UNIVERSITY OF NAIROBI

College Of Biological And Physical Sciences

LONGITUDINAL MODELING OF HIV PREVALENCE IN KENYA

BY

AIBAN RONOH

SUPERVISOR: DR. Thomas N. Achia

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL FULFILLMENT FOR A DEGREE OF MASTER OF SCIENCE IN BIOMETRY

August 2009

Declaration

I, hereby declare that this Project is my original work and has not been presented for a degree in any other University.

Aiban K. Ronoh Reg.No. I56/8701/2005

Sign: ..... Date: 02/9/2009

This project has been submitted for examination with my approval as the University Supervisor.

Dr Thomas N. Achia
School of Mathematics
University of Nairobi

Sign: ..... Date: 02/09/09

Acknowledgment

I am grateful to my supervisors, Dr Thomas Achia for his continuous guidance, advice and valuable comments throughout all the stages of this project. I also wish to especially thank Dr. Thomas Achia for his help in building the model. I also wish to extend my appreciation to all my lecturers and other members of staff of the School of Mathematics, University of Nairobi without whose support and encouragement my studies may not have been the same. To all my classmates, Hillary Kipruto, Festus Muriuki.Joan Thiga, Alex Mwaniki, George Emukule.

Special thanks goes to the Kuresoi Members of parliament, present and past and the Constituency office for their financial assistant.I also wish to thank posthumously, the late Hon Kipkalya Kones for his fatherly advice and monetary support he gave without which I could not have reached this far.

Special thanks to my wife Linah, daughter Melinda, family members, colleagues at work and friends who provided moral support, prayers and encouragement throughout the study period.My brothers and sisters for standing alongside my struggles. Last but not least, the Almighty God for protection, strength, guidance, wisdom, good health and courage to undertake the studies successfully.

Dedication

To my wife Linah and daughter Melinda for their support and endurance during my absence.

My dear mother Recho for her prayers and support during this period.

Contents

List of Tables

List of Figures

Abstract

HIV/AIDS have remain the leading global health challenge.Its dynamics and spread is the concern of all the sectors of the society.In research, many studies continue to be carried out to really try and understand the key determinants of its distribution, which areas and groups are most vulnerable. This is aimed at designing effective intervention measures and seeking cure and development of a vaccine. The goal of modeling is to extract much information from the available data in order to provide an accurate representation of knowledge and uncertainty of the epidemic.Many models have been put forward to understand the level of prevalence, which include a mathematical model called the back calculation, the WHO and UNAIDS have developed a computer program called the EPP and spectrum to provide projections and mortality due AIDS. The sentinel surveillance data from ANCs still remain the crucial source of prevalence data though they are reports suggesting that it normally overestimate the level of prevalence.Other modeling techniques can be developed to give short term projections of the prevalence level in various settings of the pandemic.Mixed effect models if well fitted can give a useful insight into the prevalence. This is where certain covariates are held as fixed while others are random for example, the rural and urban settings could be random while thinks like clinics are fixed and so on. Other time covariates should also be considered, for instance the incorporation of things like condom use, circumcision, coverage of ARVs and awareness campaigns. One of the key to modeling time varying data is the consideration of the correlation in the model.

Abbreviations

AIDS	:	Acquired Immune Deficiency Syndrome
ANC	:	Anti-Natal Clinic
EPP	:	Estimation and Projection Package
AIC	:	Akaike inclusion criterion
HIV	:	Human Immune Virus
NACC	:	National AIDS Control Council
NLME	:	Non linear mixed effect model
WHO	:	World Health Organization
DHS	:	Demographic Household Surveys
AIS	:	AIDS Indicator Surveys

Chapter 1

GENERAL INTRODUCTION

1.1 Background

HIV/AIDS remains major health and development concern globally. Millions of people have died as result of the epidemic and millions more are infected with the virus. This has elicited concerted efforts by government, UNAIDS, non-government organizations and world health organizations to monitor the prevalence and device control measures to control the spread and provide treatment of the AIDS pandemic.

In Kenya HIV prevalence has been monitored through annual sentinel surveillance in ante-natal clinics (ANC) since 1990 by the National Aids and STD Control Program (NASCOP). The data collected in ANCs is useful in estimation of national prevalence. According to Cheluget, Baltazar, Orege at al (Evidence of level declines in Adult HIV prevalence In Kenya) 2006, there has been a general decline from 10% in the late 1990s to under 7% today. Women attending ANCs are recruited for sentinel surveillance up to a period of three months each year from 1990 to 2004 are captured through testing following internationally recommended protocol using fourth generation ELISA Test.

It is now clear that HIV prevalence is now declining in Kenya. This is partly because there has been high incidence of deaths compared to reports of new infection according to the same report. The studies carried out have mainly been targeting to show that there has been a general decline in reported cases of new infection allegedly due to increase awareness,

behavior change, and availability of treatment and prevention of mother to child infection. Such studies should however be treated with caution since certain areas have not shown any change or prevalence have been on increase. Further analysis should also be carried to verify reports by groups with vested interests like government, NGOs, Pharmaceutical companies who may be out to drive their own agenda.

1.1.1 STATEMENT OF THE PROBLEM

Many studies describing the change in HIV prevalence have been carried out. However, modeling change including both long and short term projection of the future prevalence poses a major challenge. There is need to explore various modeling techniques in order to adequately describe the behavior of HIV/AIDS epidemic for purposes of planning intervention measures over time. Within subject characterize the individual region change over time while the between subject characterize the inter regions patterns and attempt to give the short term projection .

1.1.2 Rationale

There is need to develop of a statistical prediction tools that will be useful to the authorities and stake holders in planning intervention measures and provision of treatment for the HIV epidemic.It is also necessary to establish the critical HIV prevalence determinants so that targeted control measures are instituted.

1.1.3 OBJECTIVES OF THE STUDY

The main objective is to describe the change in pattern of HIV Prevalence in Kenya. The specific objectives are:-

- To establish the covariance structure of HIV prevalence data
- To fit a mixed effect model to the national sentinel HIV prevalence data
- To use mixed effect model to Predict the short term HIV prevalence
- Compare the results with EPP Model projections.

1.1.4 Literature Review

G.P Garnet (2002) described the importance of mathematical models in sexually transmitted diseases for purposes of planning and policy. Various types of models and their suitability are also explained including their functions. The global HIV model is also explored.
UNAIDS has generated a flexible model which provides a flexible framework for analysis of sero- prevalence data worldwide. The estimation of model parameters is done by maximum likelihood methods from the available prevalence data. The initial rate of growth of the epidemic depends on the transmission coefficient r, and the epidemic peak is depend on the initial fraction at risk f_o, and where it stabilizes depend on the strength of changes in the recruitment to the risk population I n response to AIDS death. The model has been used in Uganda and Benin where the reliability of the HIV prevalence curves depend upon availability and validity of data. The process of describing the spread of HIV epidemic depends largely on:-

1. assumptions made
2. Data available to estimate the parameter values

As a results the limitation of forecasting the future trend are brought about by-

1. Poor quality of available data
2. Uncertainty about the parameter values
3. Non-Linearity in the system
4. Chance events

This paper acknowledges the challenges of theoretically understanding the behavior of infections with the dynamics of sexual partnerships.
Solomon et al (2001) -Modeling HIV/Aids epidemics in sub Saharan Africa using sero-prevalence data from ante natal clinics, developed a maximum likelihood approach for the estimation of model parameters and used numerical simulation to obtain uncertainty intervals around the estimates. Traditional method for modeling HIV is called back calculation or back projection. This involves production of statistical solutions from a set of equations that relate to number of AIDS diagnoses over time to past trends in HIV infections, and

the distribution of incubation period. This method is mostly used in developed countries where the data on AIDS is complete unlike in developing countries. This method cannot be used in Kenya currently because of the scarcity of reliable of information on Incidence of AIDS. A modified framework was developed by WHO to reconstruct incidence curves and develop short-term projections based on HIV prevalence rather than AIDS notifications. This was in the form of software called the EPIMODEL. Epimodel uses an input estimate of point prevalence in reference year, combined with assumptions about HIV/AIDS progression rates and start year of epidemic. The assumption here is that the HIV prevalence follows a parametric curve over time based on gamma distribution. The shape of the curve and the position on the curve in the anchor year is required input. This model is deterministic in nature. This seeks to improve modeling of HIV in sub Saharan Africa and develop estimates of prevalence and mortality over time that includes ranges of uncertainty. There is also a back calculation method where AIDS cases as a result of infection with HIV infection is followed by an incubation period. This method does not require keeping track of any specific risk group or modes of transmission. The back calculation is based on the underlying relationship between the number of new AIDS cases at time t and t+dt which is denoted by a(t) and the number of new HIV infections h(s) at time s since the start of the epidemic at time $s = 0$. Let u be the time between the infection and diagnosis and f(u) be the density function of the incubation period u. Then

$$a(t) = \int_0^\infty h(t-u)f(u)\ du \tag{1.1}$$

So that if h and f are known then the distribution of the number of AIDS diagnosis in the period upto time t can be established.However information on h is mostly not available and this equation cannot be used directly. So $a(t)$ can be obtained by adjusting a parametric model to the new AIDS cases and this can be used to predict future AIDS cases. The analytical objective is to estimate a set of parameter values that are most likely to have produced the observed prevalence data. The assumption being that the prevalence data is normally distributed. The findings of the study were that the prevalence was generally high in urban areas compared to the rural areas. The study addressed the major criticism of the Epimodel that was based on the unmodified gamma distribution which gives a poor representation of the decline of the epidemic after its peak.

The key assumptions were that sentinel data from pregnant women can be extrapolated to the entire population. This was however dependent on whether:

- prevalence rate in antenatal clinic sites represent the general population rates in the areas among women of same age as those who attend the clinic;
- the prevalence rate in ANC represent the general adult population.
- prevalence rate at sentinel sites represent the national prevalence rate.

Montana et al (2008) seek to compare the prevalence from ANC and DHS and found out that ANC prevalence tend to overestimate the prevalence rate. However the estimates are similar if the study is restricted women and men or women only residing in the ANC catchment zone.Geographical information systems were used to map ANC surveillance sites to DHS/AIS survey clusters.National DHS/AIS prevalence estimates for men and women were then compared with the estimates from ANC surveillance.In all the occasions, national DHS/AIS were lower than the ANC estimates. According to Somi et all (2006) the Estimations and Projections Package(EPP) estimates and project the number of people living with HIV and AIDS cases using ANC data which can the be used to calculate the number of deaths as a result of AIDS. This software was developed by jointly by WHO and UNAIDS for countries with heterosexual epidemics of HIV infection. Again the assumption is that, the data from ANC surveillance can be used to represent the whole population. This assumption is based on the comparison of large number of studies of HIV prevalence among the pregnant women in community surveys. The EPP uses the available surveillance data to estimate the time trend of adult prevalence at national level. EPP estimates the time trend of HIV prevalence by fitting a simple epidemiological model to surveillance data provided by HIV sentinel surveillance. The modeling and projections has determined the model should be suitable with four parameters namely:-

- The starting year of the epidemic
- The force of infection in which a large value will cause high prevalence increase
- The fraction of the initial population at risk which determines the peak of the epidemic
- The behavior adjustments parameters which determines the proportion of the new entrants in the adult population at risk, if this parameter is negative,people reduce

their risk in response to the epidemic and the curve shows sharper decline after the peak, if is zero then the risk remains constant and when positive, the risk increase over time and prevalence falls less quickly or stabilizes at a high level.

In the study carried out in Tanzania for the data covering 1980 to 2010 for the whole country and rural and urban areas, the HIV prevalence in the urban areas increased from 0 percent in 1981 to a peak of 12.6 percent in 1992 and levels out to between 10.9 percent to 11.8 percent in 2003 to 2010 period. The rural curve showed a steep increase in HIV prevalence trend until 1995 when it peaked at 7 Percent in 1995 and then a gradual decline to reach 5.2 percent in 2004 and then stabilized between 5.1 and 5.3 percent from 2005 and 2010. The main limitation of EPP projections is that it cannot address the issue of non representativeness of the HIV prevalence data and it tries to minimize this by reducing HIV prevalence in the rural areas by a factor of 20 percent. The other weakness of EPP is that it only fits the curves and therefore becomes difficult to deal with certain issues of determinants of the prevalence like behavior change and also subjective to the understanding of the user of the epidemic.It cannot give certain parameter estimates associated with the model fit and therefore cannot give the low and high future scenarios based on parameter fits.

Chapter 2

Methodology

In this particular section, methods for fitting a suitable model for HIV longitudinal data are described and how they are applied. This is began by highlighting various exploratory data analysis that includes graphical techniques like box plot, trellis plots, data summary and covariance structure that tells us the nature and behavior of HIV prevalence data over time. Covariance structure is important in determining whether the data is correlated and therefore the model which is eventually fitted should account for the correlation.Non linear mixed effect model are popular in the predictive scenarios beyond the observed data. This is useful in making more reliable and non controversial short term projections of the HIV prevalence bearing in mind other important prevalence determinants.

2.1 Definition

Longitudinal data are observations taken from an experimental unit over time. There is therefore variation between units and within units.Mixed logistic regression with random effects can be used to study the response change over time and effects of the explanatory variables on the response.Longitudinal data may exhibit correlation between successive measurements.In the case of HIV prevalence data, the experimental unit is the clinics and also province level.Prevalence from the same site is likely to be similar compared to another site because of the socio economic and demographics dynamics which differ from one region to

another. For a valid statistical inferences, correlation should be included in the model.For HIV prevalence data, exploratory analysis shows that prevalence are normally distributed and correlated. There are two approaches for modeling this Longitudinal data:-

1. Dropping the assumptions of independence between the response Y_i and modeling correlation structure explicitly. This method is similar in conclusions with generalized linear models for independent outcomes.

2. The second one is multilevel modeling

The main issues that arise from modeling longitudinal data are:-

(a) Methods for exploratory data analysis

(b) Risk of using inappropriate model

(c) Missing data

The preferable exploratory analysis is data summary which consist of summary by a small number of descriptive statistics based on the assumptions of independent.

2.1.1 Longitudinal Model Goals

The goals for Longitudinal data analysis are:

- To characterize patterns for example subject responses over time
- To investigate the effects of the important covariates on these patterns

2.1.2 Types of covariates

The main types of covariates in longitudinal modeling are:

1. Non time varying covariates for example gender between subjects.

2. Time varying covariates for instance the percentage coverage of anti retro viral treatment over time.

2.1.3 Advantages of Longitudinal studies over Cross Sectional studies

The advantages of longitudinal studies compared to cross-sectional studies are:-

1. Can separate the cohort and time effect in population studies. The time effect change over time within units and the cohort effects is the difference between unit.

2. Cross sectional studies only give a single response for each unit.

Suppose N study units with n_i measurement for the i-th subject. That is, longitudinal observations for units i in n. Let y_i denote the vector of responses for subject i and y denote the vector of responses for all subjects.

$$\mathbf{Y_i} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{2.1}$$

assuming a normal linear model for y then

$$E(y) = X\beta = \mu \tag{2.2}$$

where y is normally distributed with variance V where

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{2.3}$$

and

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} \tag{2.4}$$

Xi is a n x p design matrix for unit i and β is a parameter vector of length P. The variance -covariance matrix for the subject i will be

$$\mathbf{V}_i = \begin{bmatrix} \sigma_{i11} & \sigma_{i12} & \sigma_{i13} & \dots & \sigma_{i1ni} \\ \sigma_{i21} & \sigma_{i22} & \sigma_{i23} & \dots & \sigma_{i2ni} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{ini1} & \sigma_{ini2} & \sigma_{ini3} & \dots & \sigma_{inini} \end{bmatrix} \tag{2.5}$$

and the overall variance covariance matrix has block diagonal form

$$\mathbf{V} = \begin{bmatrix} V_1 & O & O & \dots & O \\ O & V_2 & O & \dots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & \dots & V_N \end{bmatrix} \tag{2.6}$$

Assuming that the response are independent for different experimental units and **O** is a matrix of zeros and **V** are assumed to have the same format for all the units. if **V** elements are known constants then beta can be estimated either method of least squares or the maximum likelihood estimator obtained by solving the score function

There are commonly used forms of matrix **V**i

1. All of diagonal elements are equal that is

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & \rho & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} \tag{2.7}$$

   This is a case of equal correlation for example samples from the same experimental unit, and here rho is the intra class correlation coefficient and if it can be expressed as

$$\frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}$$

   the matrix is of **compound symmetry**

2. The off diagonal terms are exchangeable is called the equi-correlation or spherical matrix where p depends on the distance between observation,j and k for measurements

at time **tj** and **tk**

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho_{1_2} & \rho_{1_3} & \dots & \rho_{1_n} \\ \rho_{2_1} & 1 & \dots & \dots & \rho_{2_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n_1} & \rho_{n_2} & \dots & \dots & 1 \end{bmatrix} \tag{2.8}$$

This kind of correlation is used in the first order autoregressive models such that

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \dots & \rho^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \dots & \dots & \dots & 1 \end{bmatrix} \tag{2.9}$$

3. All correlation terms are different and this is called **unstructured correlation matrix** involves no assumptions between measurements but all the vectors y_i must be the same length n

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho_{1_2} & \rho^2 & \dots & \rho_1 n \\ \rho_{2_1} & 1 & \dots & \dots & \rho_{2_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n_1} & \dots & \dots & \dots & 1 \end{bmatrix} \tag{2.10}$$

2.1.4 Estimation and Projection Package 2007

This is a WHO and UNAIDS software that was built to estimate and project time trend of adult prevalence using the available surveillance data in this case the sentinel data from ANCs countrywide. The HIV prevalence time trend was fitted using data from twenty five sentinel sites that were categorized either rural or urban and then the results used to give a national projection. The parameters of interest in this method are:

- The rate of growth of the epidemic r, which is unique to every country, if this rate increase then the epidemic grows rapidly and vice versa.where $0 = r = 30$
- Fraction of new entrants to the risk category f_0 . This is important in determining where the epidemic levels off.Larger values lead to epidemic with higher stable HIV prevalence levels.where $0 = f_0 = 1$

- The start time of the epidemic t_0,is the point at which HIV is introduce to the population.In Kenya The first case of HIV was reported in 1985.

- The behavior change parameters ϕ, which is depended on the intervention measures carried out. Positive values of ϕ means more people are brought into the risk population and the prevalence becomes high.Negative values reduces the prevalence. This value range from -10,000 to 10,000.

EPP explores the full range of curves allowed by the following procedure for fitting the model:

- Large number of randomly generated values of r, f_0,t_0, and ϕ.

- Measure of fit is calculated and the curves weighted according to the measure of fit. Those with best fit receive high weight values and those that miss the data are given very low weights.

- Resampling all the curves according to the weight based on the measure of fit. The curves that fit the data are selected many times and those that do not fit the data are not re sampled at all.

- The re sampled curves are the used to estimate the uncertainty.

The EPP deterministic model uses a Bayesian melding approach. This is a combination of inputs and outputs of the model to a generalized epidemic. Bayesian is the inference which starts by quantifying prior knowledge about the values of the prevalence. After specifying the prior distribution, they are used to generate a set of possible epidemic curves based on EPP and satisfy the constrains imposed on the epidemic prevalence. Data and information measurement errors are used to calculate a measure of likelihood, an epidemic curve that is similar to the levels and trends in observed prevalence has a high likelihood of representing the true prevalence. Combining prior distributions with likelihood gives a posterior distribution of the prevalence.In this method of Bayesian melding, the sample of country specific epidemic curves describing HIV prevalence over time is derived based on the time series of ANC prevalence data and the general parameters that describes the epidemic. The prevalence trends at ANC are calibrated to population based on the estimates from

the national surveys.For countries with no national population based estimates, a general calibration method is developed.

2.1.5 Non Linear Mixed Effects Model

The main characteristic of the longitudinal data is the repeated measurement from the same experimental unit over time. This calls for a special statistical method because of the inter correlation of the set of observations from the same unit. This correlation must be taken into account in order to arrive at valid scientific inference. Non Linear mixed effect model provides an effective approach to longitudinal data analysis.Non linear mixed effect models are mixed effect models in which some or all the fixed and random effects occur non linearly in the model function.Non linear models are generally mechanistic, that is a model for mechanism that produces the responses. This kind of model normally give rise to parsimony since they use less parameters than the linear mixed effect models.Non linear model also give a more reliable prediction of the response outside the observed data compared to polynomial models.NLME require starting fixed estimates of co-efficients which is often done by intuition although Bates and Watts(1988) re commended some general guidelines.

Some of the assumptions of NLME model are:

1. Correct mean function.
2. homogeneity of variance
3. Normally distributed measurement error

2.1.6 Single Level NLME Model

Let the j-th observation in the i-th group be modeled as

$$y_{ij} = f_i(\phi_{ij}, \upsilon_{ij}) + \epsilon_{ij} \ , \ i = 1, \ldots, M, \ j = 1, \ldots, N \tag{2.11}$$

where $\phi_i = \mathbf{A}_i\beta + \beta_i\mathbf{b}_i$. $\mathbf{y_i}' = \left[\ y_{i1} \ldots y_{in_i} \ \right]$, and $\phi_\mathbf{i}' = \left[\ \phi_{i1} \ldots \phi_{in_i} \ \right]$.

We also have that

$$\mathbf{f_i}(\phi_\mathbf{i}, \upsilon_\mathbf{i})' = \left[\ f(\phi_{i1}, \upsilon_i) \ldots f(\phi_{in_i}, \upsilon_{in_i}) \ \right]$$

$v_i' = \begin{bmatrix} v_{i1} \dots v_{in_i} \end{bmatrix}$; $\mathbf{A_i}' = \begin{bmatrix} A_{i1} \dots A_{in_i} \end{bmatrix}$; $\mathbf{B_i}' = \begin{bmatrix} B_{i1} \dots B_{in_i} \end{bmatrix}$

- **M** is the number of groups.
- n_i is the number if observations on the ith group.
- f is the general real values differentiable function of group specific parameter $\phi_i j$ and the covariate vector v_{ij} and ϵ_{ij}
- $\mathbf{b}_i$ are independent random effects for the i-th subject

The function f is at least non linear in one component in the group-specific vector ϵ_{ij}, which is modeled as

$$\phi_{ij} = \mathbf{A}_{ij}\beta + \beta_i \mathbf{b}_i \tag{2.12}$$

where β is P dimensional vector of fixed effects and $\mathbf{b_i}$ is a q-dimensional random effect vector associated with i-th group but not varying with j, with variance covariance matrix ψ. A_{ij} and B_{ij} are depended on the group and possible values of some covariates at the j-th observation. This generalization of the model can the allow the incorporation of the time varying covariates in the fixed effect or the random effect of the model. The main assumptions of the NLME are:

- Observations between groups are independent
- within group errors are $N(0, \sigma^2)$ and independent of $\mathbf{b}_i$

For computational purposes the representation of group-specific co-efficients could be chosen so that A_{ij} and B_{ij} are always simple incidence matrices.

2.1.7 Multi Level Non-linear mixed effects models

This is an extension of single level NLME to the data that is grouped according to the nested factors . For example the multi level version of Lindstrom and Bates(1990) model for two levels of nesting is written as two stage model in which the first stage expresses the response y_{ijk} for the k-th observation on the j-th second level of the i-th first group.

$$\mathbf{y}_{ijk} = f_i(\phi_{ijk}, v_{ijk}) + \epsilon_{ijk} \qquad where, \qquad i = 1, \dots, M, \quad j = 1, \dots, N \tag{2.13}$$

$$\mathbf{y}'_{ij} = \left[\begin{array}{c} y_{ij1} \dots y_{ijnij} \end{array} \right] \tag{2.14}$$

and

$$\phi'_{ij} = \left[\begin{array}{c} \phi_{ij1} \dots \phi_{ijnij} \end{array} \right] \tag{2.15}$$

$$\mathbf{f}_{ij}(\phi_{ij}, \upsilon_{ij}) = \left[\begin{array}{c} f(\phi_{ij1}, \upsilon_{ij1}) \dots f(\phi_{ijnij}, \upsilon_{ijnij}) \end{array} \right] \tag{2.16}$$

and

$$\upsilon'_{ij} = \left[\begin{array}{c} \upsilon_{ij1} \dots \upsilon_{ijnij} \end{array} \right] \tag{2.17}$$

$$\mathbf{A}'_{ij} = \left[\begin{array}{c} A_{ij1} \dots A_{ijnij} \end{array} \right] \tag{2.18}$$

$$\mathbf{B}'_{ij} = \left[\begin{array}{c} B_{ij1} \dots B_{ijnij} \end{array} \right] \tag{2.19}$$

where M is the number of first level groups , M_i is the number of second level groups within the i-th first level group,$n_{ij}$is the number of observations on the j-th second level group in the i-th first level group while ϵ_{ijk} is normally distributed error within group. The second stage of the model is expressed as

$$\phi_i jk = \mathbf{A}, jk\beta + \mathbf{B}_i, jk\mathbf{b}_i + \mathbf{B}_i jk\mathbf{b}_i j \tag{2.20}$$

where$\mathbf{b}_i \sim \mathbf{N}(0, \Psi_1)$ and $\mathbf{b}_i j \sim \mathbf{N}(0, \Psi_2)$ β is a p-dimensional vector of fixed effects with design matrix A_{ijk} which may incorporate time varying covariates, $\mathbf{b_i}$ is in dependently distributed q_1- dimensional vectors with variance covarianceψ_i. $\mathbf{b_i j}$ is the second level random effect q_2 dimensional independently distributed matrix ψ_2 assuming a first level random effect.$\mathbf{B}_i, jk$ and $\mathbf{B}_i jk$ are random effects depended on the first and second level groups and possibly on the values of the covariates at k-th observation.$\epsilon_i jk \sim \mathbf{N}(0, \sigma^2)$.

2.1.8 Brain Cousens Model

The model function for the Brain-Cousens model (Brain and Cousens, 1989) is

$$f(x, b, c, d, e, f) = c + \frac{d - c + fx}{1 + exp(b(log(x) - log(e)))}$$

and it is a five-parameter model. obtained by extending the four-parameter log-logistic model (LL.4 to take into account inverse u-shaped hormesis effects.

The parameters have the following interpretations

b: Not direct interpretation

c: Lower horizontal asymptote

d: Upper horizontal asymptote

e: Not direct interpretation

f: Size of the hormesis effect: the larger the value the larger is the hormesis effect. f=0 corresponds to no hormesis effect and the resulting model is the four-parameter log-logistic model. This parameter should be positive in order for the model to make sense.

Fixing the lower limit at 0 yields the four-parameter model

$$f(x) = \frac{d + fx}{1 + exp(b(log(x) - log(e)))}$$

used by van Ewijk and Hoekstra (1993). This models fits a fixed effect non linear model for the HIV prevalence data very well compared to all other NLME models.

Chapter 3

Results and Findings

3.1 Data Description

The prevalence data is obtained from the sentinel reports that are done annually by the government of Kenya through NASCOP and also from analytical reports from Kenya National Bureau of Statistics.Other sources includes KDHS data, UNAIDS reports on AIDS in Kenya and reports from NACC and National Census analytical reports. The data is described as below:

- prevalence refers to percentage of HIV infection per clinic as per ANC sentinel surveillance data
- Mixed site refers to the clinics that are located in areas that are semi urban and semi rural
- Urban sites are clinics that are located in towns.
- Rural sites are those located in the villages
- The start year of interest is 1990 and the end year is 2004

3.1.1 Exploratory Data Analysis

In this section we carry out exploratory analysis of the response variable, prevalence rates and use various predictions to explain their variable. Graphical methods were used basically to visualize the nature, trend and pattern of the prevalence data. In this study we used these methods to

- similarities in patterns are detected.
- Year to year variation can be appreciated.
- Straight line response overtime is suspect trend. Curves then enables us to choose the appropriate method of the analysis in so far as the choice of the model is concerned.
- Data summaries were also done to establish how the means by year, province and sites which in this case are the clinics, with a view of establishing how they vary across the group.

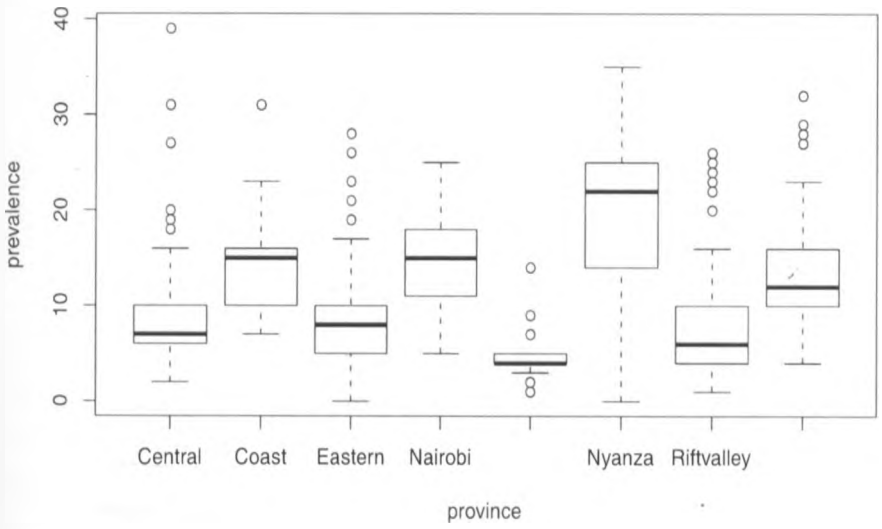


Figure 3.1: Box plot of HIV prevalence rate by province

Clearly from the box plot above, the prevalence of HIV in each province is not normally distributed with outliers. The length of the box plot differ for each province to show that the spread of the prevalence data is not constant.

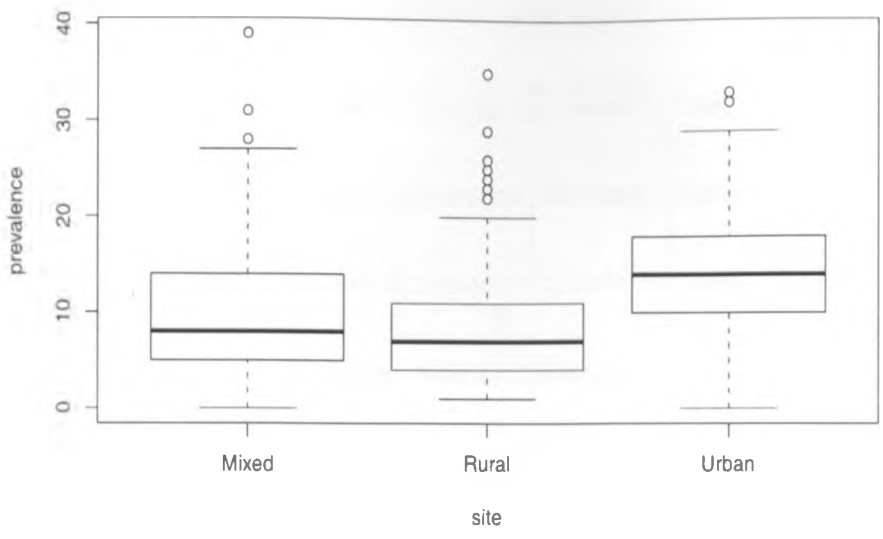


Figure 3.2: Box plot of HIV prevalence rate by site

The prevalence data per site, that is rural, mixed and urban are also not normally distributed with outliers. The spread is constant as shown by equal sizes of the box plots.

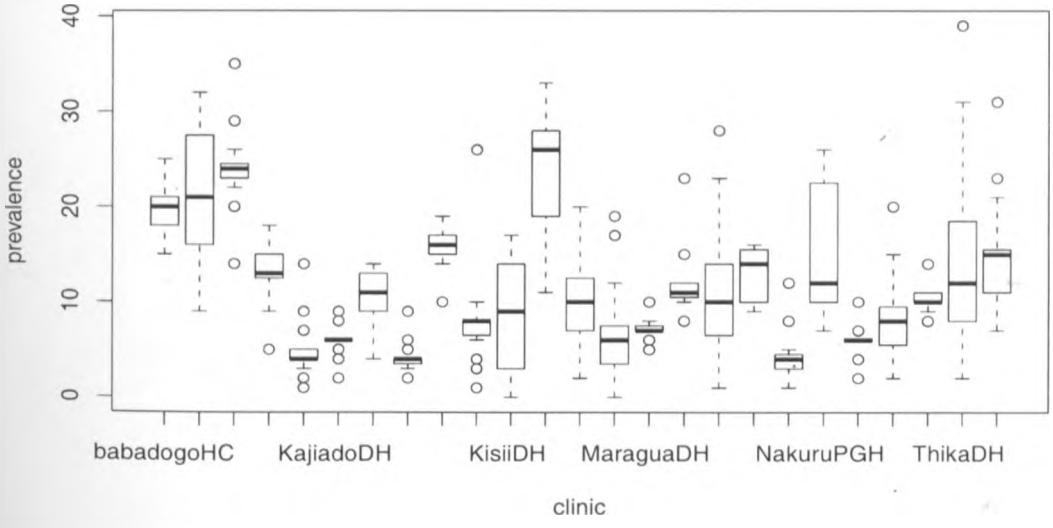


Figure 3.3: Box plot per clinic

Here the spread is not constant across each of the clinics and there are outliers. The location is also not constant. The figures below are trellis plots that helps to examine the trend of the prevalence rates for various clinics and sites. The trend seams to be a general increase

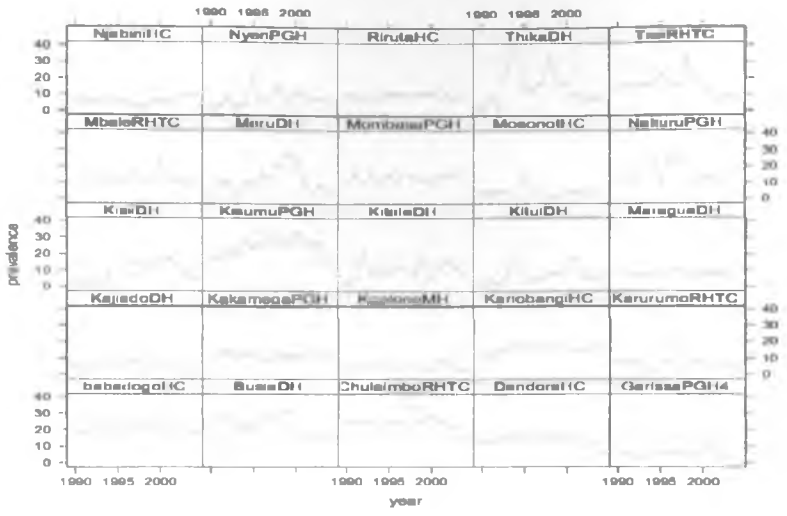


Figure 3.4: Trellis plot of HIV prevalence rate against year by per clinic

in the beginning, levels off and the a steady decline in all the sites and clinics.

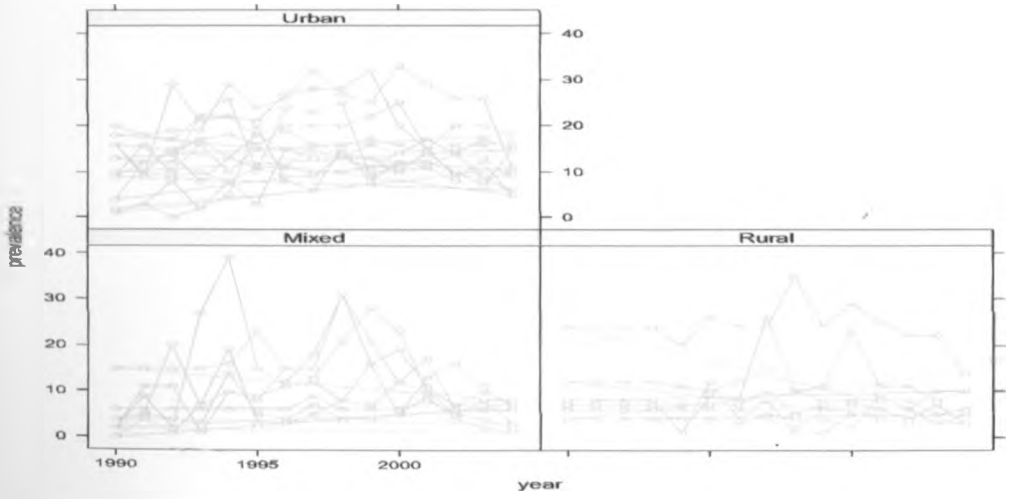


Figure 3.5: Trellis plot of HIV prevalence rate against year by site

3.2 Summary Statistics

In this section the prevalence data was summarized using mean and standard deviation for each year, province and clinics.This was particularly important in trying to identify the existence of outliers. Nyanza province had the highest mean prevalence while North Eastern province showed the lowest prevalence rates.

Table 3.1: Descriptive statistics per province

Province	N	Mean	SD
Nyanza	45	18.98	9.08
Nairobi	60	14.75	4.23
Western	45	14.60	6.44
Coast	30	14.07	4.84
Central	60	9.08	6.74
Eastern	45	8.78	6.45
Rift Valley	75	7.92	5.80
North Eastern	15	5.00	3.12

Table 3.2: Descriptive statistics per Year

Year	N	Mean	SD
1990	25	8.56	6.60
1991	25	9.72	5.41
1992	25	10.84	7.26
1993	25	11.48	7.85
1994	25	13.52	9.18
1995	25	12.40	6.84
1996	25	12.60	7.30
1998	25	14.84	9.91
1999	25	13.32	8.16
2000	25	13.64	8.06
2001	25	12.56	5.54
2002	25	10.64	6.06
2003	25	10.36	6.16
2004	25	8.24	4.65

As far as the years prevalence rates are concern then, 1998 recorded the highest prevalence rates while 2004 was the lowest. The following table is a table of means and standard deviation in each clinic. Kisumu Provincial hospital was found to record the highest prevalence mean rates while Kaplong and Mosoriot recorded lowest rates.

Table 3.3: Descriptive Statistics per Site

site	n	mean	SD
KisumuPGH	15	24.13	6.02
ChulaimboRHTC	15	24.07	4.43
BusiaDH	15	21.07	6.55
babadogoHC	15	19.80	2.43
KariobangiHC	15	15.73	2.09
NakuruPGH	15	15.20	6.74
TiwiRHTC	15	15.13	6.23
ThikaDH	15	14.93	10.57
DandoraHC	15	13.27	3.17
MombasaPGH	15	13.00	2.70
MbaleRHTC	15	12.00	3.40
MeruDH	15	11.27	7.81
KakamegaPGH	15	10.73	2.76
RirutaHC	15	10.20	1.42
KitaleDH	15	10.13	4.81
KisiiDH	15	8.73	5.86
KarurumoRHTC	15	8.13	5.50
NyeriPGH	15	8.13	4.87
MaraguaDH	15	7.07	1.22
KituiDH	15	6.93	5.36
NjabiniHC	15	6.20	1.93
KajiadoDH	15	5.87	1.55
GarissaPGH4	15	5.00	3.12
KaplongMH	15	4.20	1.61
MosoriotHC	15	4.20	2.81

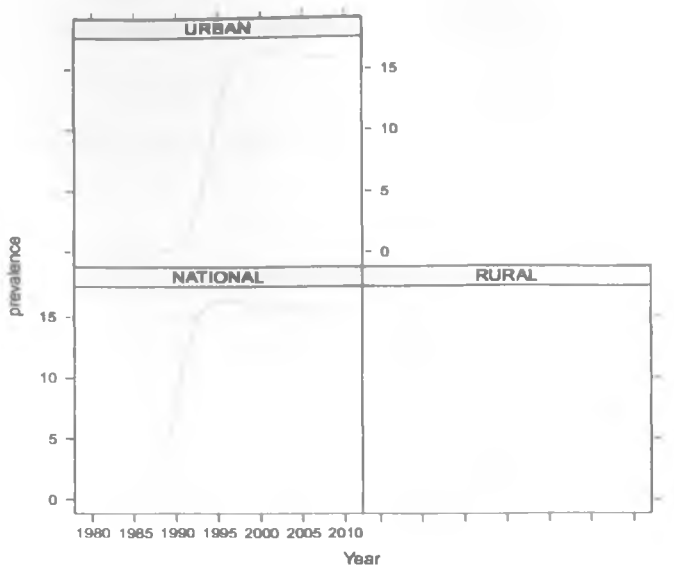
3.2.1 Covariance Structure of the data

Measurements that are repeated on the individual site and clinic are likely to be correlated. Therefore if the correlation is ignored it can impact negatively on the parameter estimation, test of hypothesis and the study design. The covariance matrix plays a key role in the analysis of the Longitudinal data. The prevalence in the years following each other are strongly correlated but the correlation becomes less as the time increase from each measurement. This strongly suggest an **autoregressive** lag one correlation structure and therefore the correlation will be important in the model. Table 3.4 gives the correlation matrix.

Table 3.4: Correlation of HIV prevalence rates

year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1990	1	0.822	0.741	0.706	0.362	0.753	0.730	0.654	0.559	0.596	0.635	0.542	0.652	0.752	0.69
1991	0.822	1	0.641	0.719	0.479	0.708	0.683	0.631	0.694	0.632	0.716	0.594	0.667	0.772	0.59
1992	0.741	0.641	1	0.551	0.326	0.715	0.763	0.624	0.577	0.773	0.535	0.493	0.698	0.734	0.64
1993	0.706	0.719	0.551	1	0.806	0.678	0.728	0.702	0.753	0.562	0.596	0.500	0.559	0.637	0.74
1994	0.362	0.479	0.326	0.806	1	0.456	0.623	0.548	0.748	0.540	0.623	0.553	0.465	0.491	0.54
1995	0.753	0.708	0.715	0.678	0.456	1	0.683	0.595	0.750	0.667	0.630	0.531	0.589	0.739	0.54
1996	0.730	0.683	0.763	0.728	0.623	0.683	1	0.840	0.810	0.817	0.772	0.675	0.744	0.768	0.75
1997	0.654	0.631	0.624	0.702	0.548	0.595	0.840	1	0.718	0.673	0.684	0.564	0.630	0.668	0.54
1998	0.559	0.694	0.577	0.753	0.748	0.750	0.810	0.718	1	0.790	0.735	0.576	0.528	0.619	0.55
1999	0.596	0.632	0.773	0.562	0.540	0.667	0.817	0.673	0.790	1	0.774	0.587	0.620	0.729	0.52
2000	0.635	0.716	0.535	0.596	0.623	0.630	0.772	0.684	0.735	0.774	1	0.738	0.684	0.752	0.56
2001	0.542	0.594	0.493	0.500	0.553	0.531	0.675	0.564	0.576	0.587	0.738	1	0.840	0.787	0.47
2002	0.652	0.667	0.698	0.559	0.465	0.589	0.744	0.630	0.528	0.620	0.684	0.840	1	0.912	0.70
2003	0.752	0.772	0.734	0.637	0.491	0.739	0.768	0.668	0.619	0.729	0.752	0.787	0.912	1	0.71
2004	0.687	0.592	0.636	0.737	0.545	0.543	0.748	0.545	0.540	0.517	0.564	0.473	0.704	0.715	1

Figure 3.6: EPP graphs for prevalence by site



3.3 Application of the EPP method on ANC data

This figure shows the curves fitted using EPP using ANC data from Kenya. The epidemic was defined into two main sub-epidemics which are the rural and urban and a general curve combining the two to give Kenya's national prevalence trends.In this particular case, the national prevalence trends appears to be similar to the urban trend compared to the rural trends in magnitude but the patterns are almost the same.

3.3.1 EPP predicted Values and Estimation Parameters

The table below were the short term projections of HIV prevalence rates for five years using EPP projection model. The figures for urban and National are almost the same, while the rural prevalence rates were lower. However the trends were similar for both sites over time.

Table 3.5: **EPP projections**

Year	National	Urban	Rural
2004	15.95	15.95	7.10
2005	15.94	15.94	7.06
2006	15.95	15.95	7.07
2007	15.96	15.96	7.06
2008	15.96	15.96	7.04
2009	15.98	15.98	7.04
2010	15.98	15.98	7.05

Table 3.6 gives the estimate prameters for Kenya using the EPP. The rate of growth for the epidemic seemed generally lower. The prevalence growth in the rural areas, has high rate given the r value of 9.44 compared to 5.54 for urban sites. In contrast, the fraction at risk in urban sites was higher compared to the rural sites. The ϕ values are positive suggesting that the prevalence could still rise. The base year t_o used was 1980, the year that the first case of the epidemic was reported.

Table 3.6: **EPP Parameters**

Fitting parameters	Urban	Rural
r	5.54	9.44
f0	0.18	0.08
t0	1980	1980
phi	7311.82	71678.89
lnL	1506.37	1092.09

3.3.2 Brain Cousens curves per Site

The trend curves shown here were fitted using the Brain Cousens model with four parameters. The trends for the rural sites were generally good giving the expected sigmoidal fits which are expected to explain the behavior of the prevalence. The curves clearly shows an initial increase over time, then a peak and the a decline.For Njabini health center, there seems to be a steep decline in the prevalence rates. Similar trends were noted for the urban

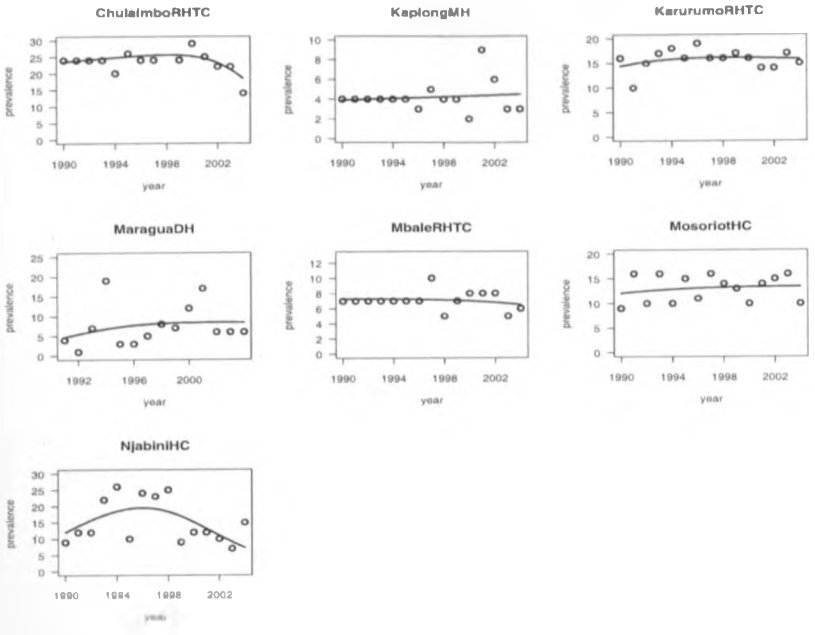


Figure 3.7: Brain Cousens Curve for Rural

and mixed sites with exceptions of a few cases where a straight line was experience which did not give a good trend of the prevalence. For example, Nakuru provincial hospital and babadogo gives straight lines for the urban sites and Kitui for mixed sites.

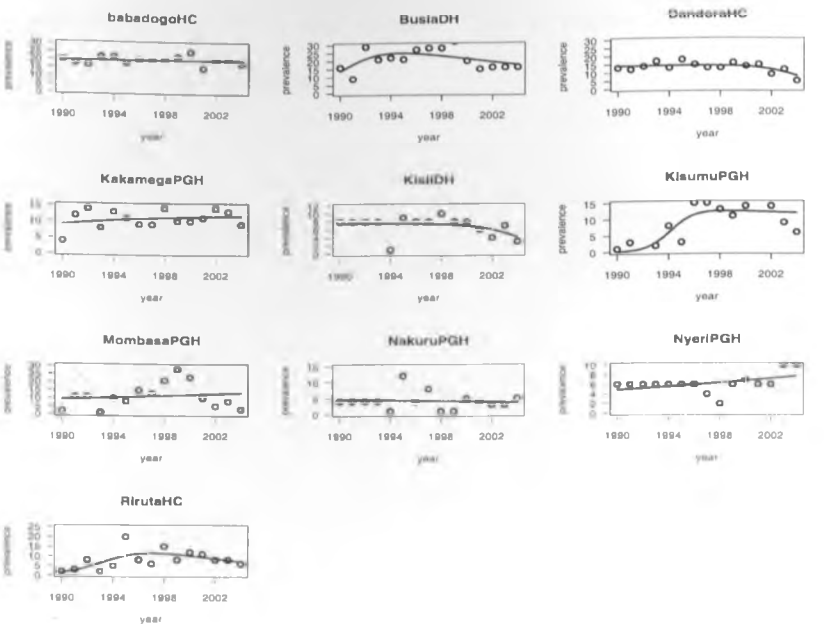


Figure 3.8: Brain Cousens Curve for Urban site

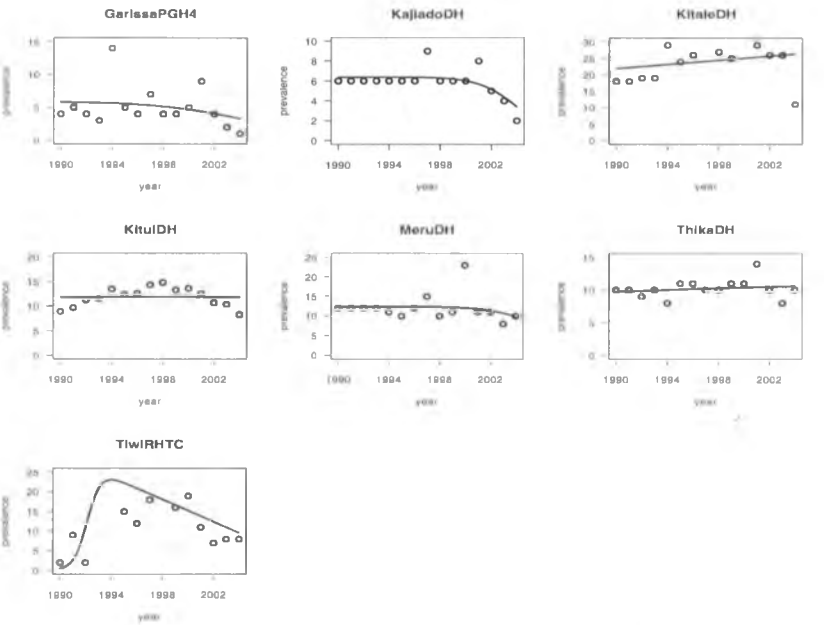


Figure 3.9: Brain Cousens Curve for Mixed

3.3.3 Mixed effect Model

The mixed effect model was fitted for HIV prevalence as a response variable and the clinics as explanatory variables. The main aim was to check if clinics as a unit can explain the trend of prevalence. The tables below shows the parameter estimates for each clinic for

the three sites, that is, rural, urban and mixed. The p values indicated that most of the parameter estimates were not significant.

Table 3.7: Rural

Site			Estimate	Std Error	t-value	p-value
Rural	ChulaimboRHTC	b	1031.93	481.23	2.14	0.0552
		d	-646.36	802.85	-0.81	0.4378
		e	2005.22	1.43	1399.16	0.0000
		f	0.34	0.40	0.84	0.4208
	KaplongMH	b	-30.77	356.04	-0.09	0.9327
		d	10.64	2256.55	0.00	0.9963
		e	2037.10	585.60	3.48	0.0052
		f	0.00	1.16	0.00	0.9996
	KarurumoRHTC	b	1187.62	1013.46	1.17	0.2660
		d	32.36	528.93	0.06	0.9523
		e	2004.38	2.44	821.08	0.0000
		f	-0.01	0.27	-0.05	0.9634
	MaraguaDH	b	455.73	734.21	0.62	0.5474
		d	-10.24	301.34	-0.03	0.9735
		e	2012.44	16.35	123.12	0.0000
		f	0.01	0.15	0.06	0.9545
	MbaleRHTC	b	988.77	867.33	1.14	0.2785
		d	-38.09	642.99	-0.06	0.9538
		e	2006.50	3.61	556.25	0.0000
		f	0.03	0.32	0.08	0.9388
	MosoriotHC	b	192.77	1981.73	0.10	0.9243
		d	-15.20	3121.03	0.00	0.9962
		e	2013.48	254.66	7.91	0.0000
		f	0.01	1.57	0.01	0.9949
	NjabiniHC	b	-68.66	17.02	-4.03	0.0020
		d	4.33	-	-	-
		e	2075.70	25.79	80.49	0.0000
		f	0.04	0.34	0.13	0.8976

Table 3.8: Urban

Site			Estimate	Std Error	t-value	p-value
Urban	babadogoHC	b	38.52	519.43	0.07	0.9422
		d	-150.79	2506.52	-0.06	0.9531
		e	2081.29	1215.02	1.71	0.1147
		f	0.09	1.26	0.07	0.9461
	BusiaDH	b	-1260.90	1141.32	-1.10	0.2928
		d	1931.45	2080.44	0.93	0.3731
		e	1990.33	1.73	1148.96	0.0000
		f	-0.96	1.04	-0.92	0.3781
	DandoraHC	b	1321.84	853.99	1.55	0.1499
		d	-50.56	563.39	-0.09	0.9301
		e	2004.30	1.29	1548.09	0.0000
		f	0.03	0.28	0.12	0.9104
	KakamegaPGH	b	-370.77	840.57	-0.44	0.6677
		d	-11.81	827.86	-0.01	0.9889
		e	1983.40	17.23	115.13	0.0000
		f	0.01	0.41	0.03	0.9777
	KariobangiHC	b	-529.93	410.72	-1.29	0.2234
		d	208.86	510.13	0.41	0.6901
		e	1983.94	5.47	362.51	0.0000
		f	-0.10	0.25	-0.38	0.7128
	KisiiDH	b	-2131.11	1447.31	-1.47	0.1717
		d	271.12	1435.53	0.19	0.8540
		e	1994.19	1.41	1415.20	0.0000
		f	-0.13	0.72	-0.18	0.8605
	KisumuPGH	b	-27.93	2595.99	-0.01	0.9916
		d	-611.54	788.18	-0.78	0.4542
		e	1463.67	43200.53	0.03	0.9736
		f	0.32	0.39	0.81	0.4370
	MombasaPGH	b	-344.31	1129.11	-0.30	0.7661
		d	66.12	886.11	0.07	0.9419
		e	1979.30	35.55	55.67	0.0000
		f	-0.03	0.44	-0.06	0.9538
	NakuruPGH	b	665.75	351.30	1.90	0.0846
		d	-4621.70	7693.63	-0.60	0.5602
		e	1999.06	6.07	329.42	0.0000
		f	2.33	3.86	0.60	0.5590
	NyeriPGH	b	-1313.46	-	-	-
		d	1645.24	-	-	-
		e	1993.64	-	-	-
		f	-0.82	-	-	-
	RirutaHC	b	-126.48	705.04	-0.18	0.8609
		d	-12.40	685.75	-0.02	0.9859
		e	1958.45	180.09	10.88	0.0000
		f	0.01	0.34	0.03	0.9732

Table 3.9: site as random

Parameter	Estimate	Std Error	Df	t-value	p-value	AIC
Intercept	-447653.9	83693.96	370	-5.348700	0	2517.723
year	448.3	83.82	370	5.348443	0	
$I(year^2)$	-0.1	0.02	370	-5.348047	0	

Table 3.10: Clinic as random effect

Parameter	Estimate	Std Error	Df	t-value	p-value	AIC
Intercept	-447653.4	54525.42	348	-8.209994	0	2273.542
year	448.3	54.61	348	8.209610	0	
$I(year^2)$	-0.1	0.01	348	-8.209001	0	

Table 3.11: EPP and Model Predicted values

	EPP			MODEL		
Year	National	Urban	Rural	National	Urban	Rural
1990	8.2	8.2	1.4	10.5	9.8	11.2
1991	11.7	11.7	2.4	9.9	9.5	10.2
1992	14.2	14.2	3.7	10.4	10.1	10.7
1993	15.6	15.6	5	10.9	10.5	11.2
1994	16.2	16.2	6	11.2	10.9	11.5
1995	16.4	16.4	6.7	11.5	11.2	11.8
1996	16.4	16.4	7.1	11.6	11.3	11.9
1997	16.4	16.4	7.2	11.7	11.3	11.9
1998	16.3	16.3	7.3	11.5	11.2	11.7
1999	16.2	16.2	7.3	11.2	10.9	11.3
2000	16.1	16.1	7.2	10.8	10.5	10.9
2001	16.1	16.1	7.2	10.2	10.1	10.3
2002	16	16	7.2	9.7	9.6	9.8
2003	16	16	7.1	9.1	9	9.1
2004	16	16	7.1	8.2	7.9	8.4

Chapter 4

Discussion and Recommendation

The initial exploratory analysis indicated that the HIV prevalence data is not normally distributed but rather skewed to the right. This is according to box plots which was used as exploratory tools. There has been an increase in the prevalence since the pandemic was reported and stabilized in the late 1990 and seams to be declining. This study could not cover the explanatory variables that could explain the decline.It is therefore recommended that future models should incorporate the time varying covariates such as condom use, availability of ART and other intervention measures that have been undertaken to curb the spread of the disease.

The mixed effect model with year as a covariate is important in the explanation for the HIV prevalence over time.However the model with clinics within site as random effect gives a better fit as it has the least AIC of 2273.542.

This is of course linear mixed effect model. The major challenge in fitting a non linear Brain Cousens model is the lack of model optimizer in the available software of which if found could give a much better fit to the data. The covariance structure of HIV prevalence data shows an autoregressive moving average of order one, which implies that the data is correlated and therefore the correlation will be important in the model.In general all correlations are moderate to large and their magnitudes suggests that correlation should not be ignored.

t -test for difference in means was used to compare the difference in means in the national projections of both the fitted model and EPP. Based on the t-test for difference in means, there is significant difference between the mean values in the national projection in the EPP model compared to that of the mixed effect model fitted on a 95 percent significant level.

4.1 Recommendation

More efforts should be directed to availing reliable HIV prevalence in this country. This should capture the complete demographics related to HIV, condom use, ART coverage so that future statistical models can capture this and more valid conclusions drawn with all important covariates incorporated in the model.

R codes used

```
mean<-tapply(prevalence,province,mean)
sd<-tapply(prevalence,province,sd)
n<-tapply(prevalence,province,length)
n<-tapply(prevalence,site,length)
mean<-tapply(prevalence,site,mean)
sd<-tapply(prevalence,site,sd)
cbind(n,mean,sd)
library(lattice)
xyplot(prevalence~year|province,data=data1,type="b")
xyplot(prevalence~year|site,data=data1,type="b")
qqnorm(prevalence)
scatter.smooth(prevalence
```

```
data1a=read.csv("data1.csv")
data13a=data1a[,c(1,2,3,4,5)]
data13a$clin1=as.numeric(data13a$clinic)
attach(data13a)
```

```
data1=data13a[clin1==1,]
data2=data13a[clin1==2,]
data3=data13a[clin1==3,]
data4=data13a[clin1==4,]
data5=data13a[clin1==5,]
data6=data13a[clin1==6,]
data7=data13a[clin1==7,]
data8=data13a[clin1==8,]
data9=data13a[clin1==9,]
data10=data13a[clin1==10,]
data11=data13a[clin1==11,]
data12=data13a[clin1==12,]
data13=data13a[clin1==13,]
data14=data13a[clin1==14,]
data15=data13a[clin1==15,]
data16=data13a[clin1==16,]
data17=data13a[clin1==17,]
data18=data13a[clin1==18,]
data19=data13a[clin1==19,]
data20=data13a[clin1==20,]
data21=data13a[clin1==21,]
data22=data13a[clin1==22,]
data23=data13a[clin1==23,]
data24=data13a[clin1==24,]
data25=data13a[clin1==25,]
```

```
library(drc)
fit1 <- drm(prevalence ~ year, data = data1, fct = LL.4(),na. Action=na.omit)
fit2 <- drm(prevalence ~ year, data = data2, fct = BC.4(),na.action=na.omit)
fit3 <- drm(prevalence ~ year, data = data3, fct = BC.4(),na.action=na.omit)
fit4 <- drm(prevalence ~ year, data = data4, fct = BC.4(),na.action=na.omit)
fit5 <- drm(prevalence ~ year, data = data5, fct = BC.4(),na.action=na.omit)
fit6 <- drm(prevalence ~ year, data = data6, fct = BC.4(),na.action=na.omit)
fit7 <- drm(prevalence ~ year, data = data7, fct = BC.4(),na.action=na.omit)
fit8 <- drm(prevalence ~ year, data = data8, fct = BC.4(),na.action=na.omit)
fit9 <- drm(prevalence ~ year, data = data9, fct = BC.4(),na.action=na.omit)
fit10 <- drm(prevalence ~ year, data = data10, fct = BC.4(),na.action=na.omit)
fit11 <- drm(prevalence ~ year, data = data11, fct = BC.4(),na.action=na.omit)
fit12 <- drm(prevalence ~ year, data = data12, fct = BC.4(),na.action=na.omit)
fit13 <- drm(prevalence ~ year, data = data13a, fct = BC.4(),na.action=na.omit)
fit14 <- drm(prevalence ~ year, data = data14, fct = BC.4(),na.action=na.omit)
fit15 <- drm(prevalence ~ year, data = data15, fct = BC.4(),na.action=na.omit)
fit16 <- drm(prevalence ~ year, data = data16, fct = BC.4(),na.action=na.omit)
fit17 <- drm(prevalence ~ year, data = data17, fct = BC.4(),na.action=na.omit)
fit18 <- drm(prevalence ~ year, data = data18, fct = BC.4(),na.action=na.omit)
fit19 <- drm(prevalence ~ year, data = data19, fct = BC.4(),na.action=na.omit)
fit20 <- drm(prevalence ~ year, data = data20, fct = BC.4(),na.action=na.omit)
fit21 <- drm(prevalence ~ year, data = data21, fct = BC.4(),na.action=na.omit)
fit22 <- drm(prevalence ~ year, data = data22, fct = BC.4(),na.action=na.omit)
fit23 <- drm(prevalence ~ year, data = data23, fct = BC.4(),na.action=na.omit)
fit24 <- drm(prevalence ~ year, data = data24, fct = BC.4(),na.action=na.omit)
fit25 <- drm(prevalence ~ year, data = data25, fct = BC.4(),na.action=na.omit)
```

Rural

```
par(mfrow=c(3,3))
plot(fit3, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="ChulaimboRHTC")
plot(fit8, broken=TRUE, ylim=c(0,10), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KaplongMH")
plot(fit9, broken=TRUE, ylim=c(0,20), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KarurumoRHTC")
plot(fit14, broken=TRUE, ylim=c(0,25), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="MaraguaDH")
plot(fit15, broken=TRUE, ylim=c(0,13), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="MbaleRHTC")
plot(fit18, broken=TRUE, ylim=c(0,20), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="MosoriotHC")
plot(fit20, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="NjabiniHC")
```

Urban

```
par(mfrow=c(4,3))

plot(fit1, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="babadogoHC")
plot(fit2, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="BusiaDH")
plot(fit4, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="DandoraHC")
plot(fit7, broken=TRUE, ylim=c(0,15), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KakamegaPGH")
plot(fit10, broken=TRUE, ylim=c(0,12), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KisiiDH")
plot(fit11, broken=TRUE, ylim=c(0,15), xlab="year", ylab="prevalence", cex=1 2, lwd=2,main="KisumuPGH")
plot(fit17, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="MombasaPGH")
plot(fit19, broken=TRUE, ylim=c(0,15), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="NakuruPGH")
plot(fit21, broken=TRUE, ylim=c(0,10), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="NyeriPGH")
plot(fit22, broken=TRUE, ylim=c(0,25), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="RirutaHC")
```

Mixed

```
par(mfrow=c(3,3))
```

```
plot(fit5, broken=TRUE, ylim=c(0,15), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="GarissaPGH4")
plot(fit6, broken=TRUE, ylim=c(0,10), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KajiadoDH")
plot(fit12, broken=TRUE, ylim=c(0,30), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KitaleDH")
plot(fit13, broken=TRUE, ylim=c(0,20), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="KituiDH")
plot(fit16, broken=TRUE, ylim=c(0,25), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="MeruDH")
plot(fit23, broken=TRUE, ylim=c(0,15), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="ThikaDH")
plot(fit24, broken=TRUE, ylim=c(0,25), xlab="year", ylab="prevalence", cex=1.2, lwd=2,main="TiwiRHTC")
```

```
coef(summary(fit11))
coef(summary(fit12))
coef(summary(fit13))
coef(summary(fit14))
coef(summary(fit15))
coef(summary(fit16))
coef(summary(fit17))
coef(summary(fit18))
coef(summary(fit19))

coef(summary(fit20))
coef(summary(fit21))
coef(summary(fit22))
coef(summary(fit23))
coef(summary(fit24))
coef(summary(fit25))
```

```
..........
 Rural
.........

logLik(fit3)
logLik(fit8)
logLik(fit9)
logLik(fit14)
logLik(fit15)
logLik(fit18)
logLik(fit20)
```

```
.........
Urban
.........

logLik(fit1)
logLik(fit2)
logLik(fit4)
logLik(fit7)
logLik(fit10)
logLik(fit11)
logLik(fit17)
logLik(fit19)
logLik(fit21)
logLik(fit22)


...........
 Mixed
...........

logLik(fit5)
logLik(fit6)
logLik(fit12)
logLik(fit13)
logLik(fit16)
logLik(fit23)
logLik(fit24)
\newpage




........
Rural
.........

pref3=predict(fit3)
```

```
pref8=predict(fit8)
pref9=predict(fit9)
pref14=predict(fit14)
pref15=predict(fit15)
pref18=predict(fit18)
pref20=predict(fit20)

pref3
pref8
pref9
pref14
pref15
pref18
pref20



prerural=cbind(pref3,pref8,pref9,pref14,pref15,pref18,pref20)



..........
Urban
.........
pref1=predict(fit1)
pref2=predict(fit2)
pref4=predict(fit4)
pref7=predict(fit7)
pref10=predict(fit10)
pref11=predict(fit11)
#predict(fit17)
pref19=predict(fit19)
#predict(fit21)
#predict(fit22)

predurban=cbind(pref1,pref2,pref4,pref7,pref10,pref11,
pref19)
```

```
.............
 Mixed
.............

predict(fit5)
predict(fit6)
predict(fit12)
#predict(fit13)
predict(fit16)
predict(fit23)
predict(fit24)


.........................................
fitting the nlmixed effect model
.........................................
norandom<-drm(prevalence~year,site, data =data13ab, fct = BC.4(),pmodels =
data.frame(clinic, 1, 1, clinic))
random<-mixdrc(norandom,random="e~1|site",data=data13ab)
names(data13a)
data13a

...........
final fits
...........
fit=lme(fixed=prevalence~year+I(year^2),random=~1|site,data=data13a)
fit1=lme(fixed=prevalence~year+I(year^2),random=~1|clinic/site,data=data13a)
cor(cbind(data13a[year==1990,1],data13a[year==1991,1],.....data13a[year==2004,1]
fit2=lme(prevalence~year+I(year^2),random=~1|clinic/site,correlation=corAR1(0.8, form =
~ 1 |province/site),data=data13a)

fit$fitted
plot(rep(seq(1,15),25),fit$fitted[,2])
\newpage
\begin{verbatim}
EPP Results for the country: Kenya

RESULTS FOR TOTAL WORKSET: msc2

Pop in baseyear = 22596510
Baseyear =  2007
Year 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
% HIV+ 0 0.01 0.02 0.05 0.11 0.25 0.56 1.22 2.54 4.9 8.23 11.68 14.2 15.58 16.19 16.4
16.41 16.36 16.27 16.2 16.12 16.05 16 15.97 15.95 15.94 15.95 15.96 15.96 15.98 15.98
Num HIV+ 0 930 2184 5114 11942 27760 63909 144090 310985 617474 1070720 1566786 1961859
2215250 2366360 2462329 2528002 2584801 2634847 2687955 2738842 2793136 2851260 2913741
2980028 3046916 3120077 3192957 3263704 3340966 3416372,
```

```
New HIV
Pop 9520748 9815833 10122311 10440587 10771055 11114082 11469934 11838644 12219732
12611756 13011870 13416020 13820161 14221680 14619905 15015708 15409205 15800672 1619273
16588070 16989219 17398345 17816846 18244737 18679974 19118769 19558317 20000000 2044837
20907721 21380385


END OF msc2



RESULTS FOR CURVEFIT: msc2\Urban:URBAN,NO

Population parameters: Fitting parameters: Epidemiological parameters.

b= 0.07876392 r= 5.04801687 Vert tran= 0.32
l15= 0.84491328 f0= 0.1805008 Fert red= 0.7
mu= 0.00808071 t0= 1980 alpha= 2
gr= 0.03640171 phi= 7311.816699 beta= 13.2123469
percent male= 0.5 lnL= 1017.034178 TimeInPop= 0
Pop=20000000
```

Year 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
% HIV+ 0 0.01 0.02 0.05 0.11 0.25 0.56 1.22 2.54 4.9 8.23 11.68 14.2 15.58 16.19 16.4
16.41 16.36 16.27 16.2 16.12 16.05 16 15.97 15.95 15.94 15.95 15.96 15.96 15.98 15.98
Num HIV+ 0 930 2184 5114 11942 27760 63909 144090 310985 617474 1070720 1566786 1961859
2215250 2366360 2462329 2528002 2584801 2634847 2687955 2738842 2793136 2851260 2913741
2980028 3046916 3120077 3192957 3263704 3340966 3416372
New HIV
Pop 9520748 9815833 10122311 10440587 10771055 11114082 11469934 11838644 12219732
12611756 13011870 13416020 13820161 14221680 14619905 15015708 15409205 15800672 1619273
16588070 16989219 17398345 17816846 18244737 18679974 19118769 19558317 20000000 2044837
20907721 21380385

END OF msc2\Urban:URBAN,NO

RESULTS FOR CURVEFIT: msc2\Rural:BOTH,NO

Population parameters: Fitting parameters: Epidemiological parameters:

b= 0.07876392 r= 9.44995671 Vert tran= 0.32
l15= 0.84491328 f0= 0.08090026 Fert red= 0.7
mu= 0.00808071 t0= 1980 alpha= 2
gr= 0.03640171 phi= 71678.8853 beta= 13.2123469
% male= 0.5 lnL= 756.0037132 TimeInPop= 0
Pop= 0
Baseyear = 2007

Year 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
% HIV+ 0 0 0.01 0.01 0.03 0.06 0.11 0.22 0.43 0.8 1.44 2.41 3.66 4.99 6 6.73 7.05 7.22
7.27 7.27 7.23 7.18 7.15 7.13 7.1 7.06 7.07 7.06 7.04 7.04 7.05
Num HIV+ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
New HIV
Pop 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
M/F ratio NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA

END OF msc2\Rural:BOTH,NO

References

1. Robert E.Weiss.(2005) Modeling Longitudinal Data
2. A.J. Dobson (2002) An Introduction to Generalized Linear Models, Second Edition
3. Cheluget et al(2006) Evidence for population level declines in adult HIV prevalence in Kenya
4. Montana L.S. Et al.(2008) Comparison of HIV prevalence estimate from ante-natal care surveillance and population based on sub-Saharan Africa
5. The policy project, the national AIDS control program, Malawi(2000) Estimating National Prevalence in Malawi from sentinel surveillance data
6. National AIDS/STI control program, Accra(2001). Estimating HIV Prevalence in Ghana using sentinel surveillance data
7. NASCOP Kenya (2005). sentinel surveilance of HIV and STDs in Kenya
8. Joshua A et al (2001). Modeling HIV/AIDS epidemic in sub Saharan Africa using data from ante natal clinics
9. Somi et al(2005) Estimating and projecting HIV prevalence and AIDS deaths in Tanzania using antenatal surveillance data
10. Oliver Schabenberger Francis J. Pierce (2002) CONTEMPORARY STATISTICAL MODELS for the Plant and Soil Sciences
11. Pinheiro and Bates(2002) Mixed effect models in S and S-PLUS
12. Brain, P. and Cousens, R. (1989) An equation to describe dose responses where there is stimulation of growth at low doses, Weed Research. 29, 9396.
13. Van Ewijk, P. H. and Hoekstra, J. A. (1993) Calculation of the EC50 and its Confidence Interval When Subtoxic Stimulus Is Present, Ecotoxicology and Environmental Safety, 25, 2532.
14. W Hladik et al(2005) HIV/AIDS in Ethiopia:where is the epidemic heading to.
15. L. Alkema et al(2008) Bayesian melding for estimating uncertainty in natural HIV prevalence estimates.

16. Lloyds J. Edwards(2002), Modern statistical Techniques for the analysis of longitudinal data in biomedical research.

17. Dr.Hector de Arazoza(2000), What percentage of Cuban HIV-AIDS epidemic is known?