAART between adolescents and children was different. The median (IQR) baseline CD4 count for children was 171.5 (51-298) cells/mm³ and 145 (50-254) cells/mm³ for adolescents (p=0.087). The response in children was significantly higher after 18 months on treatment compared to adolescents (p=0.004). Baseline CD4 counts and age group was found to predict the changes in the square root of CD4 count over time in the multivariate analysis. The increase in the square root of CD4 count over time for those who initiate HAART at adolescence stage were 0.0853 times less when compared to those initiated whilst they were still children (p=0.037) adjusting for other variables. Adjusting for other baseline variables, subjects with CD4 cell count less than 100cells/mm³ had a greater increases (beta=0.501, p<0.001) in the square root of CD4 cell count when compared to those with baseline CD4 count of more than 300cells/mm³. Subjects with baseline CD4 count of 100 to 200 cells/mm³ had a greater increase in the square root of CD4 count over time as compared to those with baseline CD4 count above 300cells/mm³ adjusting for other variables (beta=0.340, p<0.001).

Conclusion

The change in CD4 count in response to HAART between adolescents and children was different. The baseline variables which were significantly associated with an increase in CD4 count over time were baseline CD4 cell count and age group after controlling for other independent variables.

. **Keywords:** CD4 cell count, Mixed Effect Regression, Antiretroviral Therapy,

Acknowledgements

First and foremost I would like to express my deepest gratitude to my supervisors Professor Rusakaniko and Mr Mandozana, without them this dissertation would never have come to pass. I would like also to thank Mr Tinago and Mr Chikwasha for their assistance and guidance.

My sincere acknowledgement goes to my family members for the support and encouragements. Special thanks to Dr Makadzange T.A, for providing the data set for the study. Finally I would like to thank my colleagues for their encouragement.

Dedication

I dedicate this project to my brother, Murenjekwa T L, for his continued support. May God richly bless you.

Table of Contents

Abstract	ii
Acknowledgements	iv
List of tables	viii
List of Abbreviations	ix
Glossary of Terms	1
CHAPTER ONE	2
INTRODUCTION	2
1.2 Description of the PAP study	4
1.2.1 The Objectives of the PAP study were:	5
1.2.2 Study Design and Setting	5
1.2.3 Study Population	5
1.2.4 Sample size	5
1.2.5 Inclusion Criteria	6
1.3 Critical appraisal of the PAP study	6
1.4 Statement of the problem	7
CHAPTER TWO	8
LITERATURE REVIEW	8
2.2 Related Studies	9
2.4 Justification	10
2.3 Research Questions	11
2.6 Broad Objective	11
2.7 Specific Objectives	11
2.5 Research Hypotheses	12
CHAPTER THREE	13

METHODOLOGY	13
3.2 Study population	13
3.3 Inclusion criteria.....	13
3.5 Study variables	15
3.6 Data Management	16
CHAPTER FOUR.....	19
RESULTS.....	19
4.1 Descriptive Statistics.....	19
4.2 Exploratory Data Analysis	22
4.3 Model Selection.....	26
4.3.1 Selecting Between the Random Slope Quadratic and Random Intercept	27
4.3.2 Univariate Analysis: Baseline factors associated with CD4 count changes.....	28
4.5 Model Diagnostics.....	33
CHAPTER FIVE	34
5.1 Discussion	34
Appendix 1: STATA DO FILES	40
Appendix 2: Longitudinal Data analysis.....	48
Appendix 3: Approval Letter 1	52
Appendix 4: Approval Letter 2	53

List of Figures

Figure 1: Spaghetti Plot for CD4 count over time	22
Figure 2: Mean Profile Plots	23
Figure 3: Lowess Smother Graph	24
Figure 4: Box plot of CD4 count at different time points	25
Figure 5: Normal Q-Q plot of square root of CD4 count.....	33

List of tables

Table 1: Baseline Characteristics.....	19
Table 2: Median (IQR) CD4 count of for children and adolescents at different time points	21
Table 3: Random intercept vs. Random slope model	26
Table 4: LRT for Random intercept quadratic vs. Random slope quadratic model	27
Table 5: Baseline factors associated with changes in square root of CD4 count over time	28
Table 6: estimated regression coefficient for CD4 count, their standard errors for a random slope model with time squared	30
Table 7: Random Effects Parameters.....	32

List of Abbreviations

ART – Antiretroviral Therapy

HAART- Highly Active Antiretroviral Therapy.

PAP- Pediatric and Adolescents’ Project.

MTCT- Mother to Child Transmission.

HIV-Human Immunodeficiency Virus.

AIDS-Acquired Immune Deficiency Syndrome.

LRT- Likelihood Ratio Test.

PHFCC- Parirenyatwa Hospital Family Care Centre.

JREC-Joint Parirenyatwa Hospital and College of Health Sciences Research Committee.

ML- Maximum Likelihood

RML- Restricted Maximum Likelihood

Glossary of Terms

For the purpose of the study an adolescent is anyone in the age range 10-19 years and a child is anyone in the age range 5-9years.

Immunological failure is failure to significantly raise CD4 count after a continued time of effective ART.

Virological failure is failure to decrease the viral load to undetectable levels following 24 weeks of effective antiretroviral therapy or a persistent increase in viral load following a period of total suppression.

CHAPTER ONE

INTRODUCTION

1.1 Background

The Human Immunodeficiency Virus (HIV) is a major health problem world-wide and the failure to implement prevention programs against mother to child transmission (MTCT) on an appropriate scale has resulted in an increased number of preventable HIV infections among newborns [1]. Initiating highly active antiretroviral therapy (HAART) early is known to reduce morbidity and mortality in children and adolescents infected with HIV [2]. HAART refers to the use of a combination of many antiretroviral medicines to slow the rate at which HIV replicates in the body [3]. Due to the advent of HAART, many HIV infected children can survive to adolescence and adulthood. Many studies have reported that HIV infected adolescents have poor treatment outcomes compared to other age groups [4].

The three main ways of assessing response to ART are clinical, immunological and virological assessment [5]. Immunological assessment involves routine monitoring of CD4 cell counts of an HIV infected individual and virological assessment involves routine repeated measurements of viral loads. Virologic assessment is the most sensitive of the three in managing HIV infected patients and in predicting the progression of the disease in those patients but the tests are not readily available in developing countries [5]. In resource limited countries, immunologic response is used in monitoring the response to HIV treatment and CD4+ T-cell count is used in deciding whether to initiate HAART [6].

The CD4 cell count of a person not infected with HIV can be between 500 and 1800cells/mm³ [6]. But CD4 cell counts can vary a lot between people. Individual CD4 cell count fluctuate and can increase or decrease responding to different factors. Variations in CD4 cell counts in the general population may be due to the movement of CD4 cells between blood and tissue so there is need to monitor any trends in changes to CD4 cell count over time [7]. Soon after infection with HIV, CD4 cell count may drop sharply, before stabilizing and without treatment, an HIV-positive person's CD4 count will fall over time [7].

The study is a secondary data analysis of data from the Pediatric and Adolescent Project (PAP) study at Parirenyatwa Hospital Family Care Centre (PHFCC). The PAP study is retrospective cohort study on patients enrolled at PHFCC from January 2005 to December 2012. Despite an increasing trend of access to HAART there is a paucity of studies examining the changes in CD4 count over time in Zimbabwe and also no study has been done at PHFCC specifically comparing children and adolescents at the clinic. As common with many observational studies with repeated measurements the dataset used in the PAP study have missing values and methods of analysis such as analysis of variance (ANOVA), multivariate analysis of variance (MANOVA) will not generate valid results because subjects with missing values will not be included in the analysis and also the correlations between measurements taken from the same subject will not be taken into account. To evaluate the changes in CD4 cell count over time, mixed effects regression model can be used. CD4 count measurements on the same subject are highly correlated and the PAP study data is unbalanced. A better approach of modeling correlated and unbalanced data is to fit mixed effect models which provide more convenient ways for modeling error structures among the repeated dependent variables [8]. This study aims to compare the changes in CD4

count over time between children and adolescents, and also to find factors associated with changes in CD4 count over time on patients receiving HAART treatment.

1.2 Description of the PAP study

The PAP study is a retrospective cohort study of outcomes among children and adolescents enrolled in HIV care between 2004-2011 at Parirenyatwa Hospital's HIV care and ART clinic in Harare by determining clinical, immunologic and virologic failure rates, adherence rates, LTFU and retention in care and mortality rates among children receiving HIV care at the hospital. Data is extracted from the clinic charts of all children who received HAART at the clinic since January 2005 to December 2012. The study also aims to prospectively evaluate clinical and laboratory parameters that determine outcomes, assess adherence rates among individuals who are currently retained in care, and identify factors affecting attrition from care.

At registration all patients who receive care at Parirenyatwa Hospital Family Care Centre are provided with unique patient number and a paper file for their medical record. All paper records were entered into an electronic database, which facilitated rapid data extraction. Some of the variables considered in the PAP study are socio demographic data for the child and the primary care giver, child's birth history including exposure to PMTCT and regimen, baseline parameters at enrollment (including age, clinical stage, weight, CD4 count, CD4%, hemoglobin level), treatment regimen(s), treatment interruptions, treatment switches, TB history together with clinical assessments such as height and weight.

1.2.1 The Objectives of the PAP study were:

1. To determine the treatment failure rates among children and adolescents with HIV infection.
2. To evaluate the accuracy of immunologic monitoring in predicting treatment failure
3. To develop comprehensive recommendations and guidelines for identifying and managing treatment failure in a resource limited setting that will be relevant to clinicians and program managers in the region.

1.2.2 Study Design and Setting

The study is retrospective cohort study on outcomes among children and adolescents enrolled in HIV care between 2005-2012 at Parirenyatwa Hospital's HIV care and ART clinic in Harare, Zimbabwe.

1.2.3 Study Population

All children ages 0-19 years enrolled in HIV/AIDS care at Parirenyatwa Hospital Family Care Centre.

1.2.4 Sample size

A total of 2200 HIV infected children and adolescents have been enrolled into HIV/AIDS care between January 2004 and December 2012.

1.2.5 Inclusion Criteria

All HIV infected children and adolescents ages 0-19 years that have been enrolled in care at the Parirenyatwa Hospital FCC, and have available records for chart review.

1.3 Critical appraisal of the PAP study

The PAP study is a single site retrospective study of all children ages 0-19 years enrolled at Parirenyatwa Family Care centre and the design is appropriate since the data was already available from the clinical charts. The study uses a cohort of subjects not under an artificial environment which means the results can be easily generalized to other similar settings. Also various effects of antiretroviral drugs treatment and the natural history of the disease after one is initiated on antiretroviral treatment can be studied. The data set can also be used to study many outcomes. The major drawback of this study design is that of loss to follow up and missing information on some variables. The study size is 2200 which is the population of all the children in that age group enrolled at the centre. The data is readily available so there are no extra costs associated with using a larger sample size.

The dataset for the study was extracted from the clinical charts and captured into an electronic Medical Smartcard Database. Some of the information might not be captured from the charts due to omission or misinterpretation of doctors' notes since those who were extracting are not practitioners. However, errors in extracting from the charts will be minimal because there is a process of verification of data extracted before it is entered into the database.

The major problem is missing laboratory results since the data was extracted from the charts entered as far back as 2005. Some of the laboratory results and doctors notes were not available.

Like any longitudinal data, the PAP study data have some missing information and this might have a negative impact in the results and inferences. Methods of analysis must be carefully selected so as to avoid biased results due to missing information. Information on adherence was not captured and baseline viral loads were not done and these might have an influence on treatment failure. Although, height and weight measurements are supposed to be done at registration, few charts captured this information. This missing information is essential in the calculation of the body mass index (BMI) and this variable is likely to predict treatment failure.

Mixed effects models will be selected over complete case analysis because the method uses all available data to identify the set of parameter values that have the highest probability of producing the sample data.

1.4 Statement of the problem

Many studies have reported that HIV infected adolescents have poor treatment outcomes compared to other age groups, mainly due to non-adherence [9]. Africa has an emerging epidemic of adolescent survivors of HIV infection acquired by MTCT and increasing numbers of long-term survivors are presenting for care for the first time during adolescence. A study in Zimbabwe reported that HIV infection is the commonest cause of acute adolescent admission to hospital in Harare [10]. Preliminary results on PAP cross sectional study, on factors associated with treatment failure at Parirenyatwa Family Care Clinic shows that adolescents constitute 65.6% of children with unsuppressed viral load. The PAP Study longitudinal data have not been analyzed to compare the changes in CD4 count over time between children and adolescents who are on HAART and to find the factors which are associated with CD4⁺ T cell changes using a statistical method appropriate for correlated data.

CHAPTER TWO

LITERATURE REVIEW

The global burden of pediatric HIV infection is high with over 2.5 million children infected with HIV. In sub-Saharan Africa 2.3 million children are infected with HIV, and 1.84 million are in need of ART. By the end of 2010 only one in five children in need of ART was receiving it [7]. In Zimbabwe it is estimated that over 120,000 children between the ages of 0-14 years are living with HIV [11]. In 2008 only 24,958 children were registered in ART clinics in Zimbabwe [12].

The vast majority of children in HIV care acquire HIV infection at birth through mother to child transmission. The Children with HIV Early Antiretroviral Therapy (CHER) trial showed a 76% reduction in mortality and 75% reduction in disease progression among children who were initiated on ART early with relatively high CD4 counts and percentages [13]. The data available suggests that infants and adolescents are a vulnerable population and interventions should be designed to reduce the risk of mortality and improve adherence [14]. The human immunodeficiency virus (HIV) attacks the CD4⁺ T cells thus weakening the immune system.

The CD4 count is a critical measure of immune system and is used as an important marker in describing the progression to AIDS. Generally, CD4 count in children decreases with increasing in age until it approximate the levels of adult CD4 count [8]. Some studies have demonstrated that subjects who are initiated on HAART at higher baseline CD4 count levels have better chances of immune recovery compared to subjects who are initiated at lower levels [16]. Monitoring the levels of CD4 count is the standard used in decision making concerning initiation of antiretroviral therapy and response to ART over time [15].

2.2 Related Studies

A longitudinal single centre observational study by [17], supports that HAART induces a beneficial effect in terms of clinical, virologic and immunological outcomes even in children previously exposed to antiretroviral therapy. They also found that CD4 T cells increased significantly in most children (65%) irrespective of the extent of virological suppression and baseline age of children, baseline CD4 count and the number of ARVs were found to predict immunological response.[17]. Another observational cohort study reported that adolescents (10-19 years) were less likely to experience immune recovery than young adults (20-30 years). The increase in CD4 count was higher in young adults when compared to the increase in adolescents although; their baseline CD4 count cells were almost identical [18]. In another study, the response to antiretroviral therapy in adolescents both for naïve and treatment experienced patients was poorer than in younger children, or in older adults [19].

A study in Paris by Scott A et al, 2001 found that an increase in CD4 T cell recovery was significantly greater in patients with lower baseline CD4 count as compared to those with high baseline [18]. Another study which used repeated measures models, [2] reported that large improvements in CD4% attributable to HAART initiation were observed among younger children (< 5years) compared with older children (≥ 5 years). The study also found that HAART initiation was associated with 2 fold greater odds of normal CD4% compared to non initiation of HART(OR=2.26; 95% CI, 1.65-3.1) [20].

Richard D M et al showed that median change from baseline CD4+ cell count to most recent CD4 cell count was +274cell/ml with 92% of patients having an increase in CD4+ cell count, significant increases were observed in all CD4+ cell count strata during the first year but there was a lower CD4 cell count at lower baseline CD4 cell strata [19]. Overall, many studies have shown that CD4 cell count increases rapidly in the first few years on HAART, then the rate of increase start to decrease with time[18]. Some approaches that can be used to analyze repeated measurement data are analysis of variance ANOVA, multivariate analysis of variance MANOVA and mixed model. Mixed model approach is the preferred method because of its major advantages over other approaches (see appendix). It allows covariance analysis in the model which cannot be done in ANOVA and MANOVA, it can handle unbalanced and incomplete data. The ANOVA approach does not take into account time dependent correlations in the repeated measurements data, resulting in incorrect standard errors when comparing means at different times [21].

2.4 Justification

Response to antiretroviral therapy in adolescents, both for naive and treatment-experienced patients is poorer than in younger children or in older adults. While HIV disease varies considerably between adults and children, it is not well understood in adolescents [22] and the advent of ARVs has increased the number of HIV infected children reaching adolescence stage in Zimbabwe. Little is known about the immunological response to HAART in adolescents enrolled at Parirenyatwa Family Care Centre, Zimbabwe but this information is needed to help health care providers in monitoring HIV infected children and adolescents on treatment. The PAP study data have not been analyzed to compare how CD4 count changes over time between

different age groups and to find out the factors which predicts CD4 count response after initiating on HAART. Few data are available on the immunologic response to ART in Zimbabwe especially in routinely collected data and this researcher has not come across with any study done at Parirenyatwa on immunologic response to HAART using a random effects model.

2.3 Research Questions

- Are the changes in CD4+ T-cell in response to HAART the same between individuals who start ART as children (age 5-9) compared with those who initiate in Adolescence (age 10-19 years)?
- What are the baseline factors that predict CD4+ T cell recovery following initiation of ART in children and adolescents?

2.6 Broad Objective

To compare the changes in CD4+ T-cell in response to HAART between individuals who start ART as children (age 5-9) and those who initiate in adolescence (age 10-19 years) and to identify baseline factors associated with CD4+ T cell recovery from January 2005 to December 2010.

2.7 Specific Objectives

- To compare the changes in CD4+ T-cell in response to HAART between individuals who start ART as children (age 5-9) and those who initiate in Adolescence (age 10-19 years) from January 2005 to December 2010.
- To identify baseline factors that predicts CD4+ T-cell recovery following initiation of HAART in adolescents and young children from January 2005 to December 2010.

2.5 Research Hypotheses

- Adolescents have a better CD4 T-cell response to HAART as compared to young children.
- Baseline factors such as primary care giver, WHO clinical stage, history of TB infection and chronic anemia are associated with CD4+ T cell recovery following ART.

CHAPTER THREE

METHODOLOGY

3.1 Description of Dataset

A total of 2200 HIV infected children and adolescents have been enrolled into HIV/AIDS care at Parirenyatwa Hospital Family Care Centre between January 2004 and December 2012. For the purpose of the secondary data analysis, data for subjects who are HIV positive, with a minimum of one CD4 count measurements and receiving highly active antiretroviral therapy at Parirenyatwa Family care clinic was extracted from the PAP study. The study subjects consists of children (aged 5 years and above) and adolescents. Follow up time to a maximum of the first 30 months from the initiation of antiretroviral therapy (ART) was considered. Baseline variables considered for the study are WHO clinical stage, hemoglobin level, history of tuberculosis (TB), primary care giver and gender. Changes in CD4+ T-Cell count was used as a measure of immunologic response.

3.2 Study population

Children aged 5 to 9 years and adolescents aged 10 to 19 years enrolled at Parirenyatwa Family Care Centre between January 2005 and December 2010.

3.3 Inclusion criteria

Patients enrolled in the PAP study, age at least 5 years and at most 19 years, on HAART and with at least one CD4 count were selected for the study.

3.4 Sample size

512 patients who met the inclusion criteria were selected for the purpose of this study

3.5 Study variables

Variable Type	Variable Name	Variable Description	Variable format	Variable code
Outcome	CD4 count	CD4 cell counts of patients		continuous
Study	Gender	Sex of study patient	string	Coded 0=female 1=male
	Hist_tb	History of tuberculosis	string	Coded 0=No 1=Yes
	WHOstage	WHO clinical stage	string	Coded 1=stage 1 2=stage 2 3=stage 3 4=stage 4
	hgb	Hemoglobin level	string	Coded 0=normal 1=anemic
	Pri_care	Primary care giver of the patient	string	Coded 0=mother 1=father 2=grandparent 3=other

3.6 Data Management

Data set was initially stored in Medical Smart Card Database which uses Microsoft access format. The data was exported to Excel and first saved as CSV (comma delimited) for easy exportation to STATA. The data was exported and cleaned in STATA 12. Duplicates were checked using patient identification number as the unique identifier. Variables important for the study were coded in preparation for analysis. Baseline CD4 count variable was categorized into four groups, less than 100cell/mm³, between 100 and 200 cell/mm³, between 200 and 300 cell/mm³, and above 300 cell/mm³. Hemoglobin level was categorized into two groups anemic (<11gm/dl) and normal (>=11gm/dl). The dataset which was originally stored in wide format was converted to long format using the reshape command for easy analysis of longitudinal data.

3.7 Statistical Analysis

Continuous variables like CD4 count at follow up times were described using the median and inter quartile range. Frequencies were used to describe categorical variables like WHO clinical stage, gender, TB history and primary caregiver. Age at baseline was categorized into two groups, adolescents (10-19 years old) and children (5-9 years old) and hemoglobin level was also categorized into two groups, anemic (<11gm/dl) and normal (>=11gm/dl). Median values of CD4 count counts at different time points were calculated and summarized. Wilcoxon rank-sum tests were used to compare the difference in medians of CD4 count at baseline, six months, 12 months, 18 months, 24 months and 30 months.

Exploratory data analysis was done to check patterns of systematic variation across groups of subjects and aspects of random variation that distinguish individual subject. Box and whisker plot by age group and normal Q-Q plot were used to check whether CD4 count measurements were normally distributed and after noting that the normality assumption could have been violated “gladder” command in STATA was used to select the best transformation and the square root transformation was found to produce the best results when compared to the logarithmic transformation.

Graphical methods were used to explore the magnitude of subject to subject variability in CD4 cell count over time. Mean profile plots with standard errors, by age group were produced to show how the CD4 T cell changes in response to HAART over time within and across subjects. A lowess smoother graph of CD4 count versus time by age group was also constructed to provide information on the spread of observation around the smoothed line and who are outliers in terms of their pattern of progression. Box plots were plotted to show the spread of the data in the two age groups.

From the information provided by in the exploratory analysis a linear mixed effect model was fitted. A random effects model fits an overall mean slope, but it assumes that each subject’s slope is effectively random sampled from a normal distribution of slopes. A random intercept model was first fitted and it was compared to the linear regression model using the likelihood ratio test to assess whether it fits the data well. A random slope model was also fitted and the likelihood ratio test was used to check the best model comparing it with the random intercept model and the model with a slope was found to fit the data well.

To select the covariance structure which provides the best fit for the data and thus reducing the risk of model misspecification, the four structures (identity, exchangeable, independent and unstructured) were fitted and the one with the smallest model log likelihood was selected. To determine the factors associated with changes in CD4 cell count univariate analysis for each baseline factor was assessed and those found to be significant ($p\text{-value}<0.25$) were selected for the multivariate analysis. A multivariate model was fitted to estimate the average increase in CD4 count over time for the two age groups. The coefficients of the final model and their respective p-values were used to assess the association between change in square root of CD4 count and independent variables and a p value of less than 0.05 was considered significant. All variables potentially associated with increase in CD4 count over time were included as interaction terms. A histogram of residuals of square root transformed CD4 count was plotted to check whether the assumption of normality was met.

3.5 Ethical Considerations

Letter of permission to use the data was sought from the authority in charge of Parirenyatwa Opportunistic Infection Clinic. Ethical clearance was sought from Joint Parirenyatwa Hospital and College of Health Sciences Research Committee (JREC/186/13). The study is a secondary data analysis and there is no harm to the study participants because there is no direct contact with the subjects and also patient identification numbers are used in the dataset instead of names.

CHAPTER FOUR

RESULTS

4.1 Descriptive Statistics

A total of 512 subjects met the inclusion criteria and were selected for the study and more of the subjects (59.6 %) were adolescents and female gender (52.3%) was most represented. Of the females, 169 (62.9%) were adolescents and the rest were children. Out of 244 males, 136 (56%) were adolescents. Children with history of TB constitute 7% of the cohort whilst adolescents with a history of TB constitute 10.4% of the cohort. Sixty seven (32.5%) of the children and 117 (38.5%) of adolescents had baseline CD4 count less than 100 cells/mm³. Fifty seven (19.2%) of the children and 105 (34.4%) of the adolescents had hemoglobin level greater than 11 gm/dl.

The median (IQR) age was 92.5(75-105) months for children, 155.5 (137.5-179) months for adolescents and baseline median hemoglobin were 10.4(9.4-11.3) gm/dl for children and 11(10.1-11.85) gm/dl for adolescents. The median (IQR) baseline CD4 count for children was 171.5 (51-298) cells/mm³ and 145 (50-254) cells/mm³ for adolescents.

Table 1: Baseline Characteristics By Age Category.

	Children		Adolescents	
	<u>Total (n)</u>	<u>Percent (%)</u>	<u>Total (n)</u>	<u>Percent (%)</u>
Sex				
Female	99	37.1	169	62.9
Male	108	44.0	136	56.0
WHO stage				
Stage 1	29	41.4	41	58.6
Stage 2	53	45.7	63	54.3
Stage 3	101	41.4	143	58.6
Stage 4	13	31	29	69.0
Missing	10	26.3	28	73.7
TB history				
No	165	40.3	243	59.7
Yes	36	40.4	53	59.6
Missing	5	35.7	9	64.3
Caregiver relationship				
Mother	78	53.8	67	43.2
Father	32	38.5	51	61.5
Grandparent	25	43.1	38	56.9
Other	71	31.7	143	68.3
Hemoglobin level				
Hgb $\geq$ 11 gm/dl	57	35.6	103	64.4
Hgb<11 gm/dl	100	51.8	93	48.2
Missing	49	31.2	108	68.8
Baseline CD4 count				
Less than 100 cells/mm ³	67	36.4	117	63.4
Between 100 & 200 cells/mm ³	53	39.6	81	60.4
Between 200 & 300 cells/mm ³	60	44.1	76	56.9
Above 300 cells/mm ³	26	46.4	30	54.6

Comparison of CD4 count for children and adolescents at different time points.

Table 2: Median (IQR) CD4 count of for children and adolescents at different time points

Time(months)	n	Median(IQR)	n	Median (IQR)	P-value
		Children		Adolescents	
Baseline	207	171.5 (51-298)	305	145.0 (50-254)	0.087
6	76	303.5 (157-551)	80	280.5 (155-455)	0.400
12	89	461.0 (285-702)	78	396.0 (275-640)	0.316
18	82	646.0 (397-890)	61	445.0 (321-66)	0.004
24	71	678.0 (456-1024)	60	461.0 (239-618)	<0.001
30	66	641.5 (364-894)	56	460.0 (241-827)	0.040

Table 2 above, gives the medians for the two age groups at different time points and it shows that there is an increase in CD4 count from baseline measurements (time =0) to 30 months for both age groups. The median CD4 count at baseline for adolescents was 145 (50-254) cells/mm³ and for children was 171.5 (51-298) cells/mm³ and the difference was not significant (p=0.087). Significant difference (p=0.004) in medians was observed as from 18 months after initiating HAART up to 30 months of follow up.

4.2 Exploratory Data Analysis

Figure 1: Spaghetti Plot for CD4 count over time

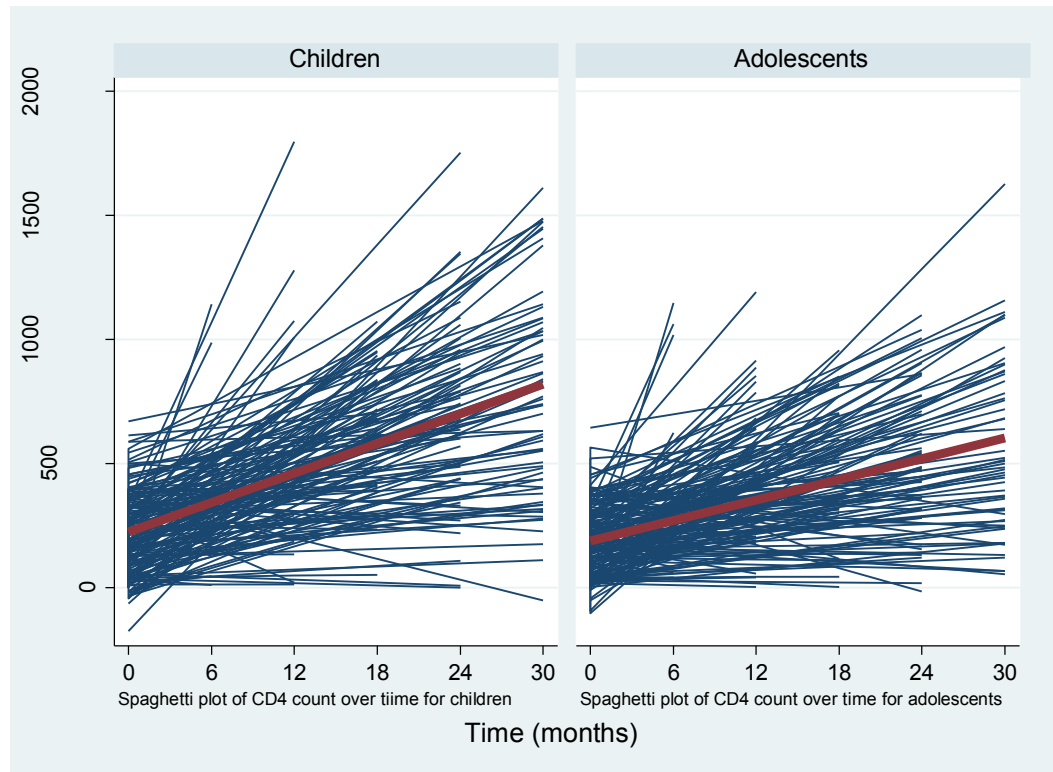
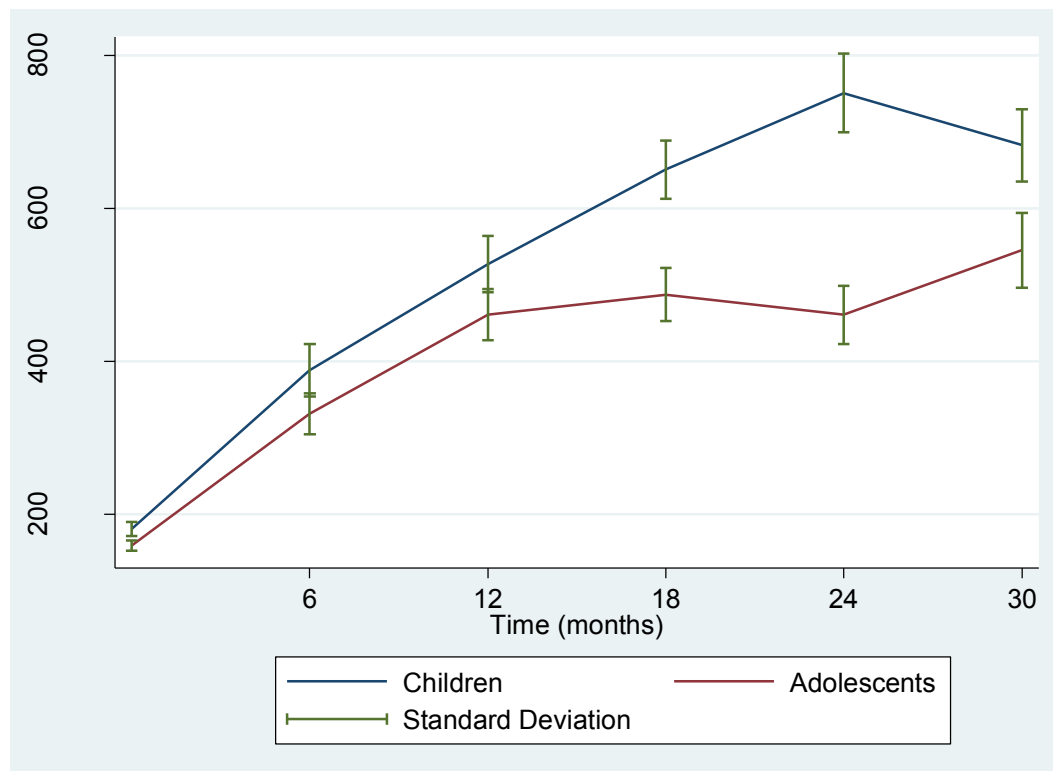


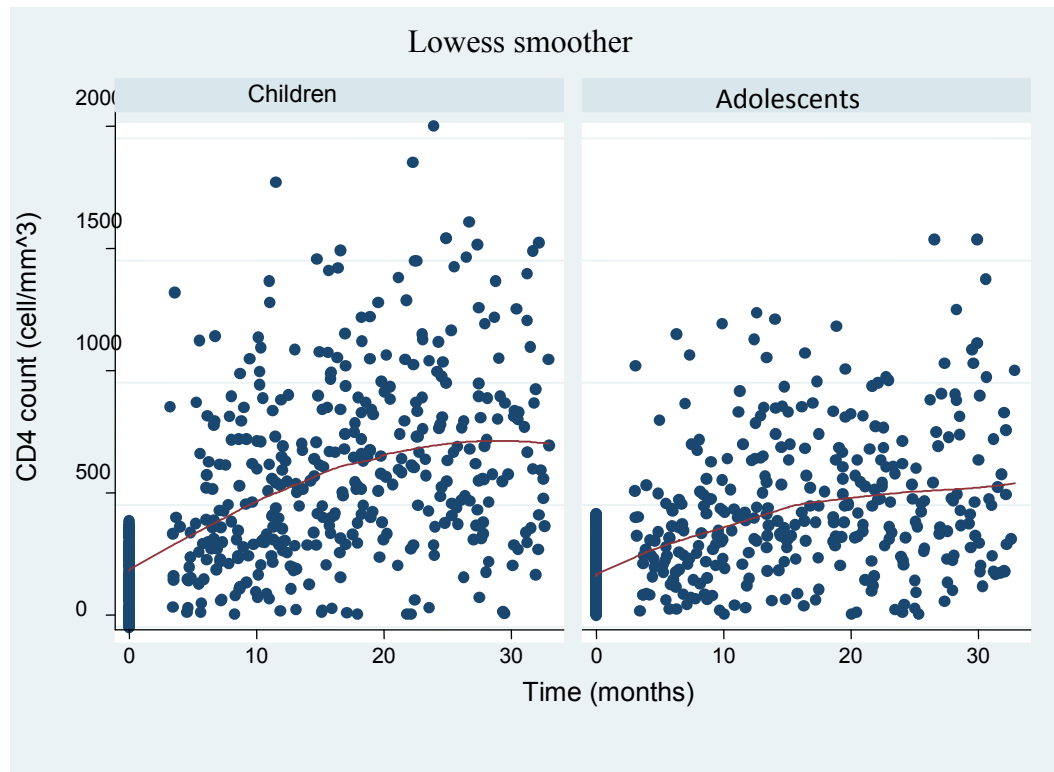
Figure 1 above shows that overall there is an increase in the level of CD4 counts over time in both children and adolescents groups and there are also outliers. The figure also shows that the numbers of CD4 count measurements decreases with time and after time 24 months the number of observations have significantly decreased in the two groups. The average CD4 count for children is generally high than that of adolescents at all time points. Variability increases from time 0 to 30 months of follow up.

Figure 2: Mean Profile Plots



The mean profile plot for the two age groups shows a sharp rise in the median CD4 count in the first six months and there is a steady increase up to thereafter up to 12 months. From Figure 2, above the mean changes of the two age groups over time are the same from baseline month up to one year. After 18 months the change in CD4 count was significantly different for the two age groups. The mean profile plot for the adolescents' increases steadily and slightly decreases after 18 months while that for children continue to increase up to 24 months. From 24 months to 30 months the profile plot for adolescents shows a steady increase in CD4 count while the graph for children shows a decline in CD4 count.

Figure 3: Lowess Smother Graph



In general, the two graphs show a decelerating positive trend. The lowess curve for the two age groups shows a gradual increase in CD4 count from baseline up to about 18 months. The graph for adolescents flattens after 20 months and that of children reaches its peak at around 25 months and then flattens. The change in CD4 count for adolescents is constantly less than the changes in children. There is much variability in CD4 count for children and outliers were also observed between the 20th and 30th month.

Figure 4: Box plot of CD4 count at different time points

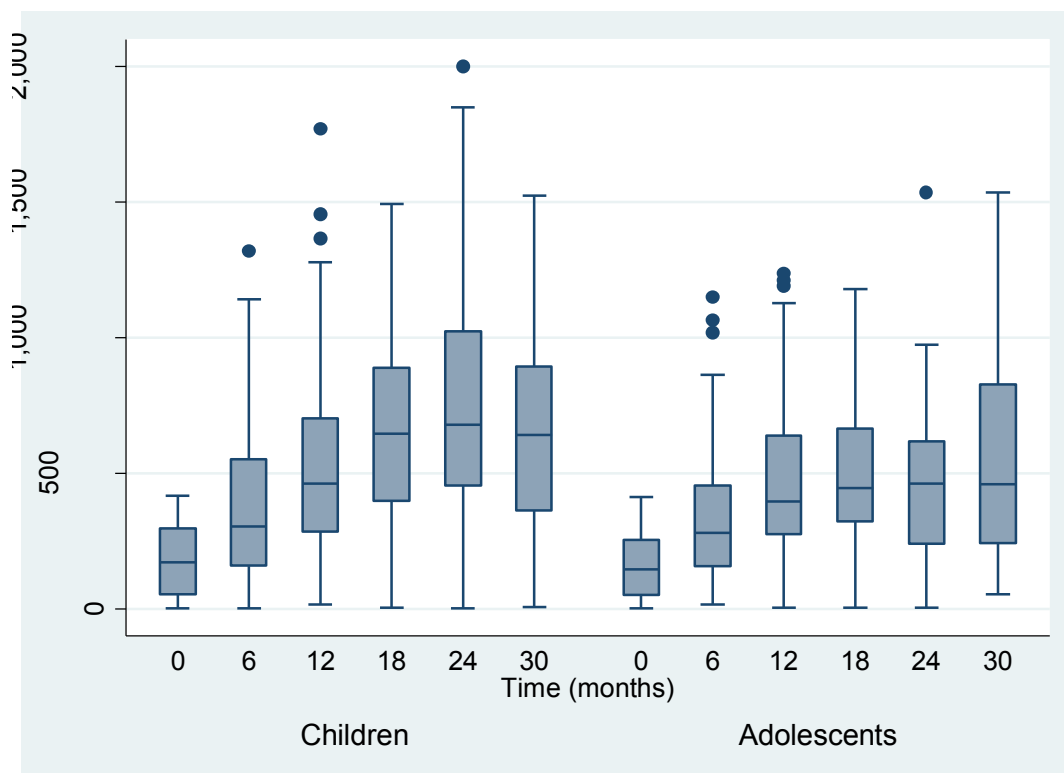


Figure 4 above shows box plots for CD4 count of adolescents and children over a period of 30 months. The median CD4 count for children is generally higher at different time points than that of adolescents and variability in CD4 counts for children is greater than that of adolescents. Box plots for the two groups' shows extreme values of CD4 count (outliers) at some time points.

4.3 Model Selection

In the selection of the best model, the likelihood ratio test was used. The unstructured covariance structure provides the best fit as it was found to have the smallest log likelihood after fitting the identity, exchangeable, independent and the unstructured covariance structures.

The positive decelerating trends on CD4 count changes from the mean profile plot and the loess graph suggest that a model with a quadratic time variable can be fitted. In fitting the model, centered time was used and this was generated by subtracting the mean time for all observation from the individual time points. Centering time was done to avoid co linearity of the time variable and the square of the time variable.

Table 3: Random intercept vs. Random slope model

Assumption	Random intercept model nested in Random slope model
Likelihood Ratio Test	233.7
p- value	<0.001

To select the best mixed effects regression model that explained the data well between random intercept model and random slope model, likelihood ratio test (LRT) was performed. The likelihood ratio tests was found to be significant ($p < 0.001$), which means the random slope fits the data well.

4.3.1 Selecting Between the Random Slope Quadratic and Random Intercept

Table 4: LRT for Random intercept quadratic vs. Random slope quadratic model

Assumption	Random intercept quadratic model nested in Random slope quadratic model
Likelihood Ratio Test	247.74
p- value	<0.001

In Table 4 above, we tested if there was need to include the random slope model with time squared and the likelihood ratio tests was found to be significant ($p < 0.001$) suggesting that the random slope model with a quadratic term is the best mixed model that explain the data well.

The results above suggest a model of the form

$$\text{Square root (CD4 count)} = \beta_0 + \beta_1 \text{time}_c + \beta_2 \text{age}_{grp} + \beta_3 \text{time}_c * \text{age}_{grp} + b_{0i} + b_{1i} + e_{ij}$$

β_0 is the average of CD4 cell count at 9 months

β_1 is the effect of time on CD4 cell count progression

β_2 is the difference in baseline measurement between the two age groups

β_3 is the effect of an age group on square root of CD4 progression

b_{0i} gives the difference between each individual subjects and the group intercept

b_{1i} is the individual slope deviation from the population slope for subject i

4.3.2 Univariate Analysis: Baseline factors associated with CD4 count changes

Table 5: Baseline factors associated with changes in square root of CD4 count over time

Independent variables		Time		Independent variables		Time Interaction	
		Coef (SE)	P-Value	Coef (SE)	P-Value	Coef (SE)	P-value
Gender	Male	0.497(0.0277)	<0.001	0.554 (0.5025)	0.270	-0.111(0.0406)	0.006
Hemoglobin level							
	Anemic	0.418(0.0360)	<0.001	-1.325(0.6044)	0.028	0.077(0.0484)	0.114
Age group							
	Adolescent	0.490(0.0289)	<0001	-1.172(0.5089)	0.021	-0.094(0.0407)	0.021
TB History	Yes	0.427(0.0230)	<0.001	-0.673 (0.6596)	0.307	0.097 (0.0554)	0.080
Primary caregiver							
	Mother		ref				
	Father			-1.021 (0.7770)	0.189	-0.003(0.0636)	0.962
	Grandparent			1.330(0.8738)	0.128	-0.084(0.0704)	0.323
	Other			-0.339(0.5955)	0.570	-0.014(0.0483)	0.767
WHO Clinical Stage							
	Stage 1	0.370(0.0546)	<0.001				
	Stage2			0.350 (0.8527)	0.681	0.014(0 .0679)	0.006
	Stage3			-1.756(0.7673)	0.022	0.094(0.0622)	0.133
	Stage4			-3.514(1.0971)	<0.001	0.238(0.0925)	0.01
Baseline CD4 count							
	> =300 cells/mm ³	0.135(0.0580)	0.020				
	200 –<300 cells/mm ³			-2.304(0.6135)	<0.001	0.183(0.0696)	0.009
	100 -<200 cells/mm ³			-6.185(0.6183)	<0.001	0.382 (0.0717)	0.001
	<100 cells/mm ³			-12.690(0.5900)	<0.001	0.511(0.06871)	0.001

Table 5 above, gives the results on the association between each baseline factor and changes in the square root of CD4 count over time in a random slope model. Factors that were significant at p-values less than 0.25 were included in the multivariate model. From the results above all variables except primary care giver were found to be significantly related to change in square root of CD4 count over time and will all be included in the multivariate model. Having a mother as a primary caregiver was associated with a greater increase in CD4 count over time but the association was not significant. History of TB, hemoglobin level less than 11gm/dl (being anemic) were associated with a greater increase in square root of CD4 count over time (beta=0.097, p=0.08 and beta=0.077, p=0.114) respectively. The increase in square root of CD4 count over time were 0.111 times less for males as compared to females (p=0.006). Patients with baseline CD4 count above 300 cells/mm³. WHO clinical stage 1 was associated with a lesser increase in the square root of CD4 count over time compared to other categories. The increase in square root of CD4 count were 0.094 (p=0.021) less in adolescents when compared to children.

4.4 Multivariate Analysis

Table 6: Estimated regression coefficient for square root of CD4 count, interaction with time and their standard errors for a random slope model

Independent variables	Independent variables		Time Interaction	
	Coef (SE)	P-Value	Coef (SE)	P-value
Time	0.222 (0.0972)	0.022		
Gender Male	-0.440 (0.4247)	0.300	-0.079 (0.0490)	0.109
Hemoglobin level				
Anemic	-0.248 (0.4300)	0.564	0.001 (0.0500)	0.990
Age group				
Adolescent	-0.715 (0.4269)	0.049	-0.0853 (0.0499)	0.037
TB History Yes	-1.072 (0.5571)	0.054	0.0672 (0.0648)	0.300
WHO Clinical Stage				
Stage 1	ref			
Stage2	1.680 (0.7050)	0.017	-0.014 (0.0804)	0.863
Stage3	1.3252 (0.6685)	0.047	-0.006 (0.0772)	0.941
Stage4	0.8158 (0.9402)	0.386	0.204 (0.1103)	0.065
Baseline CD4 count				
Above 300 cells/mm ³	ref			
200 to 300 cells/mm ³	-1.829(0.7185)	0.011	0.154(0.0791)	0.051
100 & 200 cells/mm ³	-6.237(0.7262)	<0.001	0.340(0.0810)	<0.001
<100 cells/mm ³	-12.761(0.7011)	<0.001	0.501(0.0796)	<0.001

NB: level of significant (alpha)= 0.05

Table 6 shows results of the multivariate analysis and hemoglobin level, history of TB and WHO clinical stage were not significantly associated with change in the square root of CD4 count over time. The change in the square root of CD4 count were 0.079 times less for males when compared to females, however, the association was not significant ($p=109$) . The increase in the square root of CD4 count over time for those who initiate HAART at adolescence stage were 0.0853 times less when compared to those initiated whilst they were still children ($p=0.037$) independent of other variables. The rate of increase in square root of CD4 count is higher for children as compared to adolescents.

Baseline CD4 count was significantly associated with an increase in the square root of CD4 count. The increase in square root of CD4 count for those with baseline CD4 counts of less than 100cells/mm³ was 0.501 ($p<0.001$) times higher when compared to those with baseline CD4 count of more than 300cells/mm³ controlling for other variables. Patients with baseline CD4 count of between 100 & 200 cells/mm³ had a greater increase in the square root of CD4 count over time as compared to those with baseline CD4 count above 300cells/mm³ adjusting for other variables ($\beta=0.340$, $p<0.001$). The baseline variables which were significantly associated with an increase in CD4 count over time were baseline CD4 count and age group after controlling for other independent variables.

Table 7: Random Effects Parameters

Random effects parameters	ML Estimate	95% Confidence Interval	
Var (time)	0.124	0.079	0.152
Var(intercept)	10.765	7.913	14.478
Cov (time, cons)	2.123	1.739	2.435

The variance of the intercept (10.765) is significant that is, the confidence interval does not include the null value and this means that there is significant variation of the intercepts between study subjects. The significant variances of time means that increase in square root of CD4 count over time vary across time.

4.5 Model Diagnostics

Figure 5: Normal Q-Q plot of square root of CD4 count

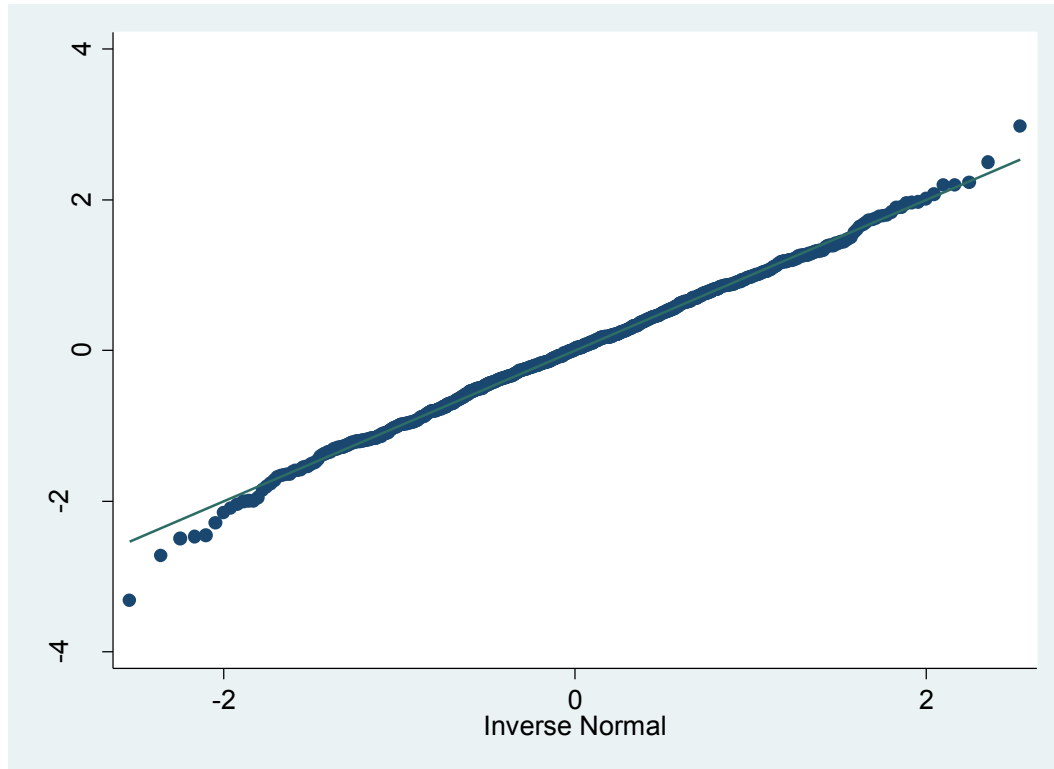


Figure 5 shows the Q-Q plot of square root of CD4 count. The residuals fall approximately on a diagonal straight line in this plot. Few outliers are observed at the extreme ends. This shows that the assumption of normality holds.

CHAPTER FIVE

5.1 Discussion

The secondary data analysis study use the statistical method of random effects modeling and this is the appropriate methodology for longitudinal studies with repeated measures. Sixty percent of all children and 64.9% of adolescents had baseline CD4 count below 200cells/mm³. The median baseline CD4 count for children was not significant different from that of adolescents (p=0.087). The median CD4 count at different time points shows an increasing trend in the level of CD4 count with time for both age groups. Generally, the increase in the level of CD4 count was consistently higher in children than in adolescents. Variables which were not significantly associated with changes in levels of CD4 count over time include level of hemoglobin, primary care giver, WHO stage, gender and history of TB. Age group and baseline CD4 count were significantly associated with an increase in the level of CD4 count over time.

The mean profile plot in Figure 2 shows that from 24 months onwards the level CD4 count for children starts to decline whilst that of adolescents' increases and this anomaly might be a result of fewer CD4 count measurements after 24 months. Children who were recovering faster might have missing CD4 count measurements whilst adolescents who were slowly recovering their CD4 count might have missing values also after the 24 months. The increasing trend in the median level of CD4 count in this study was also observed in a study by *Kiertiburanakul S et al* [27] in their prospective observational study of patients with HIV [27, 29]. The median CD4 count was found to increase over time from a low of 115 cells /mm³ to a peak of 302 cells/mm³ over a period of 36 months and in another study the increase was from 221 cells/mm³ to 314 cells/mm³ over a period of 6 years. Another study showed an increase in the median lymphocyte counts among patients taking ART in Eastern Ethiopia [28].

When HIV patients are commenced on HAART their level of CD4 count is generally expected to increase over time up to a certain point. In this study WHO clinical stage was not a significant predictor of CD4 count over time and this was supported by another retrospective longitudinal study on predictors of change in CD4 lymphocyte count among HIV infected patients. In that study they also found that level of hemoglobin was not an important predictor of change in CD4 count over time and this was consistent with my findings [28].

This study did not find an association between gender and a change in CD4 count over time. Other studies [30, 31], have also demonstrated a lack of association between gender and immunologic response. However, another study found a relationship between gender and immunologic response contradicting the findings above [32]. The increase in the level of CD4 count was found to be higher in those with lower baseline CD4 count and these results were similar to results from other studies [31, 33, 34]. Other studies have found a higher increase in CD4 count over time in those with higher baseline CD4 count [35, 36].

5.2 Conclusion

The results show that there is an increase in the levels of CD4 count cells over time in children and adolescents who are on HAART. The rate of increase in CD4 count over time was higher in children than in adolescents. The baseline variables which were significantly associated with an increase in CD4 count over time were baseline CD4 count and age group after controlling for other independent variables.

5.3 Limitations

There are limitations to the study. One of the factors which affect the responds to HAART is adherence to treatment and the information on this variable was not captured. If a patient is not drugs at the right time interval and in prescribed amounts this will results in an increase in the levels of HIV virus in the body which will in turn destroy the body cells with time. The virus also builds resistance resulting in reduced treatment option. Also viral loads at baseline were not done because of limited resources.

REFERENCES

1. UNAIDS: (2012) Global Report 2012: UNAIDS report on the global AIDS epidemic. Geneva: UNAIDS.
2. Kunjal p et al; Long Term Effects of HAART on CD4+ cell evolution among children and adolescents : Clinical Infectious Diseases 2008; 46:1751–60
3. <http://www.webmd.com/hiv-aids/tc/hiv-highly-active-antiretroviral-therapy-haart-topic-overview>
4. Agwu AL and Fairlie L. Antiretroviral treatment, management challenges and outcomes in perinatally HIV-infected adolescents. Journal of the International AIDS Society 2013, 16:18579.
5. Guidelines for Antiretroviral drug therapy in Kenya. 2005
6. Antiretroviral Treatment of Adult HIV Infection 2010 Recommendations of the International AIDS Society–USA Panel
JAMA. 2010; 304(3):321-333. doi:10.1001/jama.2010.1004.
7. AIDSMED. http://www.aidsmeds.com/articles/TCellTest_4727.shtml
8. Minggao S et al (1995). Analysis of Pediatric CD4 counts for AIDS patients using flexible curves.
9. Agwu AL and Fairlie L. Antiretroviral treatment, management challenges and outcomes in perinatally HIV-infected adolescents. Journal of the International AIDS Society 2013, 16:18579
10. Ferrand, R., et al. Survey of children accessing HIV services in a high prevalence setting: time for adolescents to count? Bull World Health Organ. 88(6): p. 428-34
11. Skovdal, M., et al. Challenges faced by elderly guardians in sustaining the adherence to antiretroviral therapy in HIV-infected children in Zimbabwe. AIDS Care. 23(8): p. 957-64.
12. WHO, *Epidemiologic Fact Sheet on HIV and AIDS, Zimbabwe, 2008*.
13. Marston M. et al. Estimating the net effect of HIV on child mortality in African populations affected by generalized HIV epidemics. J Acquir Immune Defic Syndr, 2005. 38(2): p. 219-27.
14. Faye A. et al. Early versus deferred antiretroviral multidrug therapy in infants infected with HIV type 1. Clin Infect Dis, 2004. 39(11): p. 1692-8.

15. Andrea S et al. Prediction Based classification for longitudinal biomarkers. Analysis of Applied Statistics; 2010 Vol 4. No 3. 1476-97
16. Jenniffer Hoffman . Role of the CD4 count in HIV management; CD4 count in resource limited settings. HIV Therapy 2010; 4(1): 27-39
17. Machado DM et al. Factors associated with clinical, Immunological and Virological Response in Protease-Inhibitor- Experienced Brazilian Children Receiving HAART. The Brazilian Journal of infectious Diseases 2007; 11(1): 16-19
18. Nachega JB et al. Antiretroviral Therapy adherence, virologic and immunologic outcomes in adolescents compared with adults in Southern Africa. *J Acquir Immune Defic Syndr*. 2009 May ; 51(1): 65–71.
19. Giaquinto C. Collaboration of Observational HIV Epidemiological Research Europe [COHERE] Study group. Response combination antiretroviral therapy: variation by age. European Collaborative Study ,AIDS 22(12), 1463-1473(2008)]
20. Kunjal p et al. Long Term Effects of HAART on CD4+ cell evolution among children and adolescents : Clinical Infectious Diseases 2008; 46:1751–60
21. Scott et al. CD4+ Cell count 6 years after commencement of HAART in person with sustained virological suppression: clinical infectious diseases, 2007
22. Puthanakit, T. et al. Hospitalization and mortality among HIV-infected children after receiving highly active antiretroviral therapy. Clin Infect Dis, 2007. 44(4): p. 599-604
23. Daniel P. O'Brien et al. Baseline Immunological Status among Young Children Receiving Non nucleoside Reverse Transcriptase Inhibitor–Based Antiretroviral Therapy in Resource-Limited Settings Clinical Infectious Diseases 2007; 44:1245–8
24. Nash D et al. Long-term immunologic response to antiretroviral therapy in low-income countries: a collaborative analysis of prospective studies. AIDS. 2008 November; 22(17): 2291–302.
25. Thai S et al. Predictors of Immune recovery and the association with mortality while on antiretroviral treatment in Cambodia.
26. Florence E, Laundgren J, et al. (2003). Factors associated with a reduced CD4 lymphocytes count response to HAART despite full viral suppression in the EuroSIDA study. HIV med 4: 255-262

27. Kiertiburanakul S et al. Trends of CD4 count levels at the initiation of antiretroviral therapy over time and factors associated with late initiation of antiretroviral therapy among Asian HIV-positive patients. *JIAS* 2014, 17:18804
28. Reda AA et al (2013) Predictors of change in CD4 Lymphocyte count and weight among HIV infected patients on ART in Ethiopia: A Retrospective Longitudinal Study. *PLoS ONE* 8(4): e58595. doi: 10.1371/journal.pone.0058595.
29. Tang H et al. (2014) Baseline CD4 counts of newly diagnosed HIV cases in China: 2006-2012. *PLoS ONE* 9(6): e96098. doi: 10.1371/journal.pone.0096098.
30. Moore A, et al. Gender differences in virologic response to treatment in an HIV-positive population: a cohort study. *J Acquir Immune Defic Syndr* 2001; 26:159-63
31. Smith C J et al. Factors influencing increases in CD4 cell counts of HIV-positive persons receiving long term HAART. *JID* 2004; 190: 1860-8
32. Pounstone K E et al. Differences in HIV disease progression by injection drug use and by sex in the era of highly active antiviral therapy. *AIDS* 2001;15:1115-23
33. Staazeski S et al. Determinants of sustainable CD4 lymphocyte count increases in response to ART. *AIDS* 1999; 13:951-6.
34. Cozzi-Lepri A et al. When to start highly active antiviral therapy in chronically HIV infected patients: evidence from ICONA study. *AIDS* 2001; 15:983-90.
35. Kauffman GR et al. The extend of HIV-1-related immunodeficiency and age predict the long term CD4 T cell response to potent antiviral therapy. *AIDS* 2002; 16:359-67.
36. Paddam LA et al. Effect of baseline CD4 count on immunological response to antiretroviral therapy in therapy naïve HIV 1 infected individuals. *International AIDS society*, 2002; 26: 159-63

Appendix 1: STATA DO FILES

```
// working directory

use "C:\Users\welly\Desktop\data1\desertation dset.dta", clear

log using "welly thesis `date'.log", replace

*****
keep if artatlastvisit=="Yes"

drop if age<60

//formatting Registration date

gen Regdate=date(regdate, "MDY" , 2012)

format Regdate %td

label variable Regdate "Registration Date"

//formatting date of birth

gen Dateofbirth=date(dob, "MDY", 2012)

format Dateofbirth %td

label variable Dateofbirth "Date of birth"

*****generating visiting times*****

// baseline visit

gen CD4date0=date(cd4date, "MDY" , 2013)

format CD4date %td

label variable visitdate0 "Date of baseline measurements"

g t0=(visitdate0-visitdate0)/30.4

//formatting CD4 date1

gen visitdate1=date(cd4date1, "MDY" , 2013)

format visitdate1 %td
```

```
label variable visitdate1 "Date of First Visit"
```

```
g t1=(visitdate1-visitdate0)/30.4
```

```
//formatting CD4 date2
```

```
gen visitdate2=date(cd4date2, "MDY" , 2013)
```

```
format visitdate2 %td
```

```
label variable visitdate2 "Date of 2nd Visit"
```

```
g t2=(visitdate2-visitdate0)/30.4
```

```
//formatting CD4 date3
```

```
gen visitdate3=date(cd4date3, "MDY" , 2013)
```

```
format visitdate3 %td
```

```
label variable visitdate3 "Date of 3rd Visit"
```

```
g t3=(visitdate3-visitdate0)/30.4
```

```
//formatting CD4 date4
```

```
gen visitdate4=date(cd4date4, "MDY" , 2013)
```

```
format visitdate4 %td
```

```
label variable visitdate4 "Date of 4th Visit"
```

```
g t4=(visitdate4-visitdate0)/30.4
```

```
//formatting CD4 date5
```

```
gen visitdate5=date(cd4date5, "MDY" , 2013)
```

```
format visitdate5 %td
```

```
label variable visitdate5 "Date of 5th Visit"
```

```
g t5=(visitdate5-visitdate0)/30.4
```

```
*****
```

```
//generating age category
```

```
gen age_grp=.
```

```
replace age_grp=0 if (age<120 & age>=60) & age!=.
```

```
replace age_grp=1 if (age>=120 & age<=228) & age!=.
```

```
label variable age_grp "Age Group category"
```

```
label define age_grp 0 "60-<120 months" 1 "120-228 months"
```

```
label values age_grp age_grp
```

```
//Recoding gender
```

```
gen Gender=.
```

```
replace Gender=1 if gender=="Male"
```

```
replace Gender=0 if gender=="Female"
```

```
label variable Gender "GENDER"
```

```
label define Gender 1 "Male" 0 "Female"
```

```
label values Gender Gender
```

```
//Generating new caregiver variable
```

```
gen pri_care=.
```

```
replace pri_care=0 if caregiver=="Mother"
```

```
replace pri_care=1 if caregiver=="Father"
```

```
replace pri_care=2 if caregiver=="Grandmother"
```

```
replace pri_care=3 if caregiver=="Other"
```

```
label variable pri_care "Primary Caregiver"
```

```
label define pri_care 0 "Mother" 1 "Father" 2 "Grandmother" 3 "Other"
```

```
label values pri_care pri_care
```

```
//Recoding history of TB
```

```
gen hist_tb=.
```

```
replace hist_tb=0 if tbhistory=="No"
```

```
replace hist_tb=1 if tbhistory=="Yes"
```

```
label variable hist_tb "History of TB"
```

```
label define hist_tb 0 "No" 1 "Yes"
```

```
label values hist_tb hist_tb
```

```
// generating WHO stage
```

```
gen WHOstage=.
```

```
replace WHOstage=1 if whostage=="1"
```

```
replace WHOstage=2 if whostage=="2"
```

```
replace WHOstage=3 if whostage=="3"
```

```
replace WHOstage=4 if whostage=="4"
```

```
label variable WHOstage "WHO clinical stage"
```

```
label define WHOstage 1 "WHO stage 1" 2 "WHO stage 2" 3 "WHO stage 3" 4 "WHO stage 4"
```

```
label variable WHOstage WHOstage
```

```
// hemoglobin level
```

```
g hgb=.
```

```
replace hgb=0 if hb0n>=11 & hb0n!=.
```

```
replace hgb=1 if hb0n<11 & hb0n!=.
```

```
label variable hgb "Hemoglobin level"
```

```
label define hgb 0 "normal" 1 "anemic"
```

```
label values hgb hgb
```

```
//Reshaping the data from wide to long format
```

```
reshape long cd4count cd4percent visitdate t, i( id ) j(visits)
```

```
sort id basedate time dic
```

```
quietly by id basedate time dic: gen dup = cond(_N==1,0,_n)
```

```
drop if dup==2
```

```
drop dup
```

```
sort id time
```

```
quietly by id time: gen dup = cond(_N==1,0,_n)
```

```
drop if dup==2 | dup==3
```

```
drop dup
```

```
tsset id time
```

```
tsfill
```

```
//declaring data to be panel Data
```

```
xtset id visitnr
```

```
lowess cd4count visits
```

```
xtline cd4count, overlay
```

```
estat recovariance, correlation
```

```
///Normality test
```

```

pnorm cd4count

predict resid, residuals

predict resid_std, rstandard

qnorm resid_std


//graphs

graph box root_cd4, over(age_grp)

xtgraph cd4count, av(mean) bar(se) t1("mean, se") group(age_grp)

ksm cd4count4 Time, lowess by(age_grp)

spagplot cd4count time, id(id)


//////// centering time////////

g ave_time=9.98738

g time_c=(Time-ave_time)

g time_sq=time_c*time_c

g t_agegp=Time*age_grp

g root_cd4=sqrt(cd4count)


///checking covariance structure to use: lowest likelihood

xtmixed cd4count time || id: time , mle cov(independent)

xtmixed cd4count time || id: time , mle cov(exchangeable)

xtmixed cd4count time || id: time , mle cov(identity)

xtmixed cd4count time || id: time , mle cov(unstructured)


/// selecting a suitable transformation

```

```
ladder cd4count
```

```
gladder cd4count
```

```
graph box root_cd4
```

```
////LIKELIHOOD RATIO TESTS//////////
```

```
///LRT 1
```

```
xtmixed root_cd4 age_grpn time_c t_agegp || id:, mle var
```

```
estimates store rin
```

```
xtmixed root_cd4 age_grpn time_c t_agegp || id:time_c, mle cov(un) var
```

```
estimates store rs
```

```
lrtest rin rs
```

```
///LRT2: random intercept quadratic vs. slope
```

```
xtmixed root_cd4 age_grpn time_c time_sq t_agegp || id:, mle var
```

```
estimates store rin1
```

```
xtmixed root_cd4 age_grpn time_c time_sq t_agegp || id:time_c, mle cov(un) var
```

```
estimates store rs1
```

```
lrtest rin1 rs1
```

```
///examine cov of random slopeseffects//
```

```
estat recovariance
```

```
*****testing equality of medians*****
```

```
ranksum cd4count0, by(age_grp)
```

```
ranksum cd4count1, by(age_grp)
```

```
ranksum cd4count2, by(age_grp)
```

```
ranksum cd4count3, by(age_grp)
```

```
ranksum cd4count4, by(age_grp)
```

```
ranksum cd4count5, by(age_grp)
```

```
**** median cd4 count at each time point by age group****
```

```
tabstat cd4count5, s(med p25 p75) by(age_grp)
```

```
tabstat cd4count4, s(med p25 p75) by(age_grp)
```

```
tabstat cd4count3, s(med p25 p75) by(age_grp)
```

```
tabstat cd4count2, s(med p25 p75) by(age_grp)
```

```
tabstat cd4count1, s(med p25 p75) by(age_grp)
```

```
tabstat cd4count0, s(med p25 p75) by(age_grp)
```

```
// univariate analysis
```

```
foreach var of varlist bCD4 age_grpn gendern pri_caren hist_tb WHOstage hgb {
```

```
xtmixed root_cd4 time_c time_sq `var' ||id:Time, mle
```

```
}
```

```
////final model
```

```
xtmixed root_cd4 i.age_grpn time_c time_sq t_agegp i.gendern i.WHOstage i.hgb|| id:time_c,  
mle cov(un) var
```

```
xtmixed root_cd4 i.age_grpn time_c time_sq t_agegp i.gendern i.hgb|| id:time_c, mle cov(un) var
```

```
log close
```

Appendix 2: Longitudinal Data analysis

Longitudinal data are repeated measurements taken on each of number subjects over time. The aims of longitudinal data analysis are to study and compare responses with time.

Some advantages of longitudinal data analysis

- Subjects can serve as their own control
- They allow researchers to distinguish between aging effects and cohort effects.

Measurements taken from the same subjects are usually correlated therefore the method of analysis must consider the existence of possible correlation among measurements from the same subject.

Mixed models are a design with both fixed effects and random effects. They are also known as multilevel models, linear mixed-effects models, random-effects models, random-coefficient models, hierarchical linear models. Fixed effects are parameters associated with an entire population or repeatable levels of experiments and random effects are associated with individual experimental units drawn at random from the population. Mixed models provide more convenient ways for modeling error structures among the repeated dependent variables. Mixed effects models can also handle unbalanced data that is, we do not necessarily require the same number of observations on each subject or that the measurements are taken at the same times. Mixed Effects Model can be used to model both linear and nonlinear relationships between dependent and independent variables.

Main stages in modeling are;

- Model the structure of means using fixed effects

- Specify a covariance structure both between subjects and within subjects
- Fit the means model accounting for the covariance structure specified
- Then make tests and inferences

The general linear mixed mode is in the form

$$\mathbf{Y} = \mathbf{XB} + \mathbf{ZU} + \mathbf{e} \quad \text{where}$$

$\mathbf{Y}$ is the $n \times 1$ response vector

$\mathbf{X}$ is the $n \times p$ fixed-effects design matrix

$\mathbf{B}$ is the $p \times 1$ vector of fixed effects

$\mathbf{Z}$ is the $n \times q$ random-effects design matrix

$\mathbf{u}$ is the $q \times 1$ vector of random effects

$\mathbf{e}$ is the $n \times 1$ vector of errors.

Random-Intercepts Model assumes that all variability in subject-specific slopes can be attributed to treatment differences. This has subject specific intercepts, but the same slopes within each treatment group. Random slopes involve freeing the variance of the slope parameter from being constrained to zero—i.e., allowing the slope parameter to have a variance. This means that we are allowing the slope of the regression line to take on a different value across the values of the level 2 variables.

The parameters in the covariance structure of the random error and the random effects are estimated using the maximum likelihood (ML) or the restricted maximum likelihood estimation (RML). Both methods are based upon the likelihood principle and the two only differs in the construction of the likelihood function. The methods use all available data to identify the set of parameter values that have the highest probability of producing the sample data. ML estimate is a value that is most likely to have resulted in the observed data. The estimation process uses a mathematical function called a log likelihood to quantify the standardized distance between the observed data points and the parameters of interest and the goal is to identify parameter estimates that minimize these distances. This option is preferable for a limited set of models for which the software is available. ML estimation provides estimates for fixed effects, whereas REML, by itself, cannot. RML variance components are less biased in small samples since they incorporate degrees of freedom used to estimated fixed effects.

LIKELIHOOD RATIO TEST

Is a generalization of the optimal test of simple null and alternative hypothesis that was developed by Neyman. It is based on the likelihood function $f(x_1, \dots, x_n / \theta)$, and the intuition that the likelihood function tends to be highest near the true value θ . The likelihood ratio test is the ratio of two likelihood functions, the numerator is the likelihood function maximized over parameter space restricted under the null hypothesis and the denominator is the likelihood function maximized over the unrestricted parameter space. The test can also be presented as a difference in the log likelihoods that is,

$$\begin{aligned}
\text{Likelihood Ratio Test} &= -2[\log_e (L_{\text{reduced}}) - \log_e (L_{\text{full}})] \\
&= -2\log_e (L_{\text{reduced}}) + 2\log_e (L_{\text{full}})]. \\
&= \text{deviance}_{\text{reduced}} - \text{deviance}_{\text{full}}.
\end{aligned}$$

This test follows a chi square distribution with degrees of freedom equal to the number of parameters between the two models.

Models with a lower deviance fit better than models with a higher deviance. If two models are nested, meaning that a specific model can be derived from a more general model by removing parameters from that general model, the deviances of the two models can be used to compare their fit statistically. The chi square test of deviance can also be used to explore the importance of a set of random effects by comparing a model that contains these effects against a model that excludes them.

Appendix 3: Approval Letter 1

DEPARTMENT OF MEDICINE

P. O. Box A 178
Avondale
HARARE, Zimbabwe

Telephone: 263-4-791631
Fax: 263-4-251017
Telegrams: UNIVERSITY

Chairperson: Prof M.Z Borok, MRCP (UK)
Email: medicine@medsch.uz.ac.zw



COLLEGE OF HEALTH SCIENCES

UNIVERSITY OF ZIMBABWE

July 14, 2013

To Whom It May Concern:

Re: Wellington Murenjekwa MSc Biostatistics masters research project

This is to confirm that Wellington Murenjekwa is an MSc Biostatistics student at the University of Zimbabwe who is working on his thesis project using Parirenyatwa Hospital Family Care Centre data. He will be working on his thesis project with my supervision. He will be conducting secondary analysis on data that has been obtained through the Pediatric and Adolescent Project (PAP) Study. The study has obtained all appropriate ethical approvals from JREC and MRCZ.

If you have any further questions please do not hesitate to contact me.

A. Tariro Makadzange, MD PhD
Lecturer University of Zimbabwe College of Health Sciences
Co-PI PAP Study

Appendix 4: Approval Letter 2



Joint Parirenyatwa Hospital And College of Health Sciences Research Ethics Committee

5th Floor College of Health Sciences Building
Telephone: +263 4 708140 Email: medirural@medsch.uz.ac.zw



University of Zimbabwe
College of Health Sciences

APPROVAL LETTER

Date: 2nd August 2013

JREC Ref: 186/13

Name of Researcher: Wellington Murenjekwa
Address: University of Zimbabwe, Department of Community Medicine

Re: Comparison Of CD4+T-Cell Changes In Response To Highly Active Antiretroviral Therapy In Adolescents And Children Enrolled At Parirenyatwa Hospital Family Care Centre (2005-2010).

Thank you for your application for ethical review of the above mentioned research to the Joint Research Ethics Committee. Please be advised that the Joint Research Ethics Committee has reviewed and approved your application to conduct the above named study.

- APPROVAL NUMBER: JREC/186/13
- APPROVAL DATE: 2nd August 2013
- EXPIRATION DATE: 1st August 2014
- TYPE OF MEETING: Expedited Review

This approval is based on the review and approval of the following documents that were submitted to the Joint Ethics Committee:

- a) Completed application form
- b) Full Study Protocol
- c) Informed Consent in English and/or appropriate local language
- d) Data collection tool version:

After this date the study may only continue upon renewal. For purposes of renewal please submit a completed renewal form (obtainable from the JREC office) and the following documents before the expiry date:

- a. A Progress report
- b. A Summary of adverse events.
- c. A DSMB report

- **MODIFICATIONS:**
Prior approval is required before implementing any changes in the protocol including changes in the informed consent.

- **TERMINATION OF STUDY:**

On termination of the study you are required to submit a completed request for termination form and a summary of the research findings/ results.

Yours Faithfully,



Professor MM Chidzonga
JREC Chairman