STATISTICAL ANALYSIS OF FACTORS ASSOCIATED WITH MORTALITY AMONG HIV INFECTED ADULT PATIENTS UNDER ANTIRETROVIRAL THERAPY (ART) IN HOSSANA DISTRICT QUEEN ELLENI MOHAMAD MEMORIAL HOSPITAL, SNNPR, ETHIOPIA

By: Gizachew Gobebo

A Thesis submitted to the Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University as a Partial Fulfillment of the Requirements of Masters of Science (MSc) in Biostatistics

October, 2015
Jimma, Ethiopia

STATISTICAL ANALYSIS OF FACTORS ASSOCIATED WITH MORTALITY AMONG HIV INFECTED ADULT PATIENTS UNDER ANTIRETROVIRAL THERAPY (ART) IN HOSSANA DISTRICT QUEEN ELLENI MOHAMAD MEMORIAL HOSPITAL, SNNPR, ETHIOPIA

MSc Thesis

By: Gizachew Gobebo

Advisor: Wondewosen Kassahun (PhD)
Co-advisor: Mersha Filate (MSc)

October, 2015
Jimma, Ethiopia

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES, DEPARTMENT OF STATISTICS

As thesis research advisors, we herby certify that we have read the thesis prepared by GIZACHEW GOBEBO MEKEBO under our guidance, which is entitled "**Statistical Analysis of Factors associated with Mortality among HIV Infected Adult Patients under Antiretroviral Therapy (ART) in Hossana District Queen Elleni Mohamad Memorial Hospital, SNNPR, Ethiopia**", in its final format it is consistent and acceptable; fulfills university and department style requirements; the final manuscript is satisfactory to the graduate committee.

Wondwosen Kassahun (PhD) _____ _____

Main Advisor Signature Date

Mersha Filate (MSc) _____ _____

Co-advisor Signature Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____ _____ _____

Name of Chairman Signature Date

_____ _____ _____

Name of Main Advisor Signature Date

_____ _____ _____

Name of Co-advisor Signature Date

_____ _____ _____

Name of Internal Examiner Signature Date

_____ _____ _____

Name of External Examiner Signature Date

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I somberly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Gizachew Gobebo

Date:_____

Signature:_____

Jimma University, Jimma

DEDICATION

This thesis is dedicated to my beloved father, Gobebo Mekebo who passed away when I was grade 10 student and my beloved mother, Almaz Ersulo who passed away when I was first degree graduate class student.

ACKNOWLEDGEMENT

Above all, I would like to thank my almighty God for giving me the opportunity to pursue my study at Department of Statistics, Jimma University.

Secondly, I would like to thank my main advisor Wondewosen Kassahun (PhD) for his important comments and encouragements throughout my study. I would also like to thank my co-advisor Mersha Filate (MSc) for his valuable suggestions and comments.

Thirdly, I would like to extend my gratitude to Mr. Dechasa Bedada for his friendly advice and provision of important materials.

Fourthly, my heartfelt thanks go to my family, especially to my dear sister, Misgana Gobebo, for their encouragement and support in my study period and thesis work.

I would also like to thank all my classmates for their encouragement and support that they gave me during my study period and this thesis work.

Finally, I would like to thank Wondimu and Tagesse; data clerks of Hossana District Queen Elleni Mohamad Memorial Hospital ART clinic for helping me in collecting the data for this study.

ABSTRACT

Even though the use of ART has brought a significant reduction in the mortality and morbidity of patients living with HIV/AIDS, a number of patients still die after the start of ART. The main objective of this study was to identify and describe factors associated with mortality among HIV infected patients who are taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital. The data for the study was obtained from Hossana District Queen Elleni Mohamad Memorial Hospital ART clinic. The HIV infected patients $\geq$ 15 years of age and who are under ART from March 2009 up to May 2015 were included in the study. So as to meet our objective logistic regression and multilevel logistic regression models were used. A total of 400 adult HIV infected patients who are taking ART were included in the study. Out of these patients, 18.75% of them were dead. The results obtained from standard logistic regression analysis showed that age, level of education, alcohol, baseline weight, woreda, TB status and baseline CD4 count were found to be significant factors of mortality of HIV infected patients taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital. The multilevel logistic regression analysis showed that the variance of the random component related to the intercept term was found to be statistically significant implying variation in mortality of HIV infected patients among the woredas where they live. It was also found that age, level of education, alcohol, baseline weight, TB status and baseline CD4 count were significant determinants of variations of mortality of HIV infected patients among woredas. However, the factors that had significant effect on mortality of the patients did not show underlying variations across woredas.

Key Words: Mortality of HIV infected patients, Multilevel Logistic Regression model, Random Intercept Logistic Regression Model.

TABLE OF CONTENTS

LIST OF TABLES

ACRONYMS

AIC	Akaike's information criterion
AIDS	Acquired immunodeficiency syndrome
ART	Anti retroviral treatment
ARV	Antiretroviral
AZT	Zidovudine
BIC	Bayesian Information Criterion
CI	Confidence interval
d4t	Stavudine
3TC	Lamivudine
EFV	Efavirenz
FHAPCO	Federal HIV/AIDS Prevention and Control Office
FMOH	Federal Ministry of Health
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
NVP	Neverapine
PLWHA	People Living with HIV/AIDS
SNNPR	Southern Nations Nationalities and People's Region
TB	Tuberculosis
TDF	Tenofovir Disoproxil Fumarate
UN	United Nations
UNAIDS	United Nations Program of HIV/AIDS
WHO	World health organization

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

AIDS-related illnesses remain one of the leading causes of death globally and are projected to continue as a significant global cause of premature mortality in the coming decades (UNAIDS, 2008).There are 34.2 million people living with HIV, 2.5 million new HIV infections and 1.7 million deaths due to AIDS in 2011 worldwide. Out of the total number of people living with HIV/AIDS in 2011, 30.7 million are adults, 16.7 million are women and about 3.4 million are children below the age of 15. From the total of 1.7 million deaths in 2011, about 230,000 are children below 15 years of age while the rest are adults (UNAIDS/WHO, 2012).

Sub-Saharan Africa has the most serious HIV and AIDS epidemic in the world. In 2010, about 68% of all people living with HIV resided in sub-Saharan Africa, a region with only 12% of the global population. Sub-Saharan Africa also accounted for 70% of new HIV infections in 2010, although there was a notable decline in the regional rate of new infections. The epidemic continues to be most severe in southern Africa, with South Africa having more people living with HIV (an estimated 5.6 million) than any other country in the world. Almost half of the deaths from AIDS-related illnesses in 2010 occurred in southern Africa. AIDS has claimed at least one million lives annually in sub-Saharan Africa since 1998. Since then, however, AIDS-related deaths have steadily decreased, as free antiretroviral therapy has become more widely available in the region. The total number of new HIV infections in sub-Saharan Africa has dropped by more than 26%, down to 1.9 million from the estimated 2.6 million at the height of the epidemic in 1997(UNAIDS, 2011).

Regional HIV and AIDS statistics shows that there are 23.5 million (22.2 million-24.7 million) adults and children living with HIV, 1.7 million adults and children newly infected with HIV, 1.2 million adult and child death due to AIDS and 4.8% adult prevalence of HIV in sub Saharan Africa in 2011 (UNAIDS, 2012).

In 2011, over half (56%) of HIV patients of sub-Saharan Africans in need of ART were receiving it (UNAIDS, 2012), in 2012 this increased to 68 % (UNAIDS, 2013). It is widely

acknowledged that increasing access to ART will dramatically decrease the impact of HIV in this region (Thirumurthy *et al.*, 2012).

In Ethiopia since the first two AIDS case reported in 1986, the prevalence rate has continuously increased until the year 2000 when it begun to show some decline (Merso, 2008). Adult HIV prevalence in 2009 was estimated to be between 1.4% and 2.8% in the country. Prevalence was 1.8% for males and 2.8% for females, and women accounted for 59% of the HIV-positive population. There were an estimated 131,145 new HIV infections and 44,751 AIDS-related deaths of which females accounted for 57% of the total infections and deaths. The total estimated number of HIV-positive pregnant women and annual HIV positive births in the same year were 84,189 and 14,140, respectively. There were an estimated 72,945 children less than 15 years old living with HIV, out of which 20,522 needed ART. Due to the combined effect of poverty and AIDS, more than 5.4 million children under the age of 18 years were orphaned out of which 855,720 lost at least one parent due to AIDS (FHAPCO, 2010).

The Antiretroviral (ARV) drugs improve the quality of HIV infected persons by helping them to stay well much longer than they otherwise would. The drugs slow down the replication of HIV within the body. Although the treatments are not a cure and continue to present new challenges with respect to side-effects and drug resistance ART as disease modifying therapy for established HIV infection has produced dramatic effects on morbidity and mortality among HIV AIDS patients. As a result of the widespread use of ART, the HIV AIDS pandemic which was once regarded as an infectious disease with an almost universal fatal outcome has been transformed into a manageable chronic infectious disease (WHO, 2003; Palella *et al.*, 2003).

A fundamental component of working towards the goal of providing, by 2010, universal access to antiretroviral treatment for patients with acquired immunodeficiency syndrome (AIDS) is an increased and secured production of antiretroviral drugs (ARTs) in order to meet the increased demand from lower- and middle-income countries. The vast majority of adults (96%) were reported to be receiving first-line regimens. Reporting compliance was very high for this group, with information on the specific regimens used available for 97% of this set of patients. The programs reported that 95% of all adults receiving first line regimens were using regimens consistent with the preferred first-line approach including: Stavudine (d4T) lamivudine (3TC) Nevirapine (NVP) (61%), Zidovudine (ZDV) +3TC+NVP (16%), ZDV+3TC+ efavirenz (EFV)

(9%), and d4T+3TC+EFV (8%). Less than 1% of these groups were reported to be taking either alternative first-line regimens, including the triple nucleoside combinations of ZDV+3TC+abacavir (ABC) and d4T+3TC+ABC, or taking regimens not considered or not recommended by WHO(Françoise *et al*., 2006).

There are two main approaches regarding when to start antiretroviral treatment as out lined by the guidelines. The more aggressive approach recommends starting when the CD4 counts fall below 500, and the second approach is more used in the United States which recommends starting antiretroviral medications immediately in all patients regardless of the patients CD4 counts (WHO protocols for CIS countries 2004).

The effectiveness of antiretroviral therapy is determined by its ability to rapidly reduce viral load and to sustain low levels of viral activity. This viral activity is what has an independent effect on increasing or decreasing susceptibility to opportunistic infections (Jamison *et al*., 2006).

Logistic regression is widely used to model the outcomes of a categorical dependent variable. The logistic regression model is supported by variety of link functions, which include the logit, clog-log, log and probit. For logistic regression, least squares estimation is not capable of producing minimum variance unbiased estimators for the actual parameters. Instead maximum likelihood estimation is used to solve for the parameters that best fit the data. Binary logistic regression is the form of regression which is used when the dependent variable is dichotomous and the independent variables are of any type (Hosmer and Lemeshow, 2000).

Multilevel modeling is applied to logistic regression and other generalized linear models in the same way as with linear regression: the coefficients are grouped into batches and a probability distribution is assigned to each batch. Or, equivalently, the error terms are added to the model corresponding to different sources of variation in the data (Gelman, 2006).

1.2. Statement of the problem

Even though many governmental and non-governmental organizations are doing their best to prevent people from HIV infection, people are still newly being infected by the virus. Since HIV/AIDS has no cure, these infected people have to be treated and given care to improve their quality of life. Most of the researches conducted previously in Ethiopia focused more on the prevention of people from infection by HIV. But it seems that little attention has been given to

study factors that facilitate mortality of those people living with HIV/AIDS. And those researches that have been conducted on factors that facilitate mortality of HIV infected patients taking ART were done using only standard single-level logistic regression model. This study in addition to standard single-level logistic regression model, it uses multilevel logistic regression model and considering woredas as clusters in which the patients are nested, tries to show variations in mortality of the HIV infected patients within and between woredas and identifies demographic, socioeconomic, behavioral and clinical factors of mortality of the HIV infected patients taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital.

This study will answer the following basic questions:

- ✓ Which model is appropriate for analyzing the predictors of the mortality of HIV infected patients taking ART?
- ✓ Which factors significantly affect the mortality of HIV infected patients taking ART?
- ✓ Is there variation in mortality of HIV infected patients taking ART among woredas?

1.3. Objective of the study

1.3.1. General objective

The general objective of this study is to identify and describe factors associated with mortality among HIV infected patients who are taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital.

1.3.2. Specific objectives

The specific objectives of this study were:

- ✓ To identify the effect of demographic, socio-economic, behavioral and clinical factors on mortality of HIV infected patients taking ART.
- ✓ To identify the factors that may explain the variation in mortality among HIV infected patients taking ART.
- ✓ To provide information for concerned bodies on the factors that facilitate mortality of HIV infected patients taking ART.

1.4. Significance of the study

The results of this study will have the following benefits:

- ✓ It helps to identify the potential risk factors of mortality among HIV infected patients taking ART.
- ✓ It gives information to concerned bodies in setting policies and strategies for improving quality of life of HIV infected patients.
- ✓ It can also be used as a basis for further studies on HIV infected patients.

CHAPTER TWO
2. LITERATURE REVIEW

2.1. HIV/AIDS in Ethiopia

Ethiopia is one of the countries in Sub-Saharan Africa highly affected by HIV/AIDS and its associated socioeconomic impacts. HIV in Ethiopia is predominantly spread through unprotected heterosexual intercourse, which accounts for approximately 88% of all HIV infections. In Ethiopia Since the year HIV was first claimed to be the cause of AIDS in 1984, AIDS has claimed the lives of millions and has left behind hundreds of thousands of orphans (FHAPCO, 2007).

As one of the countries in Africa, Ethiopia's GDP has increased in the last couple of years with double-digit economic growth. With the current effort of achieving the Millennium Development Goals, the country has made significant lead in several development sectors as in Health, Education and poverty reduction. Despite this dedicated effort to avert the burden of poverty, the hindrance caused by HIV/AIDS associated morbidity and mortality has posed an obstacle in the productive part of the society. The fact that the pandemic is predominately affecting part of individuals between the ages of 14-59, the productive age group, is a significant loss of labor supply. The protracted morbidity and eventual mortality resulting from HIV/AIDS causes significant lost time to illness, reduced productivity, shortage of manpower, increased absenteeism in the workplace and rising medical costs (AIDS Resource Center *et al.*, 2005).

HIV/AIDS has the greatest challenges to the Ethiopian health system, as elsewhere in sub-Saharan African countries. It has remained among the major causes of deaths over the past two decades. In 2010, more than one million people were estimated to be living with HIV in Ethiopia of whom nearly 397,818 need ART care and treatment (National Factsheet 2010).

In Ethiopia a total of 5.8 million people (53% male) received HIV counseling and testing in 2008/09. ART coverage increased from 46% in 2008 to 53% in 2009. As of end of 2009 there were a total of 241,236 people ever started ART. Females accounted for 57.9% of ART clients. A total of 11,000 children ever started ART as of December 2009. ART coverage for children was 43%. However, lost to follow up to ART service was 28% by the end of 2008 which is a major challenge (FHAPCO, 2010).

2.2. ART

The goal of treating HIV infected patients with antiretroviral drugs is to inhibit viral replication while minimizing toxicities and side effects associated with the available drugs. The inhibition of virus replication permits restoration of the immune system (suppression of HIV replication, as reflected in plasma HIV concentration, to as low as possible and for as long as possible, the preservation or enhancement of the immune function (CD4 restoration), thereby preventing or delaying the clinical progression of HIV disease. Viral eradication from the host genome is not achievable, thus a cure for HIV is not yet possible. By using HAART, it is possible to promote growth in children and prolong the survival of all HIV infected patients, reduce their morbidity and improve their quality of life (Elly *et al*., 2008).

A study done by Palella *et al*., (1998) in US evaluating 1255 HIV patients who were using antiretrovirals revealed that mortality among the patients declined from 29.4 per 100 person-years in 1995 to 8.8 per 100 person-years in the second quarter of 1997. The incidence of major opportunistic infections also declined from 21.9 per 100 person-years in 1994 to 3.7 per 100 person-years by mid-1997. There were also reductions in mortality and incidence of opportunistic infections regardless of sex, race, age, and risk factors for transmission of HIV.

Introduction of HAART has greatly improved the survival of HIV/AIDS patients. HAART reduces morbidity and mortality by suppression of viral replication, restoration and preservation of immune function, and prevention of drug resistance. Mortality among patients on HAART is associated with high baseline levels of HIV RNA, WHO stage III or IV at the beginning of treatment, low body mass index, severe anemia, low CD4 cell count, type of ART treatment, cotrimoxazole prophylaxis; gender, resource-poor settings and poor adherence to HAART (Elke *et al*., 2011).

A study was conducted in British Columbia, Canada by using Pearson's Chi-Square. The Cochran-Armitage and the Wilcoxon rank-sum tests have also been used to determine the degree to which antiretroviral resistance may contribute to mortality among HIV infected individuals enrolled in the centralized HIV/AIDS Drug Treatment Program in British Columbia, Canada, who had died between July 1997 and December 2001(Recsky *et al*., 2004).

The effectiveness of ART could vary from region to region (this variation is also generally a reflection of the variation that exists between and within countries and regions as regards HIV prevalence and its epidemiological patterns) because of the difference in background disease burden (such as tuberculosis or intestinal parasites), viral subtypes, and possible genetic differences in drug metabolism (Degu *et al*., 2006).

2.3. Predictors of mortality

A retrospective cohort study was done by Reda *et al*., (2013) among HIV infected patients on ART in Hiwot Fana, Jugal and Dil Chora hospitals located in eastern Ethiopia with objective of examining mortality and its predictors among a cohort of HIV infected patients on antiretroviral treatment retrospectively followed for five years. It was found that in the multivariate analysis factors such as WHO stage, weight, CD4 cell counts and level of education were predictors of mortality of HIV infected patients on antiretroviral treatment. The results revealed that WHO stage IV patients were 3 times more likely to die compared to stage I and II patients (HR 3.19; 95% CI 1.51–6.76). Patients who reported to have lost more than 10% of their weight at baseline were 5 times more likely to die compared to those patients who did not (HR 4.93; 95% CI 1.20–20.41). Patients whose CD4 cell counts between 201–300 were 60% less likely to die compared to those whose CD4 counts less than 200 (HR 0.40; 95% CI 0.17–0.93). Those patients with primary education were almost 3 times more likely to die than illiterate counterpart (HR 2.79; 95% CI 1.26–6.16).

A retrospective cohort study conducted in northern Thailand using Kaplan-Meier models to estimate mortality and Cox proportional hazards models to identify predictors of mortality revealed that sex, age group, registered year, clinical status, CD4 group, and ARV drug group were all significantly related to death in the univariable analysis, but the association with sex and age group was not significant in the multivariate analysis (Pathipvanich *et al*., 2003).

A study was conducted in Adama Hospital, Ethiopia with the objective of studying socio-economic, demographic and health factors that influence the survival/death status of HIV-Positives under ART follow up. It was found that in the multivariate analysis that factors such as Condom use, Alcohol, Drugs, Number of rooms, Base Line Weight, Base Line CD4 Count, Partners HIV status were predictors of mortality of HIV infected patients under antiretroviral treatment (Nuredin, 2007).

Lawn *et al* ., (2005) conducted a prospective cohort study in South Africa to determine the long-term incidence of TB and associated risk factors among individuals receiving HAART. The study revealed that risk of TB was independently associated with CD4 cell count < 100 cells/ml (P = 0.04), WHO stage III or IV disease (P = 0.01) and age < 33 years (P = 0.01). Risk of TB was not independently associated with plasma viral load, previous history of TB, low socioeconomic status or sex. Blood CD4 cell count increases were much smaller among patients who developed TB than among those who remained free of TB.

A retrospective cohort study conducted in Armed Forces General Teaching Hospital (AFGTH) located in Addis Ababa, Ethiopia and applied Kaplan-Meier survival curves and Log-Rank test to compare the survival experience of different category of ART patients, and employed proportional hazards Cox model to identify independent predictors of mortality. The independent predictors of mortality were low CD4 cell count at baseline, (HR = 0.995, 95% CI: 0.991 - 0.999), ambulatory and bedridden functional status, (HR=2.011, 95% CI: 1.018 - 3.973) and (HR=3.358, 95% CI: 1.734 - 6.500), respectively, WHO clinical stages III and IV (HR=7.052,95% CI: 1.677- 29.658) and (HR=12.64, 95% CI: 3.003 - 53.199), respectively, TB co-infection,(HR=1.734, 95% CI: 1.039 - 2.893) and OIs (HR=8.985, 95% CI: 1.240 - 65.085) (Ketema, 2011).

A longitudinal cohort study was done by Liu *et al* ., (2006) to identify the predictors for lower quality of life among HAART using HIV infected homosexual and bisexual men participants in four cities of USA. Random effects model was used to assess predictors. The study revealed that quality of life before HAART initiation was a strong predictor of quality of life subsequent to HAART initiation. Older age, lower socioeconomic status, less male sexual partners, no alcohol drinking, and more advanced HIV disease stage were significant predictors for lower physical health summary score. More outpatient visits, depression, amprenavir use, antiretroviral drug interruption, recreational drug use, and less social support were significantly associated with lower mental health summary score.

Health is believed to be influenced by ecological, social environment, genetics and individual characteristics. Urban and rural differences in mortality are determined by many factors, such as

differences in life style, ecological situation, and access to social and health services, unequal distribution of incomes and resources (Verheij, 1996).

The impact of HIV on an individual's probability of dying is complex and depends on many factors including sex, mode of infection, number of infections, age at infection(s), immune competence, overall health, and treatment(s), among many factors (Bajaria, 2002).

A study done by Johannessen *et al.*,(2008) in Tanzania among HIV-infected patients starting ART in one rural hospital with the objective of assessing mortality and to identify predictors of mortality. The study used Kaplan-Meier models to estimate mortality and Cox proportional hazards models to identify predictors of mortality and it was found that male sex, severe malnutrition and WHO stage IV were associated to progression to death and no such associations were found for age, religion, education level and active TB in the univariable analysis.

2.4. Multilevel Model

According to Goldstein (2003), Longford (1993) and Gibbons and Bock (1987) the multilevel model is extension of generalized linear models. In multilevel generalized linear models, the multilevel structure appears in the linear regression equation of the generalized linear model.

Multilevel models are increasingly being used to analyze hierarchical or clustered data in medical, public health, epidemiological, and educational researches. Many popular statistical packages such as MLwiN, R, SAS and Stata have the capacity to fit multilevel models. Multilevel models are also referred to as mixed effects models, random effects models, or hierarchical models (Austin *et al*., 2001; Snijders and Boskers, 1999).

Multilevel models recognize the existence of data hierarchies by allowing for residual components at each level in the hierarchy. For example, a two-level model which allows for grouping of child outcomes within schools would include residuals at the child and school level. Thus the residual variance is partitioned into a between-school component (the variance of the school-level residuals) and a within-school component (the variance of the child-level residuals). The school residuals represent unobserved school characteristics that affect child outcomes. It is these unobserved variables which lead to correlation between outcomes for children from the same school (Hox, 2010).

The multilevel regression model is more complicated than the standard single-level multiple regression models. One difference is the number of parameters, which is much larger in the multilevel model. This poses problems when models are fitted that have many parameters, and in model exploration. The second difference is that multilevel models often contain interaction effects in the form of cross-level interactions. Interaction effects are complicated, and analysts should deal with them carefully. The third difference is that the multilevel model contains several different residual variances, and no single number can be interpreted as the amount of explained variance (Hox, 2010). The multilevel model also provides a coherent model that simultaneously incorporates both individual- and group-level models as well as getting the right standard error (Gelman and Hill, 2006).

CHAPTER THREE
3. DATA AND METHODOLOGY

3.1. Study area and period

The study was conducted in Hossana Queen Elleni Mohamad Memorial Hospital in Hossana town, Ethiopia from March 2009 to May 2015. The hospital has a separate ART clinic. The clinic has one doctor, one nurse, one pharmaceutist and two data clerks. Hosanna town is administrative center for Hadiya zone and it is 235 km away from capital of Ethiopia, Addis Ababa. There are 10 woredas and one town administration in Hadiya zone. A total of 2899 HIV infected patients have visited ART clinic, 1644 ever started ART of which 152 have died while 342 are either transferred to other similar clinics or lost to follow up until May 2015 and 956 patients are currently on ART.

3.2. Study Population

The population of the study is HIV infected patients who are aged 15 years or older and taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital ART clinic.

3.3. Sample size and Sampling technique /Sampling procedures

The sample size for this study was 400. The sample was selected by using stratified random sampling technique by grouping patients by woreda where they live. The sample size in each woreda was determined in proportion to the size of the woreda (the number of patients in the woreda), termed as proportional allocation (Hossana town administration 71, Lemmo 51, Anlemmo 45, Misha 53, Angecha 37, Duna 33, Lera 45, Gombora 36, Soro 29)(Appendix A: Table A). Finally, using ART unique identification number of eligible patients, a simple random sample of patients was taken from each woreda.

3.4. Eligibility Criteria

Inclusion criteria:

- ✓ HIV infected patients aged 15 years or older who have started ART
- ✓ HIV infected patients with complete intake form, registers and follow up form

Exclusion criteria:

- ✓ Diagnosis made outside of the hospital
- ✓ Loss to follow up (withdraw, transfer out)

3.5. Data collection procedures (Instrument, personnel, data quality control)

The data were extracted from the available standard national medical registers which have been adopted by Federal Ministry of Health (FMOH) to be uniformly used by clinicians to simply identify and document clinical and laboratory variables. The registers include pre-ART register and follow up form, ART intake form, patients' card and death certificate complemented by registration by home visitors. Two days training was given for supervisors and data collectors. The overall activity was controlled by the researcher. Data quality was controlled by designing the proper data collection materials and through continuous supervision. The completed data collection forms were examined for completeness and consistency during data management, storage and analysis. The data were collected by data clerks working in the clinic and coded and analyzed using the statistical packages STATA and SAS.

3.6. Variables of the study

The variables of interest that were considered in the analysis are the response (dependent) variable and the explanatory (independent) variables.

3.6.1. Response (Dependent) variable

Response variable is the HIV infected patient mortality. It is dichotomous variable coded as 1 if a patient is dead and 0 if a patient is alive.

3.6.2. Explanatory (Independent) variables

Several explanatory variables are used as predictors of mortality of HIV infected patients taking ART. These variables are classified as Demographic factors (Age, Sex, Marital status), Socio-economic factors (Residence, Level of educational and Woreda), Behavioral factors (Tobacco and Alcohol) and Clinical factors (WHO clinical stage, Baseline CD4 counts, Baseline weight, Antiretroviral regimen, and TB status).

Table 3.1: Independent variables description and categories

Variable name	**Description**	**Categories**
Sex	Sex	0=Female 1= Male
Age	Age in complete years	0 =Below 40

		1 = 40 or above
Wrda	Woreda where patient lives	0= Hossana town administration 1=Lemmo 2=Anlemmo 3=Misha 4=Angecha 5= Duna 6= Lera 7=Gombora 8=Soro
Resd	Residence	0= Rural 1= Urban
Marst	Marital status	0= Never married 1= Married 2= Separated 3=Divorced 4= Widowed
Led	Level of Education	0= No education 1= Primary 2= Secondary or above
Tobc	Tobacco	0=No 1=Yes
Alch	Alcohol	0=No 1=Yes
Baslinwt	Base Line Weight	0=Less than 50kg 1=50kg or above
WHOstg	WHO Clinical Stage	0= Stage I 1= Stage II 2=Stage III 3= Stage IV

TBStas	TB status	0= Negative 1= Positive
Regmn	Antiretroviral regimen	0=d4t-3TC-NVP or d4t-3TC-EFV 1=AZT-3TC- NVP or AZT-3TC-EFV 2=TDF-3TC-EFV or TDF-3TC-NVP
BaslinCD4	Baseline CD4 Count	0=Less than 200 1= 200 or above

3.7. Methods of Data Analysis

3.7.1. Logistic Regression Model

Logistic regression is part of a family of models called the Generalized Linear Model used when the response variable is qualitative in nature or categorical and independent variables may be either continuous or categorical. Unlike discriminant analysis, the logistic regression does not have the requirements of the independent variables to be normally distributed, linearly related, nor equal variance within each group (Fidell and Tabachnick, 2007). There is no formal requirement for multivariate normality, homoscedasticity, or linearity of the independent variables within each category of the response variable. However, the assumptions that apply to logistic regression model include: meaningful coding, linearity through logit transformation of the dependent variable and sampling adequacy.

Let Y_{nx1} be a dichotomous outcome random vector with categories 1 (HIV infected patient taking ART is dead) and 0 (HIV infected patient taking ART is alive). Let X be an n x (k+1) matrix denote the collection of k-predictor variables of Y, i.e.

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \sim \ (k+1) \times 1$$

Where, X -is the design matrix

β - is the vector of unknown coefficients of the covariates and intercept

Let π denote the proportion of success (Death of HIV infected patient taking ART).

Then, the conditional probability that the ith patient mortality given patient characteristics $\mathbf{X}_i$ is given by:

$$\pi_i = p(y_i = 1/x_i) \quad \text{.....(3.1)}$$

In logistic regression analysis, it is assumed that the explanatory variables affect the response through a suitable transformation of the probability of the success. This transformation is a suitable link function of π_i, and is called the logit-link, which is defined as:

$$\log it(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \quad \text{.....(3.2)}$$

The transformed variable $\log it(\pi_i)$ is related to the explanatory variables as:

$$\log it(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} = X_i'\beta \quad \text{.....(3.3)}$$

Where,

$\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)'$ are parameters

$X_i = (1, X_{1i}, X_{2i}, \ldots, X_{ki})'$ i = 1, 2, 3,..., n

The probability of success is expressed as follows:

$$\pi_i = p(Y_i = 1/x_{1i}, x_{2i}, \ldots, x_{ki}) = \frac{e^{X_i'\beta}}{1+e^{X_i'\beta}} \quad \text{.....(3.4)}$$

Then, odds of success is given as follows

$$odds(Y_i = 1) = \frac{\pi_i}{1-\pi_i} = e^{X_i'\beta} \quad \text{.....(3.5)}$$

3.7.1.1. Estimation of Logistic Regression Model

The most commonly used method of estimating logistic regression parameters is method of Maximum Likelihood. The method of maximum likelihood yields values for the unknown parameters which maximize the probability of obtaining the observed set of data. In order to apply this method we first construct a function, called the likelihood function. This function expresses the probability of the observed data as a function of the unknown parameters. The maximum likelihood estimators of these parameters are chosen to be those values that maximize this function. Thus, the resulting estimators are those which agree most closely with the observed data. We describe how to find these values from the logistic regression model. Since each y_i represents a bernoulli count in the i^{th} population, the probability distribution function of Y_i is given by:

$$f(y_i) = \pi^{y_i}(1-\pi)^{1-y_i} \quad \text{.....(3.6)}$$

$Y_i = 0$ or 1 and i =1, 2, 3, …, n

Then the likelihood function is the joint probability distribution of all *n* observations:

$$l(\beta) = \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i} \quad \text{.....(3.7)}$$

The principle of maximum likelihood states that we use as our estimate of parameter the value which maximizes the expression in equation (3.7).However, it is easier mathematically to work with the log of equation (3.7). This expression, the log likelihood, is defined as:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln(\pi) + (1-y_i)\ln(1-\pi)\} \quad \text{.....(3.8)}$$

The maximum likelihood estimates are the values for β that maximize the likelihood function in equation (3.8). Through maximization of the log-likelihood function we can theoretically estimate the parameter vector β. But the equation is nonlinear in β, and as a result the estimates do not have a closed form expression. Therefore, β can be obtained by maximizing using iterative algorithm method (Agresti, 1996).

3.7.1.2. Goodness of Fit of Logistic Regression model

Once the model is fitted, we would be interested to know how effective the model is in describing the outcome variable. This is referred to as goodness-of-fit.

3.7.1.2.1. Wald Test

A Wald test is used to test the statistical significance of each coefficient (β) in the model. If the Wald test is significant for a particular explanatory variable, then we would conclude that the parameter associated with this variable is not zero so that the variable should be included in the model otherwise it should be omitted from the model (Agresti, 1996).

The hypothesis to be tested is:

$H_0 : \beta_j = 0$ Versus $H_A : B_j \neq 0$

The Wald test statistic, Z, for this hypothesis is

$$Z^2 = \frac{\hat{\beta}_j{}^2}{\text{var}(\hat{\beta}_j)} \sim \chi^2_{(1)} \quad \text{……..……..(3.9)}$$

Where, $\hat{\beta}_j$ is the estimated regression coefficient and $\text{var}(\hat{\beta}_j)$ is the variance of $\hat{\beta}_j$.

3.7.1.2.2. The likelihood ratio test

The likelihood ratio test statistic (G^2) is the test statistic commonly used for assessing the overall fit of the logistic regression model. The likelihood ratio test is computed based on - $2LL$ (-2 times log likelihood). The likelihood ratio statistic is obtained by subtracting the two times log likelihood ($-2LL$) for the full model from the log likelihood for the intercept only model. This log likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the intercept only model L_0 over the maximized value of the likelihood function for the full model L_1 . The likelihood test statistic is given by:

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = -2\left[\log(L_0) - \log(L_1)\right] = -2[LL_0 - (-LL_1)] \quad \text{……(3.10)}$$

Where LL_0 the log likelihood value of the model which is have the intercept term only and LL_1 is the log likelihood value of the full model. The likelihood ratio statistic has a chi-square distribution and it tests the null hypothesis that all logistic regression coefficients except the constant are zero. The degrees of freedom are obtained by differencing the number of parameters in the both model. It is compared with chi-square value at the difference between degree of freedom of both models. If p-value is less than 5 % level of significance leads the rejection of the null hypothesis that all the predictor effects are zero. When this likelihood test is significant, at least one of the predictors is significantly related to the response variable (Hosmer and Lemeshow, 2000).

3.7.1.2.3. Hosmer-Lemeshow Test

The Hosmer-Lemeshow test is used to check the overall model fit. In this approach, data are divided into ten groups. From each group, the observed and expected number of events are computed. Then, the Hosmer-Lemeshow test statistic is given by:

$$\hat{C} = \frac{\sum_{j}^{g}(O_j - E_j)^2}{V_j} \qquad \text{.....(3.11)}$$

Where, $E_j = np_j$

$V_j = np_j(1-p_j)$

g is the number of groups

O_j is observed number of events in the j^{th}group

E_j is expected number of events in the j^{th} group and

V_j is a variance correction factor for the j^{th} group.

If the observed number of events differs from what is expected by the model, the statistic $\hat{C}$ will be large and there will be evidence against the null hypothesis that the model is adequate to fit the data. This statistic has an approximate chi-square distribution with (g-2) degrees of freedom (Agresti, 1996).

3.7.1.2.4. R- Square Statistic

A number of measures have been proposed in logistic regression as an analog to R^2 in multiple linear regressions. The Cox and Snell R^2 are calculated using log-likelihood functions and the sample size. The maximum value that the Cox & Snell R^2 attains is less than 1. The Nagelkerke R-square is an adjusted version of the Cox & Snell R-square and covers the full range from 0 to 1, and therefore it is often preferred (Bewick and Jonathan, 2005).

The Cox and Snell R-square is given by:

$$Cox\ \&\ Snell\ R^2 = 1 - \left[\frac{-2LL_{null}}{-2LL_{full}}\right]^{2/n} \quad \text{.....(3.12)}$$

The Nagelkerke R –square is an adjusted version of the Cox & Snell R-square and covers the full range from 0 to 1 and it is given by:

$$\text{Nagelkerke}\ \ R^2 = \frac{1 - \left[\frac{-2LL_{null}}{-2LL_{full}}\right]^{2/n}}{1 - \left(-2LL_{null}\right)^{2/n}} \quad \text{.....(3.13)}$$

Where: LL_{null} is log-likelihoods of the null model or the logistic model with the constant only.

LL_{full} is log-likelihoods of the full logistic regression model (the logistic regression *model* contains all the *k* predictors).

3.7.2. Multilevel Model

Multilevel models have become popular for the analysis of a variety of problems. Multilevel models are specifically geared toward the analysis of data that have a hierarchical or cluster structure. Such data arise routinely in various fields, for instance in educational research with pupils nested within schools, family studies with children nested within families.

Multilevel analysis is a statistical approach for the analysis of data with complex patterns of variability, with a focus on nested sources of variability. The best way to the analysis of multilevel data is an approach that represents within the group as well as between group relations within a single analysis. It can take into account the variability associated with each level of the hierarchy (Dai *et al.*, 2010). It can also estimate both between group and within group variations, and help to figure out how those levels interact with each other. For this study variation in

mortality of HIV infected patients taking ART within woredas means that not only unexplained variation between HIV infected patients taking ART but also unexplained variation between woredas. Such variation can be analyzed through statistical models known as random coefficients models.

3.7.2.1. Multilevel logistic regression model

Multilevel logistic regression model can be used to predict a dichotomous dependent variable from a set of independent variables. It can be employed in the simplest case without explanatory variables (usually called the empty model) and also with explanatory variables by allowing only the intercept term or both the intercept and slopes (regression coefficients) to vary randomly, and the coefficients are assumed to follow a multivariate normal. In this study for the sake of simplicity of discussion on multilevel logistic regression model, two-level models were used.

3.7.2.2. Heterogeneous proportions

The basic data structure of the two-level multilevel logistic regression is a collection of N groups (woredas units at level-two) and within-group j (j =1, 2,…, N) a random sample of n_j level-1 units (HIV infected patients). The outcome variable, mortality of HIV infected patients, is dichotomous and is denoted by Y_{ij} for patient i in woreda j (i=1, 2,…,n_j, j=1, 2, …, N) and Y_{ij} coded as 0(alive) and 1(dead).The total sample size is denoted by $M = \sum_{j=1}^{N} n_j$.

For the proper application of multilevel analysis, the first logical step is to test heterogeneity of proportions between the groups (woredas). To test whether there are systematic differences between the groups, the well-known chi-square test for contingency table can be used. The test statistic is:

$$\chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) \quad \text{.....(3.14)}$$

Where O is observed and E is the expected count in the cell of the contingency table. This can also be written as follows:

$$\chi^2 = \frac{\sum_{j=1}^{N} n_j\left(\hat{p}_j - \hat{p}_.\right)^2}{\hat{p}_.(1-\hat{p}_.)} \quad \text{.....(3.15)}$$

Where, $\hat{p}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$ .....(3.16) is the proportion of patients who are dead in woreda j.

$\hat{p}_. = \frac{1}{M}\sum_{j=1}^{N}\sum_{i=1}^{n_j} Y_{ij}$ .....(3.17) is the overall proportion of the patients who are dead.

The statistic above in equation (3.14) (χ^2 chi-square statistic) follows approximately central chi-square distribution with N −1 degrees of freedom.

3.7.2.3. Estimation of between and within-groups variance

The true variance between the group-dependent probabilities var(p_j) can be estimated by:

$$\hat{\tau}^2 = S^2_{between} - \frac{S^2_{within}}{\tilde{n}} \quad \text{.....(3.18)}$$

Where, $\tilde{n} = \frac{1}{N-1}\left\{M - \frac{\sum_{j=1}^{N} n_j^2}{M}\right\}$

The between-groups variance is closely related to the chi-squared test statistic.

$$S^2_{between} = \frac{\hat{p}_.(1-\hat{p}_.)}{\tilde{n}(N-1)}\chi^2 \quad \text{.....(3.19)}$$

Where, χ^2 is given in equation (3.14)

The within groups' variance is a function of the group averages.

$$S^2_{within} = \frac{1}{M-N}\sum_{j=1}^{N} n_j \hat{p}_j(1-\hat{p}_j) \quad \text{.....(3.20)}$$

3.7.2.4. The Multilevel Empty Logistic Regression Model

Empty model is a model that contains no explanatory variables. The empty two-level model for a binary outcome variable refers to population of groups (level-two units) and specifies the probability distribution for the group-dependent probabilities in equation $Y_{ij} = p_j + \varepsilon_{ij}$ without taking further explanatory variables in to account. It can be expressed with logit link function as:

$$\log it(p_j) = \beta_0 + U_{0j} \quad \text{.....(3.21)}$$

Where, Y_{ij} is the outcome for individual i in group j.

p_j is the probability (average proportion of successes) in group j.

ε_{ij} is individual-dependent residual.

β_0 is the population average of the transformed probabilities and

U_{0j} is the random deviation from this average for group j.

The above model in equation (3.21) does not include a separate parameter for the level-one residual variance of the dichotomous outcome variable follows directly from the success probability $\text{Var}(\varepsilon_{ij}) = p_j(1 - p_j)$. The residual ε_{ij} 's are assumed to have mean zero and variance σ_{ε}^2 (Snijders and Bosker, 1999).

3.7.2.5. Intra-class Correlation Coefficient

The intra-class correlation coefficient (ICC) measures the proportion of variance in the outcome explained by the grouping structure. ICC can be calculated using an intercept-only model as follows:

$$\text{ICC} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2} \quad \text{.....(3.22)}$$

Where, σ_e^2 is variance of individual (lower) level units and σ_0^2 is between group variance.

Since the logistic distribution for the level-one residual variance implies a variance of $\pi^2/3 \approx 3.29$ (Snijders and Bosker, 1999), equation (3.22) can be written as follows:

$$\text{ICC}=\frac{\sigma_0^2}{\sigma_0^2+3.29} \quad \ldots\ldots(3.23)$$

3.7.2.6. The Random Intercept Logistic Regression Model

The random intercept logistic regression model expresses the logit of p_{ij}, as a sum of a linear function of the explanatory variables and random group-dependent deviation U_{0j}.

Assume k explanatory variables $X_1, X_2,..., X_k$. The values of $X_h (h=1,2,...,k)$ are indicated in the usual way by $X_{hij} (h=1,...,k; i=1,2,...,n_j; j=1,2,...,N)$ Since some or all of these variables could be level-one variables, the success probability is not necessarily the same for all individuals in a given groups. Therefore, the success probability depends on the individual as well as on the group, and is denoted by p_{ij}. The outcome variable is expressed as the sum of success probability and a residual term ε_{ij}.

$$\log it(p_{ij}) = \left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + ... + \beta_k x_{kij} \quad \ldots\ldots(3.24)$$

Equation (3.24) does not include a level-1 residual because it is an equation for the probability p_{ij} rather than for the outcome Y_{ij} (Snijders and Bosker, 1999).

By letting $\beta_{oj} = \beta_0 + U_{0j}$ in equation (3.24) we have the following:

$$\log it(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + U_{0j} \quad \ldots\ldots(3.25)$$

Or

$$p_{ij} = \frac{e^{\beta_0 + \sum_{h=1}^{k} \beta_k x_{hij} + U_{0j}}}{1 + e^{\beta_0 + \sum_{h=1}^{k} \beta_k x_{hij} + U_{0j}}} \quad \ldots\ldots(3.26)$$

Where, $\beta_0 + \sum_{h=1}^{K} \beta_h x_{hij}$ is the fixed part of the model and U_{0j} is the random part of the model.

Thus, unit difference between the χ_h values of two individuals in the same group is associated with a difference of β_h in their log-odds, or equivalently, a ratio of $\exp(\beta_h)$ in their odds. The deviations U_{0j} are mutually independent with zero mean and variance σ_0^2.

3.7.2.7. The Random Coefficient Logistic Regression Model

In logistic regression analysis, linear models are constructed for the log-odds. The multilevel analogue, random coefficient logistic regression, is based on linear models for the log-odds that include random effects for the groups or other higher level units.

Let $X_1, X_2, ..., X_k$ denote the explanatory variables which are potential explanations for the observed outcomes. The values of $X_h (h = 1,2, ..., k)$ are indicated in the usual way by X_{hij}. Since some or all of these variables could be level-one variables, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success probability depends on the individual as well as the group, and is denoted by p_{ij}. Considering a model with group-specific regressions of logit of the success probability, $\log it(p_{ij})$ on a single level-one explanatory variable X.

$$\log it(p_{ij}) = \left(\frac{p_{ij}}{1-p_{ij}} \right) = \beta_{0j} + \beta_{1j} x_{1ij} \quad \text{.....(3.27)}$$

Where, the intercepts β_{oj} and slopes β_{1j} are group-dependent. These group-dependent coefficients can be split into an average coefficient and the group dependent deviation as follows:

$\beta_{oj} = \beta_0 + U_{0j}$ and $\beta_{1j} = \beta_1 + U_{1j}$

Replacing $\beta_0 + U_{0j}$ and $\beta_1 + U_{1j}$ in place of β_{oj} and β_{1j} respectively in equation (3.27),

we get: $$\log it(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}} \right) = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j}) x_{1ij} \quad \text{.....(3.28)}$$

Where, β_1 - is the average regression coefficient

β_0 - is the average regression intercept

There are two random group effects, the random intercept U_{0j} and the random slope U_{1j}. It is assumed that the level-two residuals U_{0j} and U_{1j} have both zero mean given the value of the explanatory variable X .The term $U_{1j}x_{1j}$ can be regarded as a random interaction between group and X . This model implies that two random effects characterize the groups: their intercept and their slope. These two groups' effects (U_{0j},U_{1j}) are independent and identically distributed. The variances and covariance of the level-two random effects (U_{0j},U_{1j}) are denoted as follows:

$\text{var}(U_{0j}) = \sigma_{00} = \sigma_0^2$

$\text{var}(U_{1j}) = \sigma_{11} = \sigma_1^2$

$\text{var}(U_{0j}, U_{1j}) = \sigma_{01}$

We can extend the model for a single explanatory variable discussed above by including more variables that have random effects. Assume that there are *k* level-1 explanatory variables $X_1, X_2, ..., X_k$,the model where all X - variables have varying slopes and random intercept.

That is, $\log it(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + ... + \beta_{kj}x_{kij}$ ……(3.29)

Replacing $\beta_0 + U_{0j}$ and $\beta_h + U_{hj}$ in places of β_{oj} and β_{hj} respectively in equation (3.29), we have the following:

$$\log it(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \sum_{h=1}^{k}\beta_h x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj}x_{hij} \quad \text{……..(3.30)}$$

Where, $\beta_0 + \sum_{h=1}^{k}\beta_h x_{hij}$ is the fixed part of the model.

$U_{0j} + \sum_{h=1}^{k} U_{hj}x_{hij}$ is the random part of the model.

$U_{0j}, U_{1j}, \ldots, U_{kj}$ are assumed to be independent between groups but may be correlated within groups. Thus, the components of the vector $(U_{0j}, U_{1j}, \ldots, U_{kj})$ are independently

distributed as a multivariate normal distribution with zero mean vector and variances and co-variances matrix Ω given by:

$$\Omega = \begin{pmatrix} \sigma_0^2 & \sigma_{10} & \cdots & \sigma_{k0} \\ \sigma_{01} & \sigma_1^2 & \cdots & \sigma_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{0k} & \sigma_{1k} & \cdots & \sigma_k^2 \end{pmatrix}$$

3.7.2.8. Estimation of Multilevel Logistic Regression Model

Multilevel models are also generally estimated using maximum likelihood methods, and combining multilevel and generalized linear models leads to complex models and estimation procedures. The most frequently used methods are based on a first-or second order Taylor expansion of the link function. Marginal Quasi-likelihood (MQL) (Goldstein, 1991; Goldstein and Rasbash, 1996) and Penalized Quasi-likelihood (PQL) (Breslow and Clayton, 1993) are two of the widely applied methods of approximation procedures. Marginal quasi-likelihood (MQL) involves expansion around the fixed part of the model, whereas penalized or predictive quasi-likelihood (PQL) involves fixed plus random part (Goldein, 1991; Goldein and Rasbash, 1996).

The nonlinear function is linearized using an approximation known as Taylor series expansion. Taylor series expansion approximates a nonlinear function by an infinite series of terms. Most commonly only the first term of the series is used, which is referred to as a first order Taylor approximation. And, when the second term is also used, we have a second order Taylor approximation, which is generally more accurate. Both MQL and PQL rely on the Taylor expansion to achieve the approximation. There are different methods of parameter estimations which are implemented in various software packages like MLwiN, LISREL, STATA and SAS among others. In this study, the multilevel data analysis is supported by the software packages STATA and SAS.

3.7.2.9. Model Comparison

When fitting several models to the same set of data, it is helpful to compare those models using summary measures of fit. In this study we used the two common measures for comparing models: Akaikes Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The Akaikes Information Criterion(AIC) is defined as:

$$AIC = -2 (\ln(likelihood)) + 2k$$

The Bayesian Information Criterion (BIC) is also defined as:

$$BIC=-2(\ln(likelihood)) + k \ln(N)$$

Where,

k- is the model degrees of freedom calculated as the rank of variance–covariance matrix of the parameters.

N- is the number of observations used in estimation.

Then, the model with smallest value of the Information Criterion is considered as the best model (Schwarz, 1978).

3.8. Ethical Considerations

The official ethical clearance for this study was provided by research ethics review board of Jimma University and the Department of Statistics has written an official support letter to Hossana Queen Elleni Mohamad Memorial Hospital. The official ethical clearance was also obtained from Hossana Queen Elleni Mohamad Memorial Hospital medical director. To keep the confidentiality, data clerks working in ART clinic of the hospital extracted the data from the patients' cards. Moreover, no personal identifier was used on data collection form. The recorded data has not been accessed by a third person except the principal investigator, and has been kept confidentially.

CHAPTER FOUR

4. STATISTICAL DATA ANALYSIS AND DISCUSSION

4.1. Descriptive Analysis

A total of 400 HIV infected patients aged 15 years or older taking ART in Hosana District Queen Elleni Mohamad Memorial Hospital were eligible for this study. Among the patients, 75 (18.75%) were dead while 325(81.25%) are alive. Summary statistics of factors associated with mortality of HIV infected patients are presented in Table 4.1.

Table 4.1: Summary of descriptive statistics of mortality of HIV infected patients taking ART

Variables	**Categories**	**Dead**			
		No	**Yes**	**Total**	**Percentage of death**
Woreda	Hossana town administration	52	19	71	26.8%
	Lemmo	44	7	51	13.7%
	Anlemmo	34	11	45	24.4%
	Misha	41	12	53	22.6%
	Angecha	31	6	37	16.2%
	Duna	29	4	33	12.1%
	Lera	40	5	45	11.1%
	Gombora	30	6	36	16.7%
	Soro	24	5	29	17.2%
Sex	Female	182	37	219	16.9%
	Male	143	38	181	21.0%
Age	Below 40	241	51	292	17.5%
	40 or older	84	24	108	22.2%
Residence	Rural	159	34	193	17.6%
	Urban	166	41	207	19.8%
Marital Status	Never married	63	18	81	22.2%
	Married	198	33	231	14.3%
	Separated	15	10	25	40.0%

	Divorced	20	4	24	16.7%
	Widowed	29	10	39	25.6%
Level of education	No education	65	19	84	22.6%
	Primary	130	33	163	20.2%
	Secondary or above	130	23	153	15.0%
Tobacco	No	254	40	294	13.6%
	Yes	71	35	106	33.0%
Alcohol	No	254	47	301	15.6%
	Yes	71	28	99	28.3%
Baseline weight	Less than 50 kg	108	41	149	27.5%
	50 kg or above	217	34	251	13.5%
WHO Clinical stage	Stage I	93	11	104	10.6%
	Stage II	107	19	126	15.1%
	Stage III	102	39	141	27.7%
	Stage IV	23	6	29	20.7%
TB Status	Negative	256	34	290	11.7%
	Positive	69	41	110	37.3%
Antiretroviral Regimen	d4t-3TC-NVP or d4t-3TC-EFV	65	29	94	30.9%
	AZT-3TC-NVP or AZT-3TC-EFV	119	27	146	18.5%
	TDF-3TC-NVP or TDF-3TC-EFV	141	19	160	11.9%
Baseline CD4 count	Less than 200	123	49	172	28.5%
	200 or above	202	26	228	11.4%

According to Table 4.1, the proportion of mortality of HIV infected patients taking ART in Hossana Disrict Queen Elleni Memorial hospital varies from one woreda where they live to the other. For instance, the highest percentage of mortality was observed in Hossana town adminstration (26.8%) followed by Anlemmo woreda (24.4%) where as the lowest percentage was observed in Lera worea (11.1%) followed by Duna woreda(12.1%).

Similarly, the proportion of mortality of HIV infected patients taking ART varies by place of residence: urban and rural. The higher percentage of mortality (19.8%) of patients was observed in urban areas, and relatively smaller percentage of mortality (17.6%) is observed in rural areas.

The proportion of the mortality of HIV infected patients varies by sex and age groups. The percentage mortality of male patients is higher (21.0%) than the percentage mortality of female patients (16.9%).With regard to their age, the percentage of mortality of patients was higher for patients aged 40 or older years (22.5%) than below 40 years (17.5%).

The proportion of mortality of HIV infected patients taking ART also varies by marital status. For example, the highest percentage of mortality of patients was observed for patients who were separated (40.0%) followed by Widowed (25.6%) while lowest percentage of mortality was observed for patients who were married (14.3%) followed by Divorced (16.7%).

The proportion of mortality of HIV infected patients taking ART varies by level of education. The percentage of mortality of patients was highest for patients who have no education (22.6%) as opposed to the lowest percentage of mortality of patients for those patients whose level of education was secondary or above (15.0%).

The proportion of mortality of HIV infected patients taking ART is higher for the patients who smoke Tobacco (33.0%) than for patients who do not smoke (13.6%).Similarly; it is higher for the patients who drink alcohol (28.3%) than for patients who do not drink (15.6%).

The proportion of mortality of HIV infected patients taking ART also varies with baseline weight. The higher percentage of mortality of patients (27.5%) was observed for patients whose baseline weight was less than 50kg, and relatively small proportion of mortality (13.5%) was observed for patients whose baseline weight is 50 kg or above.

The proportion of mortality of HIV infected patients taking ART varies with their baseline CD4 count. The percentage of mortality of the patients was higher for those patients whose baseline CD4 count was less than 200(28.5%) than the percentage of mortality of the patients whose baseline CD4 count was 200 or above (11.4%).

The proportion of mortality of HIV infected patients taking ART varies with antiretroviral regimen they take and WHO clinical stage. The percentage of mortality of patients was highest

for patients who took antiretroviral regimen d4t-3TC-NVP or d4t-3TC-EFV (30.9%) followed by AZT-3TC-NVP or AZT-3TC-EFV (18.5%) and the lowest percentage was observed for patients who took TDF-3TC-NVP or TDF-3TC-EFV (11.9%).With regard to WHO clinical stage, the percentage of mortality of patients was highest for patients with WHO clinical stage III (27.7%) where as the lowest percentage of mortality of patients was observed for patients with WHO clinical stage I (10.6%).

The proportion of mortality of HIV infected patients taking ART also varies with their TB status. The percentage of mortality of patients was higher for patients who were TB positives (37.3%) while the lower percentage of mortality of patients was observed for patients who were TB negatives (11.7%).

4.2. Logistic Regression Analysis of Mortality HIV infected Patients

Multiple logistic regression analysis was used to analyze the effect of each of independent variables on mortality of HIV infected patients while controlling for the other independent variables. Stepwise method of variable selection procedure was employed to select the important determinants of mortality of HIV infected patients. The significance of individual parameter estimates are tested using Wald test. A negative sign in column labeled estimate indicates an inverse relationship of explanatory variable with the log odds of the dependent variable whereas a positive coefficient column labeled estimate indicates a positive relationship to the log odds of the dependent variable.

Table 4.2: Results of Logistic Regression Analysis for mortality of HIV Infected Patients

Parameter	Estimate	Standard error	Wald chi-quare	df	Pr > ChiSq	OR	95% C.I of OR
Intercept	-2.1281	0.4021	28.0016	1	0.024*	-	-
Age			8.1068	1	0.013*		
Below 40(Ref)							
40 or older	0.5479	0.1924	8.1068	1	0.013*	1.7296	1.1323 1.9159
Woreda			71.3591	8	0.014*		
Hossana(Ref)							

Lemmo	-1.2691	0.3016	17.7015	1	0.029*	0.2811	0.0899	0.8789
Anlemmo	-0.3937	0.2115	3.4519	1	0.442	0.6745	0.2475	1.8384
Misha	-1.7376	0.4864	12.7481	1	0.031*	0.1759	0.0315	0.7212
Angecha	-0.6617	0.6065	1.1864	1	0.275	0.5160	0.1572	1.6938
Duna	-0.9383	0.6486	2.0768	1	0.147	0.3913	0.1098	1.3949
Lera	-1.2322	0.6101	4.0691	1	0.043*	0.2917	0.0882	0.9643
Gombora	-2.8982	0.6322	21.0014	1	0.002*	0.0551	0.0103	0.6262
Soro	-0.9176	0.6496	1.9749	1	0.158	0.3995	0.1118	2.4270
Led			37.8264	2	0.028*			
No education(Ref)								
Primary	-1.2399	0.3942	9.8717	1	0.043*	0.2894	0.0346	0.7093
Secondary or above	-2.0629	0.4147	24.7295	1	0.007*	0.1271	0.0275	0.1557
Alch			14.3186	1	0.000*			
No(Ref)								
Yes	1.2099	0.3194	14.3186	1	0.000*	3.3535	1.7931	6.2715
Baslinwt			7.9649	1	0.012*			
Less than 50(Ref)								
50kg or above	-0.8612	0.3049	7.9649	1	0.031*	0.4226	0.2325	0.7682
TBStas			18.7618	1	0.000*			
Negative(Ref)								
Positive	1.3231	0.3052	18.7618	1	0.000*	3.7551	2.0644	6.8303
BaslinCD4			11.0624	1	0.032*			
Less than 200(Ref)								
200 or above	-1.4118	0.4243	11.0624	1	0.032*	0.2437	0.1922	0.5016

(* =Significant at 5% level) and (Ref =Reference category)

The result presented in Table 4.2 is interpreted in terms of odds ratio. Odds ratios greater than one indicate that the event is more likely to happen in the comparator than in the reference category, odds ratios of one indicate the event is exactly as likely to occur in the two categories, while odds ratios less than one indicate that the event is less likely to happen in the comparator than in the reference category.

According to Table 4.2, age, woreda, level of education, alcohol, baseline weight, TB status and baseline CD4 count were found to be significant predictors of mortality of HIV infected patients at 5% level of significance but Anlemmo woreda, Angecha woreda, Duna woreda and Soro woreda are insignificant when compared to Hossana town administration.

The estimated model is given by:

$$logit(\hat{p}) = \beta_0 + \sum_{h=o}^{1} \beta_{1h} Age_h + \sum_{i=0}^{8} \beta_{2i} Wrda_i + \sum_{j=o}^{2} \beta_{3j} Led_j + \sum_{k=0}^{1} \beta_{4k} Alch_k + \sum_{l=0}^{1} \beta_{5l} Baslinwt_l + \sum_{m=0}^{1} \beta_{6m} TBStas_m + \sum_{n=0}^{1} \beta_{7n} BaslinCD4_n$$

Where, $\hat{p}$ = predicted probability of mortality of HIV infected patient, β_0 = constant, Age_h = Age of Patient h, $Wrda_i$ = Patient's woreda i, Led_j = Level of education j, $Alch_k$ = Alcohol k, $Baslinwt_l$ = Baseline weight l, $TBStas_m$ = TB status m , $BaslinCD4_n$ = Baseline weight n.

The value of explanatory variable for each category is taken as 1 if the given variable falls in the corresponding category. For instance, Age_h= 1 for patient's age h and Age_h= 0 for others age groups. Similarly, each of the other variables takes value 1 if it falls within the corresponding level of category. Based on the result revealed in Table 4.2, the regression equation consisting of the significant variables is given by:

$$\begin{aligned} logit(\hat{p}) = {} & -2.1281 + 0.5479 Age_{\geq 40} - 1.2691 Wrda_{Lemmo} - 0.3937 Wrda_{Anlemmo} \\ & - 1.7376 Wrda_{Misha} + \cdots - 0.9176 Wrda_{Soro} - 1.2399 Led_{Primary} \\ & - 2.0629 Led_{Secondary} or_{Above} + 1.2099 Alch_{Yes} - 0.8612 Baslinwt_{\geq 50kg} \\ & + 1.3231 TB_{Positive} - 1.4118 BaslinCD4_{\geq 200} \end{aligned}$$

In Table 4.2, binary logistic regression analysis revealed that HIV infected patients who aged 40 years or older were 1.7296 times more likely to die than those patients who aged below 40 years.

Thus, patients who aged 40 years or older were 72.96% (OR=1.7296, 95% CI: 1.1323, 1.9159) more likely to die than those patients who aged below 40 years controlling for other variables in the model.

HIV infected patients who live in Lemmo woreda were 0.2811 times (OR=0.2811, 95% CI: 0.0899, 0.8789) less likely to die than those patients who live in Hossana town controlling for other variables in the model. Thus, patients who live in Lemmo woreda were 71.89% less likely to die compared to patients who live in Hossana controlling for other variables in the model. Also, patients who live in Misha woreda were 0.1759 times (OR=0.1759, 95% CI: 0.0315, 0.7212) less likely to die than those patients who live in Hossana town. Similarly, patients who live in Lera woreda and Gombora wored were 0.2917 times (OR=0.2917, 95% CI: 0.0882, 0.9643) and 0.0551 times (OR=0.0551, 95% CI: 0.0103, 0.6262) respectively less likely to die compared those patients who live in Hossana town controlling for other variables in the model.

HIV infected patients whose level of education was primary were 0.2894 times (OR=0.2894, 95% CI: 0.0346, 0.7093) less likely to die compared to patients who had no education controlling for other variables in the model while patients whose level of education was secondary or above were 0.1271 times (OR=0.1271, 95% CI: 0.0275, 0.1557) less likely to die compared to patients who had no education controlling for other variables in the model.

HIV infected patients whose baseline weight was 50 kg or above were 0.4226 times(OR=0.4226, 95% CI: 0.2325, 0.7682) less likely to die compared to patients whose baseline weight was less than 50 kg controlling for other variables in the model. Also, patients who were TB positive were 3.7551 times (OR=3.7551, 95% CI: 2.0644, 6.8303) more likely to die compared to patients who were TB negative controlling for other variables in the model. Similarly, patients who drink alcohol were 3.3535 times (OR=3.3535, 95% CI: 1.7931, 6.2715) more likely to die compared to those patients who do not drink alcohol controlling for other variables in the model.

HIV infected patients whose baseline CD4 count was 200 or above were 0.2437 times(OR=0.2437 , 95% CI: 0.1922, 0.5016) less likely to die compared to those patients whose baseline CD4 count was less than 200 controlling for other variables in the model.

4.2.1. Assessment of Goodness of Fit of Logistic Regression Analysis

4.2.1.1. Test of significance of relationship between dependent and independent variables

Table 4.3: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr>ChiSq
Likelihood Ratio	75.6263	15	<.0001
Score	76.1143	15	<.0001
Wald	57.8644	15	<.0001

In Table 4.3, Likelihood ratio, Score and Wald tests have P-values less than 0.05 and are significant. This indicates that the final model with explanatory variables was more effective than the null model. Therefore, there is a significant relationship between the dependent variable and the set of independent variables.

4.2.1.2. Likelihood ratio test of overall logistic regression model

Table 4.4: Result of Model fit of Statistics for Intercept only and full model

Criterion	Intercept Only	Intercept and Covariates
AIC	408.062	347.436
BIC	423.054	389.367
-2 Log L	386.062	310.436

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2[LL_0 - (-LL_1)]$$

$$G^2 = 386.062 - 310.436 = 75.626$$

This calculated deviance is compared with Chi-square at difference of full and empty model degrees of freedom (χ^2_{tab}=23.685 at df=14). Since the calculated deviance, χ^2=75.626 is greater than χ^2_{tab}= 23.685, there is an evidence against null hypothesis that there is no difference between model without predictors and model with predictors. This indicates that the fit is adequate and at least one of the predictors is significantly related to the response variable.

4.2.1.3. Hosmer-Lemeshow Goodness of Fit Test

Table 4.5: Hosmer-Lemeshow Test

Chi-square	DF	Pr>Chisq

6.925	8	0.3540

The Hosmer and Lemeshow Goodness-of-Fit Test tests the hypothesis:

H_0: The model is a good fit, vs.
H_A: The model is not a good fit

The value of the Hosmer-Lemeshow goodness-of-fit statistic is $\hat{C}$ =6.925 with the corresponding p-value of 0.3540 (Table 4.5). A large p-value (i.e, p-value >0.05) suggests that the fitted model is an adequate model. Since the p-value = 0.3540 is greater than 0.05, we do not reject the null hypothesis that the model is a good fit. Therefore, binary logistic regression of mortality of HIV infected patients fitted the data very well.

4.2.1.4. Model summary of logistic regression model

The Nagelkerke R-square in Table 4.6 is estimated at 23.2% indicating that explanatory variables are useful in predicting mortality of HIV infected patients taking ART in Hossana Queen Elleni Mohamad memorial hospital. Using binary logistic regression for analyzing those variables having significant association with mortality of patients for our data is appropriate.

Table 4.6: Model Summary of Logistic Regression Model

-2*Log likelihood	Cox & Snell R Square	Nagelkerke R Square
296.979	0.200	0.232

4.2.1.5. Classification Table

Classification table documents the validity of predicted probabilities. With the cutoff set at 0.5, about 96.0% of the patients who are alive were predicted correctly, while 34.7% of patients dead were predicted correctly and the overall correct prediction was 84.5 %(Table 4.7).

Table 4.7: Classification Table

Observed		Predicted		
		Mortality		Percentage Correct
		Alive	Dead	
Mortality	Alive	312	13	96.0
	Dead	49	26	34.7
Overall Percentage				84.5

4.3. Multilevel Logistic Regression Analysis

In the multilevel logistic regression analysis two-level clusters were used with woredas as the level-two units and HIV infected patients as the level-one units i.e. HIV infected patients are nested within woredas. This study was done basically expecting that there is variation in mortality among the woredas. The multilevel logistic regression analysis was used for analysis of the effect of demographic, socio-economic, behavioral and Clinical variables on mortality of HIV infected patients.

A chi-square test was applied to assess the heterogeneity in the proportion of mortality of HIV infected patients taking ART in Hossana Queen Elleni Mohamad memorial hospital among 9 woredas where the patients live. The test yields χ^2= 31.74 which is greater than χ^2_{tab}=15.507 at 8 degree freedom with P-value = 0.0001 implying that there is heterogeneity among woredas where the patients live with respect to mortality of HIV infected patients taking ART in Hossana District Queeen Eleni Mohamad memorial hospital.

4.3.1. Multilevel Empty Logistic Regression Analysis

The multilevel empty logistic regression model contains no explanatory variables. The deviance based on Chi-square is the difference in -2*log likelihood between an empty model without random effect (-2LL=386.06206) and an empty model with random effect (-2LL =382.02234) =4.0397 which is greater than $\chi^2_{tab}(1)$ =3.841 with p-value= 0.032 at 5% level of significance. This implies that an empty logistic regression model with random intercept is better than an empty model without random intercept.

In multilevel logistic regression model, the average overall odds of HIV infected patients mortality is estimated to be β_0 = -1.47552 and the variance of the random intercept is estimated to be σ_0^2 =0.0333485 which is significant at 5% level of significance indicating that the variations of mortality of HIV infected patients among woredas was non-zero (Table 4.8). Also, Intra correlation coefficient (ICC) in the empty model is estimated at 0.0100346 which is found to be significant at 5% level of significance suggesting about 1.003% of the variation in mortality of the patients taking ART in Hossana District Queeen Eleni Mohamad memorial Hospital could be attributed to differences across woredas.

Table 4.8: Result of multilevel empty logistic regression model

Fixed Part	**Coefficient**	**S.E**	**Z-value**	**P-value**	**[95% Conf. Interval]**
β_0 =Intercept	-1.47552	0.141608	-10.42	0.000*	-1.753066 -1.197973
Random Part	**Variance Component**	**S.E**	**Z-value**	**P-value**	**[95% Conf. Interval]**
σ_0^2	0.0333485	0.0261593	1.27	0.016*	0.0005687 0.37756
ICC(ρ)	0.0100346	0.0196158	0.52	0.024*	0.00000983 0.9927858
AIC=390.0223, BIC=398.0053, Log likelihood = -191.01117					

(*=Significant at 5% level of significance)

4.3.2. Random Intercept Logistic Regression Analysis

In the random intercept model the intercept is allowed to vary across woredas and the effects of predictors of HIV infected patients are the same for each woreda .Based on deviance based Chi-square, the difference between deviance of multilevel empty logistic regression model and multilevel random intercept logistic regression model, random intercept logistic regression model was found to give a better fit as compared to the multilevel empty logistic regression model. The deviance based Chi-square = 75.49126 is greater than χ^2 = 14.0671 at 7 degree of freedom with P-value = 0.0000. This indicates that after controlling all predictor variables, the intercept varied across the woredas i.e. the variations of mortality of HIV infected patients among woredas where they live was non-zero. The variance of random intercept of the multilevel random intercept logistic regression model was estimated to be 0.0324 which is decreased by about 0.0009 as compared to that of multilevel empty logistic regression model (Table 4.9). This reduction indicates that there is a contribution of inclusion of fixed independent variables on variations of mortality of HIV infected patients across woredas. The intra correlation coefficient for this model is found to be 0.00976 implying 0.976% of the variation in mortality of HIV infected patients can be explained by grouping patients in woredas.

Since the multilevel random intercept logistic regression model has smaller AIC and BIC than that of multilevel empty logistic regression model, the multilevel random intercept logistic

regression model is a better fit as compared to multilevel empty model for predicting variation of mortality of HIV infected patients among woredas where they live.

Table 4.9: Result of parameter estimate of multilevel random intercept logistic model

Fixed Part	**Coeff.**	**Std.Err.**	**Z**	**P>|Z|**	**OR**	**[95% CI of OR]**
Age						
Below 40(Ref)						
40 or older	0.5483	0.1901	2.88	0.006*	1.7303	1.0273 1.9242
Led						
No education(Ref)						
Primary	-1.2468	0.3914	-3.19	0.023*	0.2874	0.0189 0.8315
Secondary	-2.0819	0.4086	-5.10	0.002*	0.1247	0.0433 0.2578
Alcohol						
No(Ref)						
Yes	1.2482	0.3167	3.94	0.000*	3.4841	1.7874 6.4526
Baslnwt						
Less than 50kg(Ref)						
50 kg or above	-0.8927	0.3032	-2.94	0.012*	0.4095	0.2251 0.9287
TBStas						
Negative(Ref)						
Positive	1.3642	0.3045	4.48	0.000*	3.9126	2.1179 6.9107
BaslnCD4						
Less than 200(Ref)						
200 or above	-1.4321	0.4125	-3.47	0.003*	0.2388	0.2043 0.6004
Constant	-2.6258	0.4030	-6.52	0.014*		
Random part	**Variance component**	**Std.Err.**	**Z**	**P>|Z|**	**[95% Conf. Interval]**	
σ_0^2	0.0324423	0.029782	1.09	0.034*	0.0271	0.7325

ICC (ρ)	0.0097646	0.029411	0.33	0.041*	0.0002 0.4771
AIC	328.5311, BIC=372.4372 , Log likelihood = -153.26554				

(*= Significant at 5%) and (Ref =Reference category)

According to Table 4.9, age, level of education, alcohol, baseline weight, TB status and baseline CD4 count were found to be significant predictors of mortality of HIV infected patients at 5% level of significance indicating effects on mortality of HIV infected patients and also contributing to mortality variations of HIV infected patients among woredas where they live.

The result in Table 4.9 revealed that the odds of death of HIV infected patients who aged 40 or older years is 1.7303 times (OR=1.7303, 95% CI: 1.0273, 1.9242) higher than patients aged below 40 years controlling for other variables in the model. The odds of death of patients who drink alcohol is 3.4841 times (OR=3.4841, 95% CI: 1.7874, 6.4526) higher than patients who do not drink controlling for other variables in the model. Similarly, the odds of death of patients who are TB positive is 3.9126 times (OR=3.9126, 95% CI: 2.1179, 6.9107) higher than patients who are TB negatives controlling for other variables in the model.

The odds of death of patients whose level of education are primary and secondary or above are reduced by 71.26% and 87.53% respectively compared to patients with no education controlling for other variables in the model. The odds of death of patients whose baseline weight is 50 kg or above is reduced by 59.05% compared to patients whose weight less than 50 kg controlling for other variables in the model. Similarly, the odds of death of patients whose CD4 count is 200 or above is reduced by 76.12% patients whose CD4 count is less than 200 controlling for other variables in the model.

4.3.3. Random coefficient multilevel Logistic regression Analysis

In random coefficient multilevel logistic regression model both intercept and slopes are allowed to vary across woredas. We analyzed for each of the explanatory variables separately to check the significance of effect of those variables on mortality of HIV infected patients. The deviance based chi-square test was used to test whether the effect of age, level of education, drinking alcohol, baseline weight, TB status, and baseline CD4 count varies across woredas. The deviance based chi-square test statistic is calculated as two times the difference in the log likelihood values between the model with and without the random slope.

The values of deviance based chi-square for age, level of education, alcohol ,baseline weight , TB status and baseline CD4 count are 1.1854(with the corresponding p-value of 0.3137), 0.0003(with the corresponding p-value of 0.8005), 3.0119(with the corresponding p-value of 0.2971), 0.5173(with the corresponding p-value of 0.9104), 0.6241(with the corresponding p-value of 0.5325) and 0.0023(with the corresponding p-value of 0.7413) respectively (Appendix B: Table B2). For each the p-value is large (i.e., greater than 0.05) indicating that the variance of random slope of each variable is not significantly different from zero at 5% level of significance. This implies that the coefficients of all variables do not vary across woredas.

4.3.4. Multilevel Logistic Regression Model Comparison

Table 4.10: Comparison of Multilevel Logistic Regression Models with respect to AIC and BIC

Model	AIC	BIC
Empty model	390.0223	398.0053
Random Intercept	328.5311	372.4372
Random coefficient(Age)	351.3457	383.2347
Random coefficient(Led)	352.5308	384.4198
Random coefficient(Alcohol)	349.5192	381.4082
Random coefficient(Baslinwt)	352.0138	383.9028
Random coefficient(TB status)	351.9070	383.7960
Random coefficient(BaslinCD4)	352.5308	384.4198

The model with smallest AIC and BIC is considered as the best model. Accordingly, random intercept logistic regression model is the best model to explain variation of mortality of HIV infected patients taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital among woredas as compared to other multilevel models.

4.4. Discussion

This study was aimed to identify and describe factors associated with mortality among HIV infected patients who are taking ART in Hossana Queen Elleni Mohamad Memorial Hospital. The results of the study revealed that, out of a sample of 400 HIV infected patients taking ART from March 2009 up to May 2015, 18.75% were dead while 81.25% are alive.

The study revealed that factors age, level of education, alcohol, baseline weight, TB status and Baseline CD4 count and Woreda had significant effect on mortality of HIV infected patients at 5% level of significance.HIV infected patients who live in Lemmo, Misha, Lera, Gombora woredas were less likely to die compared to patients who live in Hossana town, while mortality of HIV infected patients who live in Anlemmo, Angecha, Duna, Soro woreda were not significantly different compared to those patients who live in Hossana town.

In the multilevel logistic regression analysis empty model, random intercept model and random coefficient model were fitted so as to explain variations in mortality of HIV infected patients among woredas where they live. Multilevel model building started from empty model which is fitted without any covariates but with random intercept only. This model was used to see the significance of multilevel model over empty model without random effect. The second model is random intercept model which include covariates and random intercept. Finally, we built random coefficient model that both intercept and slope vary. This study has shown that the multilevel random intercept model fitted the data set very well. The result of multilevel random intercept model revealed that the overall mean of mortality of HIV infected patients varied across the woredas where they live.

The study revealed that the risk of death of older HIV infected patients was significantly different from younger patients. The older patients were more likely to die compared to the younger patients. This is consistent with the studies done previously by Lawn *et al.,* (2009) and Liu *et al*., (2006). This may be due to older aged patients progressed to AIDS at a faster rate than younger aged patients and older patients may have a reduced capacity to generate new CD4 cells in response to viral killing.

The study also revealed that the HIV infected patients whose level of education was secondary or above were less likely to die compared to those patients with no education. This result is consistent with studies done previously by DeSilva *et al*., (2009). Similarly, our study revealed that the patients whose level of education was Primary were less likely to die compared to those patients with no education which is inconsistent result with previous study by Reda *et al*., (2013) revealed that patients with primary education were more likely to die than illiterate patients (patients with no education).

According to previous studies by Liu *et al.*, (2006) and Moattia *et al.*, (2000) alcohol and other substances abuse were associated with mortality, non adherence to medication and lower quality of life of HIV infected patients. Our study showed similar result that HIV infected patients who drink alcohol were more likely died compared to those patients who do not drink alcohol. This might be due to ARV non-adherence in addition to the complications that alcohol brings in to one's health.

Similar with the previous study by Gezahegn (2011) and Moha *et al.*, (2007), this study confirmed that HIV infected patients who were TB positives at ART initiation were more likely to die compared to those HIV infected patients who were TB negatives. This might be due to the fact that TB is the leading cause of death worldwide in HIV-infection and mycobacterium tuberculosis is a virulent organism that can produce disease in HIV-infected persons at any stage of disease even when the immuno suppression is minimal.

The study also revealed that the HIV infected patients with low baseline CD4 count were more likely to die compared to those patients with high baseline CD4 count. The result is consistent with findings from the studies done previously by Moha *et al.*, (2007), Ferradini *et al.*, (2006) and Pathipvanich *et al.*, (2003). Similarly, our study revealed that patients with low baseline weight were more likely to die compared to those patients with high baseline weight. This result confirmed similar findings from previous study by Nuredin (2007).

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This study revealed that the effect of demographic, socio-economic, behavioral and clinical factors on mortality of HIV infected patients. The result of multiple logistic regression revealed that age, level of education, alcohol, baseline weight, TB status and Baseline CD4 count and Woreda had significant effect on mortality of HIV infected patients who are taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital.

The multilevel logistic regression models was applied and showed that there were mortality variation of HIV infected patients among woredas where they live. This variation among woredas is accounted by the random intercept of the model. Moreover, the overall variance of the constant term in both multilevel empty logistic regression model and multilevel random intercept logistic regression model was found to be significant implying the presence of mortality variation across woredas. Age, level of education, alcohol, baseline weight, TB status and baseline CD4 count were significant determinants of variations of mortality of HIV infected patients among woredas.

Among the models that were applied in this study, the multilevel random intercept logistic regression model best fits to the data for mortality of HIV infected patients who are taking ART in Hossana District Queen Elleni Mohamad Memorial Hospital.

5.2. Recommendation

Based on the results of this study the following recommendations are suggested.

- ✓ Health workers should be cautious when a patient has lower baseline CD4 and lower baseline weight.
- ✓ Level of education of the patients has an important role in increasing their quality of life (reduction of mortality). Health workers need to support those patients with no or little education by continuous awareness creation of taking care of themselves and knowing what factors facilitate death.
- ✓ Patients who drink alcohol need to be given advice to reduce excessive drinking.

REFERENCE LISTS

Agresti, A. (1996). An Introduction to Categorical data Analysis. New York. John Wiley and Sons.

AIDS Resource Center, Johns Hopkins University, Bloomberg School of Public Health, Center for Communication Programs 2005, *Ethiopian National ART Strategic ommunication Framework*, website: http://www.etharc.org.

Austin, P., Goel, V. and van Walraven, C. (2001). An introduction to multilevel regression models. Canadian journal of public health, 92(2), 150.

Bajaria, S. (2002).Dynamics of Naïve and Memory CD4+ T Lymphocytes in HIV-1 Disease Progression. Journal of Acquired Immune Deficiency Syndrome, **30**:41-58.

Bewick, L. and Jonathan, B. (2005). Statistics Review 14: Logistic Regression.

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

Degu, J., Endale, A., Hailu, Y., Lindtjorn, B. (2006). Predictors of early death in a cohort of Ethiopian patients treated with HAART. BMC Infectious Diseases; **6**:136.

DeSilva, M., Merry, S., Fischer, P., Rohrer, J., Isichei, C., & Cha, S.(2009). Youth, unemployment, and male gender predict mortality in AIDS patients started on HAART in Nigeria. AIDS care, 21(1), 70-77.

Elke, K., Yirga, A., Katherine, D., Peter, G, Tesfaye, A. and Bud, C. (2011). Implications of adopting new WHO guidelines for antiretroviral therapy initiation in Ethiopia. Bulletin of the World Health Organization; **90**:659-663. doi: 10.2471/BLT.11.089599.

Elly,T., Katabira, F., Moses, R., Kamya, M., Med, M., MPH, I., Med, A.(2008).Ugandan Antiretroviral Treatment and Care Guidelines for Adults, Adolescents, and Children. Kampala, Uganda.

Federal HIV/AIDS Prevention and Control Office (2010). Report on progress towards implementations of the UN declaration of commitment on HIV/AIDS.Ethiopia, Addis Ababa.

Federal HIV/AIDS Prevention and Control Office (2007).*Report on progress towards implementation of the UN Declaration of Commitment on HIV/AIDS*.

Ferradini, L., Jeannin, A., Pinoges, L., Izopet, J., Odhiambo, D., Mankhambo, L., Karungi, G.,Szumilin, E., Balandine, S., Fedida, G., Carrieri, M. P., Spire, B., Ford, N., Tassie, J.M., Guerin, P.J., Brasher,C. (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. Lancet; **367**: 1335–42

Fidell, B. and Tabachnick, L. (2007). Using multivariate statistics (5th ed.). Boston . Montreal: Pearson/A & B. ISBN 0-205-45938-2.

Françoise, R., Chris, D., Stephen, K., Sigrid, T., Joseph, P. (2006). Use of antiretroviral therapy in resource-limited countries: distribution and uptake of First and second line regimens. Department of HIV/AIDS, World Health Organization, CH-1211 Geneva 27,Switzerland; Email: theryf@who.int.

Gelman, A. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. New York : Cambridge University Press, 2006.

Gelman, A. and Hill, J. (2006). Data Analysis Using Regression and Multilevel or Hierarchical Models. Columbia University, Journal of the American Statistical Association.

Gezahegn, A. (2011). Survival Status among patient living with HIV AIDS who are on ART treatment in Durame and Hossana Hospitals. DOI: 10.1371/journal.pone.0087392 .doi:10.1371/journal.pone.0087392.t002.

Gibbons, R. and Bock, R. (1987). Trend in correlated proportions. Psychometrika, 52:113–124.

Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data.Biometrika.**78**: 45-51.

Goldstein, H. (2003). Multilevel Statistical Models. 3rd Ed. Oxford University Press.

Goldstein, H. and Rasbash, J. (1996).Improved approximations for multilevel models with binary responses.*J. Roy. Statist.Soc. A* 159, 505-513. guideline for a public health approach. 2003 revision.

Hosmer, W. and Lemeshow, S. (2000), Applied Logistic Regression.2nd Ed., John Wiley and Sons, New York.

Hox, J. (2010). Multilevel Analysis Techniques and Applications.2nd Ed. New York. Utrecht University.Routledge.

Jamison, D., Breman, J., Measham, A. (2006). Disease Control Priorities in Developing Countries, 2nd edition.

Johannessen, A., Naman, E., Ngowi, B., Sandvik, L., Matee, M., Aglen, H., Gundersen, S., Bruun, J. (2008). Predictors of mortality in HIV-1 infected patients starting antiretroviral therapy in a rural hospital in Tanzania. BMC Infectious Diseases, **8**:52.

Ketema, K. (2011). Determining factors that affect the survival rate of HIV-infected patients on ART: The case of Armed Forces General Teaching Hospital, Addis Ababa, Ethiopia. *Ethiop. J. Health Dev*; **22**(1).

Lawn, S., Little, F., Bekker, L., Kaplan, R., Campbel, E., Orrell, C. and Wood, R. (2009).Changing mortality risk associated with CD4 cell response to antiretroviral therapy in South Africa. AIDS; **23**:335–342.

Lawn, S., Badri, M., Wood, R. (2005). Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort AIDS; **19**:2109–2116.

Liu,C., Johnson, L., Ostrow, D., Silvestre, A., Visscher, B., Jacobson, L. (2006).Predictors for Lower Quality of Life in the HAART Era Among HIV-Infected Men. J Acquir Immune DeficSyndr; **42**:470- 477.

Longford, N. (1993). Random Coefficient Models. Oxford, Clarendom Press.

Merso, F. (2008). Women & Girls and HIV/AIDS in Ethiopia.

Moattia, J., Carrieria, M., Spirea, B., Gastauta, J., Cassutod, J., Moreaue, J.(2000). Adherence to HAART in French HIV-infected injecting drug users: the contribution of buprenorphine drug maintenance treatment. AIDS; **14**:151-155.

Moha, R., Danela, C., Messoua, E., Ouassaa, T., Gabillarda, D., Anziana, A., Aboa, Y., Salamona, R., Bissagnenea, E., Seylera, C., Eholie, S. and Anglareta, X. (2007).Incidence and determinants of mortality and morbidity following early antiretroviral therapy initiation in HIV-infected adults in West Africa. AIDS, **21**:2483–2491.

National Factsheet (2010). Website http://www.etharc.org/resources/healthstat/nationalfactsheet /13-nationalfactsheet2010.

Nelder, J. and Wedderburn, R. (1972). Generalized Linear Models, Journal of the Royal Statistical Society, Series A, vol. 135, pp.370-384.

Nuredin, I. (2007). Evaluation of factors affecting chance of survival/death status among HIV positive people under the Anti- Retroviral Treatment Program: The Case of Adama Hospital.

Palella, F., Delaney, K., Moorman, A., Loveless, M., Fuhrer, J., Satten, G., Aschman, D., Holmberg, S. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection.*N Eng. J Med*.;**338**:853-860.

Palella, F., Deloria-Knoll, M., Chmiel, J., Moorman, A., Wood, K., Greenberg, A.and Holmberg, S.D. (2003).Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. Ann Intern Med.; **138**(8):620-626.

Pathipvanich, P., Ariyoshi, K., Rojanawiwat, A., Wongchoosie, S., Yingseree, P., Yoshiike, K., Warachit, P., Sawanpanyalert, P. (2003).Survival Benefit from Non-Highly Active Antiretroviral Therapy in a Resource-Constrained Setting. JAIDS Journal of Acquired Immune Deficiency Syndromes; **32**:157-160.

Raudenbush, S., Yang, M. and Yosef, M. (2000).Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation, Journal of Computational and Graphical Statistics 9, 141 – 157.

Recsky, M., Zabrina, L., Keith et al., (2004).Antiretroviral resistance among HIV-infected persons who have died in British Columbia, in the era of modern antiretroviral therapy.The Journal of Infectious Diseases; **190**:285-92.

Reda, A., Biadgilign, S., Deribew, A., Gebre, B., Deribe, K. (2013). Predictors of Change in CD4 Lymphocyte Count and Weight among HIV Infected Patients on Anti-Retroviral Treatment in Ethiopia: A Retrospective Longitudinal Study. *PLoS ONE;***8**(4):e58595. doi:10.1371/journal.pone.0058595.

Schwarz, G. (1978).Estimating the dimension of a Model Annals of Statistics 6, 461–464.

Snijders, T. and Bosker, R. (1999). Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling. London/ Thousand Oaks/ New Delhi: Sage Publications.

Thirumurthy, H., & Lester, R. T. (2012). M-health for health behavior change in resource-limited settings: applications to HIV care and beyond. *Bulletin of the World Health Organization*, *90*(5), 390-392.

UNAIDS (2008).Report on Global AIDS Epidemic-Google Book: Website:http://www.unaids.org/en/media/unaids/contentassets/dataimport/pub/global report/2008/jc1510.

UNAIDS (2010). Global report: UNAIDS report on the global AIDS epidemics. Geneva, Switzerland

UNAIDS (2011).World AIDS Day Report. Geneva, Switzerland.

UNAIDS (2012). Global Report: UNAIDS Report on the Global AIDS Epidemic 2012.

UNAIDS (2013).Global Report:UNAIDS Report on the Global AIDS Epidemic 2013.

UNAIDS/WHO (2012). *Core Epidemiology Slides*. Geneva, Switzerland.

Verheij, R. (1996). Explaining urban-rural variations in health: A review of interactions between individual and environment. Social Science and Medicine**42**: 923-935.

WHO (2003). Scaling-up antiretroviral therapy in resource-limited settings. Treatment guideline for a public health approach .2003 revision.

WHO protocols for CIS countries (2004). HIV/AIDS treatment and care clinical protocols for the WHO European Region. *DK-2100 Copenhagen, Denmark*. http://www.euro.who.int/pubrequest.

APPENDICES

Appendix A: Proportional sample size allocation to different woredas

$$n_j = \frac{n}{N} N_j$$

Where, n is sample size of the study

n_j is sample size of woreda j

N is population size

N_j is size of woreda j(the number of patients in woreda j). j=1, 2, 3, …, 9

Table A: Proportional sample size allocation to the different woredas

Woreda	n_j	N_j
Hossana town administration	71	151
Lemmo	51	109
Anlemmo	46	98
Misha	53	113
Angecha	35	74
Duna	34	72
Gombora	36	77
Lera	45	96
Soro	29	62

Appendix B: Output of multilevel logistic regression analysis

Table B1: Output of single-level empty logistic regression analysis

logit Mortality

Iteration 0: log likelihood = -193.03103

Iteration 1: log likelihood = -193.03103

Logistic regression Number of obs = 400

LR chi2(0) = 0.00

Prob > chi2 = .

Log likelihood = -193.03103 Pseudo R2 = 0.0000

Mortality	Coef.	Std.Err.	Z	P>|z|	[95% Conf.	Interval]
_cons	-1.466337	.1281025	-11.45	0.000	-1.717413	-1.215261

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	-193.031	193.031	1	408.0621	423.0535

Note: N=Obs used in calculating BIC; see [R] BIC note

Table B2: Output of random coefficient logistic regression model (random part result)

For Age

Log likelihood = -152.67283

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(Age)	.0793801	.1491413	.0019973	3.154785
var(_cons)	.0006489	.0133159	.000221	1.910124
cov(Age,_cons)	.0071768	.0690936	-.1282443	.1425978

LR test vs. logistic regression:chi2(3) =11.51 Prob >chi2 =0.0072

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-152.6728	13	351.2347	383.2347

Note:N=Obs used in calculating BIC; see [R] BIC note

For Level of education

Log likelihood = -153.2654

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(Led)	.0000228	.0026747	.0004629	1.78001
var(_cons)	.0492352	.1712031	.000054	2.88621
cov(Led,_cons)	-.0010593	.0636452	-.1258016	.123683

LR test vs. logistic regression:chi2(3) =14.32 Prob >chi2 =0.0203

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-153.2654	13	352.5308	384.4198

Note:N=Obs used in calculating BIC; see [R] BIC note

For Alcohol

Log likelihood = -151.75959

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(Alch)	.5239328	.6204579	.0514347	5.336978
var(_cons)	.0011904	.0144072	.0000594	2.3914107
cov(Alch,_cons)	.0249735	.1436481	-.2565716	.3065187

LR test vs. logistic regression:chi2(3) =15.33 Prob >chi2 =0.0060

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-151.7596	13	349.5192	381.4082

Note:N=Obs used in calculating BIC; see [R] BIC note

For Basline weight

Log likelihood = -153.0069

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(Baslinwt)	.0710496	.1942037	.0003349	4.07288
var(_cons)	.161622	.2495861	.0078348	3.334073
cov(Baslinwt,_cons)	-.1071596	.2172089	-.5328812	.3185619

LR test vs. logistic regression:chi2(3) =11.84 Prob >chi2 =0.0304

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-153.0069	13	352.0138	383.9028

Note:N=Obs used in calculating BIC; see [R] BIC note

For TB Status

Log likelihood = -152.95349

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(TBStas)	.3763487	.6208049	.014842	2.543056
var(_cons)	.1583554	.2804204	.0049238	3.092904
cov(TBStas,_cons)	-.2015864	.3724357	-.9315469	.5283742

LR test vs. logistic regression:chi2(3) =13.09 Prob >chi2 =0.0089

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-152.9535	13	351.907	383.796

Note:N=Obs used in calculating BIC; see [R] BIC note

For Basline CD4

Log likelihood = -153.2644

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
wrda: Unstructured				
var(BaslinCD4)	.0000129	.0015682	.0000004	5.029315
var(_cons)	.0451922	.1424722	.0000937	4.80469
cov(BaslinCD4,_cons)	.0007634	.045548	-.088509	.0900358

LR test vs. logistic regression:chi2(3) =14.92 Prob >chi2 =0.0105

Note: LR test is conservative and provided only for reference.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	400	.	-153.2644	13	352.5308	384.4198

Note:N=Obs used in calculating BIC; see [R] BIC note