Multilevel Modeling of HIV in Swaziland Using Frequentist and Bayesian Approaches

Sifiso E. Vilakati

Thesis Submitted in Fulfillment of the Requirement for the

Degree of Master of Science in Statistics

School of Mathematics, Statistics and Computer Science

University of Kwa Zulu Natal

South Africa

2012

Declaration

I, Sifiso Vilakati, declare that this thesis titled, 'Multilevel Modeling of HIV in Swaziland Using both Frequentist and Bayesian Approaches' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- No part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Mr. Sifiso Vilakati	Signed	Date .....
Dr.Thomas Achia	Signed	Date .....
Prof. Henry Mwambi	Signed	Date .....

Acknowledgements

I am deeply indepted to my supervisors Dr. T. Achia and Professor H. Mwambi for their guidance throughout my masters. Without their assistence this work could not have been completed. I am thankful to the College of Agriculture, Engineering and Science for the bursary I received. I am also thankful my employer, the Ministry of Education in Swaziland for releasing me from duty so that I can further my studies.

Abstract

Multilevel models account for different levels of aggregation that may be present in the data. Researchers are sometimes faced with the task of analysing data that are collected at different levels such that attributes about individual cases are provided as well as the attributes of groupings of these individual cases. Data with multilevel structure is common in the social sciences and other fields such as epidemiology. Ignoring hierarchies in data (where they exist) can have damaging consequences to subsequent statistical inference.

This study applied multilevel models from frequentist and Bayesian perspectives to the Swaziland Demographic and Health Survey (SDHS) data. The first model fitted to the data was a Bayesian generalised linear mixed model (GLMM) using two estimation techniques: the Integrated Laplace Approximation (INLA) and Monte Carlo Markov Chain (MCMC) methods. The study aimed at identifying determinants of HIV in Swaziland and as well as comparing the different statistical models. The outcome variable of interest in this study is HIV status and it is binary, in all the models fitted the logit link was used.

The results of the analysis showed that the INLA estimation approach is superior to the MCMC approach in Bayesian GLMMs in terms of com-

putational speed. The INLA approach produced the results within seconds compared to the many minutes taken by the MCMC methods. There were minimal differences observed between the Bayesian multilevel model and the frequentist multilevel model. A notable difference observed between the Bayesian GLMMs and the the multilevel models is that of differing estimates for cluster effects. In the Bayesian GLMM, the estimates for the cluster effects are larger than the ones from the multilevel models. The inclusion of cluster level variables in the multilevel models reduced the unexplained group level variation.

In an attempt to identify key drivers of HIV in Swaziland, this study found that age, age at first sex, marital status and the number of sexual partners one had in the last 12 months are associated with HIV serostatus. Weak between cluster variations were found in both men and women.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is estimated that thirty eight million people are living with HIV/AIDS in the world and that over 20 million have died since the beginning of the pandemic [19]. The Sub-Saharan Africa has the highest prevalence rate of HIV/AIDS infection. Two thirds (64%) of the people living with the virus are in Sub-Saharan Africa [33]. The HIV prevalence varies considerably accross the the Sub-Saharan Africa region, with the driving forces of the disease being varied and diverse. The variation of the epidemic has been thought to be a product of local and social determinants among which are culture, religion, poverty, gender, and migration. Other drivers include biological, environmental and behavioral [33].

In Sub-Saharan Africa, the Southern African region has the highest prevalence rate of HIV infection. Estimates suggest that South Africa counts

more than one thousand new infections on a daily basis, while in Botswana, Lesotho, Namibia, Swaziland and Zimbabwe at least one in five adults has HIV. There are multiple epidemics in Africa. The continent is home to a number of different epidemics, each of them with its own epidemiological and social characteristics and variations. Majority of persons infected with HIV are living with the HIV-1 virus. HIV-2, which is rare, occurs in parts of West Africa. In Southern Africa HIV-1 is the predominant type which is causing serious havoc [46].

1.1 Literature Review

1.1.1 Background

Swaziland is the smallest landlocked nation in Southern Africa, bordered by Mozambique in the east and by South Africa in the west, north and south. Its landscape creates four distinct ecological zones ranging from the tropical lowveld to the temperate highveld. The country's sub-tropical climate, with summer rainfall patterns, creates risks for both drought and flooding. The country is also divided into four administrative regions being Hhohho, Manzini, Shiselweni and Lubombo [23]

The population of the country is around one million of which 44% are

under the age of fifteen. Like other countries in the Southern African region, Swaziland is affected by the HIV/AIDS pandemic. It is estimated that the prevalence rate of HIV/AIDS in Swaziland stands at 42.6%. The highly affected regions being Manzini (42.5%) and Lubombo (41.9%). The people of Swaziland known as Swazis are homogeneous in their language, culture and tradition. The majority of the residents speak SiSwati. SiSwati and English are the country's official languages. About 80% of the countrys population live in rural areas and depend largely on subsistence farming for their livelihood [30].

Prevalence of HIV in Swaziland

Swaziland has the highest prevalence rate of HIV/AIDS in the Sub-Saharan region. The first HIV/AIDS case in Swaziland was reported in 1987. Since that time, the virus has kept spreading in the population. By 2004 the prevalence rate was standing at 42.6%. This has been a great concern for the government and the general public. Results from the HIV sero-surveillance among people attending antenatal clinics show that the overall level of HIV infection in pregnant women increased more than 10 times, that is from 3.9% in 1992 to 42.6% in 2004. By 2006 it however went down to 39.2% [40].

1.1.2 Drivers of HIV Transmission

This study uses a theoretical framework proposed by Johnson et al [19]. Drivers of HIV/AIDS can be categorized into the following; behavioral, biological, socioeconomic, and demographic and residential characteristics.

Behavioral factors

Behavioral risk determinant factors associated with HIV/AIDS include condom use, frequency of alcohol use, number of sexual partners the individual had in the past (eg. past year) and exchange of sex for money.

According to SHDR[3], more than 40% of adults in Swaziland used condoms. It was also found that there were cases where both educated and non educated women were forced to have sex without a condom. Among tertiary students 72% reported to be using condoms, and 84.9% of secondary school students reported condon usage. Only 49.9% of youth out of school used condoms. In a study by Johnson et al [19] in Kenya, it was found that sex related behavioral factors did not have a significant contribution in explaining HIV infection. A Ugandan study by Ahmed et al [3] found that consistent condom use provides protection against HIV/AIDS while inconsistent or non-condom use is not protective. In their study condom use was found to be higher among males, younger unmarried and better edu-

cated individuals, and those reporting multiple sex partners or extramarital relationships.

The SHDR [40] reported that among the youth, sex with non-regular partners was as low as 15.7% for school going youths. This was not the case with out of school youths as their rate was at 49%. Students in institutions of higher learning had practiced this high risk behaviour of having more than one sexual partners at a rate of about 44.1% within the last 12 months of the survey. In a simulation study by Epstein and Morris [32], concurrent partnership was found to exponentially increase the number of infected individuals and the growth rate of the epidemic. It was also found that concurrent partnerships increases the speed with which the virus spreads in the population. The regression estimates indicated that each 10% increase in average number of concurrent partnership increase the rate of spread of the virus by about 12%. Concurrency is the case where an individual has more than one sexual partners at the same time. According to Epstein and Morris [32] the strength of evidence linking concurrecy to HIV epidemic severity in the southern and eastern Africa led to the Joint United Nations Programme on HIV/AIDS and SADC in 2006 to reach the conclusion that high rates of concurrent sexual partnerships, combined with low rates of male circumcision and infrequent use of condoms, are major drivers of the AIDS epidemic in Southern Africa.

In a study done by Kalichman et al [20] in Cape Town, association be-

tween alcohol and HIV/AIDS was found. The association between alcohol use and and sexual risk behaviour may be accounted for by sensation seeking personality . Johnson et al [19] found that there are higher risks of HIV infection for women who have ever consumed alcohol compared with women who report they never have. On the males side, it was found that those who took alcohol for 11 to 19 days in the last past month were 2.5 times as likely to be HIV positive than those who have never taken alcohol. Buseh, [8] studied patterns of sexual behaviour among secondary school students in Swaziland. In his study he found that less than half (43.2%) of the students indicated that they consumed alcohol bevarages. More males (51.2%), in contrast to females (35.7%), reported alcohol use. More urban students reported alcohol consumption than students from rural places. Among those students who were currently sexually active, 18% reported taking alcohol at their most recent sexual intercourse. Male students were significantly more likely than female students to have used alcohol at the time of last intercourse. Alcohol and drug use with sexual intercourse increased with age.

Many women are faced with economic problems, so sex becomes a strategy for survival, with women selling sex to meet financial obligations such as paying of school fees and buying of food. Reward for sexual services may range from occasional cash payments to supplementation of income with gifts [1]. Weiser et al [44] established the association between food insufficiency

and exchange of sex for money or gifts. In their study they found that food insufficiency was associated with over 2 times the odds of engaging in sex exchange. After adjusting for known covariates, food insufficiency was still associated with nearly 2 times the odds of sex exchange. Also a dose response relationship between alcohol use and sex exchange was observed.

Socioeconomic factors

Evidence of associations between socioeconomic status and the spread of HIV in different settings and at various stages of the epidemic are is still rudimentary [11]. Socioeconomic factors related to the spread of HIV/AIDS can be categorized as follows;

In their study Gillespie et al [11] concluded that the assertion that poverty is the main driver of HIV is too simplistic. Relative wealth appears to have a mixed influence on HIV risk depending on an array of contextual factors. Johnson et al [19] found that in both men and women, wealth was positively related to HIV risk, education on the other hand did not show the same relationship. Lopman et al [25] studied the association between wealth index which is based on household ownership and HIV incidence, HIV mortality and sexual behaviour. The study found that the greatest decrease in HIV prevalence occurred in the highest wealth index tercile in both men and women. In men HIV incidence was lowest at the top wealth index tercile.

Contrary to the findings of Lopman et al [25], Mishra et al [31] established that in all the eight countries they studied, adults in the wealthiest quintiles had higher prevalence of HIV than those in the poorer quintile. However, the positive association between wealth and HIV risk was not established in multivariate models.

Similar to the findings of Lopman et al [25] are the results of Hangreaves et al [18] who noted that among men, there was litte evidence that HIV seroconversion was associated with any socioeconomic factor. For women negative relationship was found between education and HIV. In a study by Glynn et al [12] in four African cities, association between schooling and HIV infection was not found in Kisumu or Ndola. Women of Yaounde and men of Cotonou with more schooling were less likely to be HIV positive. In all the cities studied, men and women with more education tended to report less risky sexual behaviour. Similarly Barninghausen et al [5] established that an increase in educational attainment by one grade reduced the hazard of HIV seroconversion by approximately 7%.

According to Lurie et al [26], the precise manner in which migration contributes to sexually transmitted diseases is complex and not well understood. The link between mobility and the spread of HIV is determined by the structure of the migration process, the conditions under which it occurs, including poverty, exploitation, separation from partners and families and separation

from the sociocultural norms that guide behaviours within communities [11].

According to Gillespie et al [11], there is convincing evidence linking the spread of HIV and mobility. In sub-Saharan Africa the risk of contracting HIV has been found to be higher in migrants and mobile individuals like truck drivers. In a study by Mayer [29], migrant men were found to be 26 times more likely to have been infected by an outside partner than their own wives. Also, Lurie et al [26] found that migrant men were more likely than non-migrant men to have at least one current casual sexual partner. Women were reluctant to disclose information on their casual sex partners. More sexual risk activity and an increased HIV risk were observed not only in mobile persons, but also in partners staying behind [21].

A study by Shisana et al [41] demostrated that there is a link between marital status and HIV infection. Married people are less likely to be infected with HIV compared to unmarried people. The relationship between HIV/AIDS and marital status is a complex one. In their study [41] the relationship diminished after controlling for other socio-demographic risk factors for HIV. Johnson et al [19] found that marital status is a significant factor for women, in particular those widowed or divorced.

A Malawian study by Trinitapoli et al [43] found that men belonging to Pentecostal churches consistently reported lower levels of both HIV risk and perceived risk. Regular attendance to religious services is linked with

reduced odds of reporting extra marital partners with lower level of perceived risk of infection. Interestingly Johnson et al [19] found that for both men and women those who considered themselves to be at a low risk of contracting HIV were infact the most likely to be HIV positive.

Biological Factors

The main biological factors that are related to HIV risk include circumcision or lack of it, presence of an STI, and recent birth status [19].

Male circumcision is the surgical removal of the foreskin which is thought helps to reduce the risk of HIV infection [40]. Johnson et al [19] found that the outstanding biological factor associated with HIV risk was circumcision. Men who were not circumcised were 4 times likely to be HIV positive than circumcised men. It was found that circumcision in men has a protective effect against HIV infection. Similar results were obtained by Weiss et al [45] who established that there is compelling evidence that male circumcision has a protective effect against HIV infection in sub-Saharan Africa, and especially in high risk populations. The same results were achieved by other studies [15, 4] done on male circumcision.

Fertility is greatly reduced in HIV-1 infected women compared to those who are not HIV positive [16]. A study by Johnson et al [19] found that women who had a birth 5 years before the survey was conducted were 30%

less likely to be HIV positive than those who did not have a birth in the past 5 years.

There is positive association between non-HIV sexually transmitted infections, particularly those that are ulcerative with HIV [19]. Strong evidence indicates that both ulcerative and non-ulcerative STIs promote HIV transmission by augmenting HIV infectiousness and HIV susceptibility via a variety of biological mechanisms.

Demographic and Residential Factors

A study by Gouws et al [14] established that the patterns (in terms of age) of HIV infection among SADC countries are similar. The prevalence of HIV in these countries was found to increase after the age of 15, and more rapidly among women than men, reaching the peak among women in their twenties and men in their thirties. Similar results were reached by Johnson et al [19] who found that men in the ages 35 to 44 years had the highest risk of being HIV positive, whilst women in the ages 25 to 29 had the highest risk of being HIV positive.

Johnson et al [19] found significant associations between region and HIV risk in Kenya. Men and women from the province of Nyanza were significantly at higher risk than those from Nairobi. Kleinschmidt et al [22] studied the spatial variation of HIV/AIDS in South Africa. The spatial maps

showed pronounced variations in HIV prevalence in the provinces. Within the Lubombo district of Swaziland, variations in the intensity of the virus from community to community was also observed [23].

A study in India found differences in HIV prevalence in rural and urban communities [42]. The prevalence in rural areas was reported to be 7.2% and 7.0% in urban areas. Differences in HIV prevalence were also observed in a study by Boerma et al [6]. In the small geographic area which they studied striking differences were seen; HIV prevalence in the trading centre was twice that in the area surrounding it (within 2km) and three to four times than that of rural villages 8 km away from the trading centre. Johnson et al [19] found that, among women those living in the rural areas were 0.5 times likely to be HIV positive than those in urban areas. In addition a clear association between being HIV positive and the number of one's children that died was found. Those who experienced the death of one child were twice as likely to be HIV positive, whilst those who experieced the death of two or more children were even more likely to be HIV positive.

1.2 Problem Statement

The exceptionally high prevalence of HIV infection in most of the South African countries, in particular Swaziland, has raised complex questions

about the factors that have contributed to the rapid spread of the virus in the region and about the eventual prevalence the epidemic might reach [26]. "Risk factors for the acquisition and transmission of HIV through a heterosexual route are well characterized, but their relative importance varies by location, and the relationship between biological, behavioral, and social risk are incompletely understood" [29]. In order to use intervention programmes to curb the spread of HIV/AIDS the key drivers of the pandemic must be clearly identified and understood.

1.3 Objectives

Majority of studies on the risk factors associated with HIV infection have largely focused on individual risk factors. Sociological theories, however, have long suggested that individuals health and behaviour is shaped not only by individual risk factors but also by the structure of the environment in which they live [27]. Statistical development has made it possible for researchers to test these theories. These statistical models allow researchers to examine the additive and interactive effects contextual factors that affect sociological and health outcomes at the population and individual level. Multilevel models, in particular, have been identified as highly appropriate.

This study has the following specific objectives:

- To account for heterogeneity in the distribution of HIV in Swaziland by fitting a Bayesian generalised linear mixed model (GLMM) and Multilevel models.

- To identify the key determinants of HIV in Swaziland.

- To compare the results of the different statistical models used to analyse the data.

- To compare the results of the estimation techniques in fitting Bayesian GLMMs namely, Markov Chain Monte Carlo (MCMC) methods and the Integrated Nested Laplace Approximation (INLA) approach.

1.4 Organization of the study

In this chapter, we have introduced some background information about the study. A review of literature on the determinants of HIV has been done and the objectives have been outlined. The rest of the thesis will organised as follows.

In Chapter 2 we introduce the SDHS data to be used in this study. The manner in which the data was collected is described. An exploratory data analysis is performed to identify potential covariates of HIV infection. The chapter ends with some cross tabulation of HIV serostatus with some po-

tential covariates. Statistically significant covariates are later used in the modeling of HIV in Swaziland.

Bayesian generalised linear mixed models are fit to the data in Chapter 3. In this chapter, two estimation procedures, namely the Markov Chain Monte Carlo (MCMC) and the Itegrated Laplace Approximation (INLA) are considered. The two estimation techniques results are compared.

In Chapter 4, multilevel models are discussed. The frequentist multilevel model and the Bayesian multilevel models are considered. The two models are fit to the data and their results compared. We also discuss multilevel modeling using the idea of pooling and as a result a complete pooling model is also fit. The last chapter focuses on discussing the results from the different models and give a summary of the key determinants identified and possible interventions. Future research areas are also highlighted.

Chapter 2

Data

In this chapter, we introduce the data set used in this thesis: the Swaziland Demographic and Health Survey (SDHS). Also, we perform some simple descriptive statistics to explore the relationships that exist between the variables. A χ^2 test of association is used to identify possible covariates related to HIV status.

2.1 Background

This study used data from the 2006-7 Swaziland Health and Demographic Survey (SDHS). The survey was designed to provide health estimates and demographic indicators at the national level, for urban and rural places across the four regions of Swaziland. A two-stage sample design was used. Sample

points were selected from a sample frame of enumerating areas (EAs) defined in the 1997 Swaziland Population and Housing Census from the Central Statistics Office (CSO). In total, 275 clusters were drawn from the list, 111 clusters were from urban areas and 164 from rural places.

Households from the selected clusters were listed and a systematic sample of households was drawn for a total of 5500 households. From these households, eligible men and women aged 15-49 were identified and interviewed. For HIV testing, all eligible men and women from the selected households were asked for their consent to be anonymously tested for the virus. Samples for testing were obtained by collecting blood drops from a sterile fingerprint onto a filter paper card [39].

2.1.1 The Data Sets

The 2006 DHS was the first to have included voluntary testing of HIV in Swaziland. Prior to this survey, HIV prevalence has been primarily derived from sentinel surveillance of pregnant women. The national sentinel surveillance system consists of 17 sites in government and mission health facilities which are selected to represent the differrent groups, regions, rural and urban populations.

While the rate of HIV infection in pregnant women has been shown to

be a reasonable proxy for the HIV prevalence in the combined male and female adult population, it has several limitations. The first limitation is that antenatal clinics data does not capture any information on HIV prevalence in non-pregnant women and from those women who do not attend clinic for pregnancy care or receive antenatal care at facilities not represented in the surveillance system. Pregnant women are more at risk of HIV infection than women who may be avoiding both HIV and pregnancy through the use of condoms. Also, there may be biases in the antenatal clinics surveillance data because HIV infection reduces fertility in women and knowledge of HIV status may influence fertility choices.

The SDHS data comprised of three data sets; for males (4675), for females (5301) and the third one for HIV for both males and females. The HIV data set was merged with the male and the female data sets. In both data sets, cases where HIV test results were missing were deleted. Thus our analysis uses complete case analysis. Out of the 5301 eligible women only 4987 were tested for HIV and 4156 of the 4675 men were tested for HIV. HIV results were missing for individuals who refused to be tested and who were absent at the time of the blood collection. Variables of interest that were not categorical were recoded. These include age at first sex, number of ttimes away from home in the last 12 months, number of sexual partners in the last 12 months and number of births in the past 5 years. The variable

religion had more than 10 categories. To reduce the number of categories it was recoded into 5 categories. The key outcome variable of interest is HIV status and a summary of the individual level variables is given below.

- Demographic and residential characteristics: current age, region, and type of place of residence.
- Social factors: wealth index, educational level, number of trips in the last 12 months, current marital status and religion.
- Biological factors: circumcision, recent birth status, and recent experience with STI.
- Behavioural factors: number of sexual partners, alcohol use, and condom use.

Cluster level variables are not directly available from the SDHS data set. We constructed cluster level variables by aggregating individual characteristics at the cluster level. Cluster level variables are either averages or proportions depending on the nature of the variable being created, for instance we have cluster level average wealth and cluster level proportion HIV positive. The cluster level variables were created in SPSS and Microsoft Access was used to merge the individual level and the cluster level data sets. Cluster level variables used in the study are cluster level aver-

age wealth(cwealth.full), cluster level average age(cage.full) and cluster level proportion HIV positive(chiv.full).

2.2 Exploratory Data Analysis (EDA)

A fundamental initial step in data analysis is to perform an exploratory analysis. EDA provides a better understanding of the relationships or associations that exist between the variables in the data set and also highlights any anomalities therein. In doing the exploratory data analysis, we used the svy function in STATA to compute weighted prevalence rates for selected covariates determined from the literature review.

Table 2.1 presents the results of cross classifying HIV status with selected covariates. The results suggest that the overall prevalence rate in males is at 19.5% compared to 31.4% among females. The difference is huge hence the need to model the two data sets separately. The overall HIV prevalence in the Swazi adult population is at 26%.

Table 2.1: Weighted HIV Prevalence by Selected Covariates

	Males		Females	
Variable	Percent	N	Percent	N
Age				
15-19	1.9	20	10.0	121
20-24	12.4	91	38.4	365
25-29	27.7	141	49.1	325
30-34	43.7	154	45.0	256
35-39	44.6	142	37.6	174
40-44	40.9	92	27.7	117
44-49	28.0	64	21.5	80
Region				
Hhohho	23.0	192	33.5	374
Manzini	18.2	181	30.3	410
Shiselweni	16.0	123	29.0	309
Lubombo	20.8	208	31.0	345
Residence type				
urban	25.6	289	36.8	488
rural	17.4	415	29.1	950
Marital status				
never married	9.8	233	25.7	596
married	33.1	275	28.0	406
living together	48.2	100	46.8	212
widowed	68.5	33	55.9	146
divorced	51.2	13	47.1	8
not living together	55.0	50	51.5	70
Educational level				
no education	31.1	88	39.0	161
primary	17.9	237	33.6	528
secondary	18.4	318	28.7	668
Higher	23.1	61	25.5	81
Wealth				
poorest	19.8	104	31.5	237
poorer	19.8	114	31.9	263
middle	16.9	122	31.3	287
richer	20.8	161	31.5	313
richest	20.2	203	29.2	338
Age at first sex				
never had sex	2.1	22	5.2	42
14 or less	14.5	19	38.6	149
15-17	28.8	231	36.5	644
18+	28.4	432	36.2	603
Had an STI in last 12 months				
no	17.7	601	29.7	1301
yes	48.6	98	57.5	134
don't know	1 25.1		7.3	1
Number of sexual partners				
no partner	5.0	68	19.5	1301
1 partner	27.9	461	35.7	134
2+ partners	36.1	175	52.4	1
Condom Use				
no	31.5	94	45.7	202
Yes	26.7	198	44.0	324
Any births				
none	-	-	25.8	688
Yes	-	-	37.8	750
Overall prevalence	19.5	703	31.4	1439

HIV prevalence peaks at 49.1% for women in the age group 25-29. For men, HIV prevalence peaks at 44.6% in the age group 35-39 as also shown in Figure 2.1.

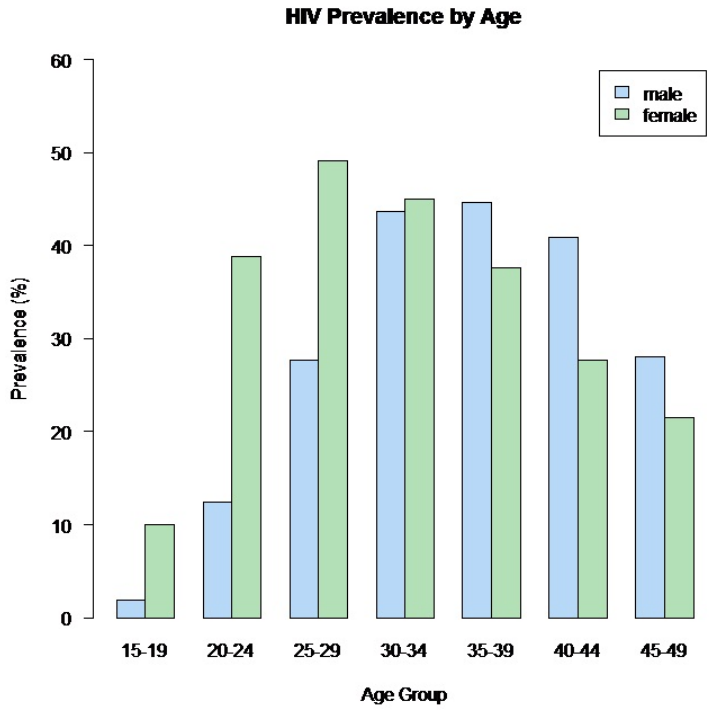


Figure 2.1: HIV Prevalence by Age

For both males and females the prevalence rate is low at the lower age groups but it then increases sharply (especially for women) which is then followed by a decline. HIV prevalence is higher in urban areas than in rural areas for both men and women. The HIV prevalence rate is at 25.6% for males and 36.8% for women in urban areas compared to 17.4% for males and

29.1% for females in rural areas.

However, there is no much regional variation in the prevalence rate. Hhohho has the highest prevalence rate at 23.0% for males and 33.5% for females followed by the Lubombo region at 20.8% for males and 31.0% for females. The Shiselweni region has the lowest HIV prevalence rate at 16.0% among males and 29.0% among females as shown in Table 2.1.

Females who started having sex at the age of 14 or less have HIV prevalence rate at 38.6% while males on the same category have prevalence rate at 14.5%. Similarly, for the category 15–17 years, among the females the HIV prevalence rate is at 36.5% compared to 28.8% among males. Regardless of the age first sex was encountered females have higher HIV prevalence accross all ages of first sex as summarised by Figure 2.2 below.

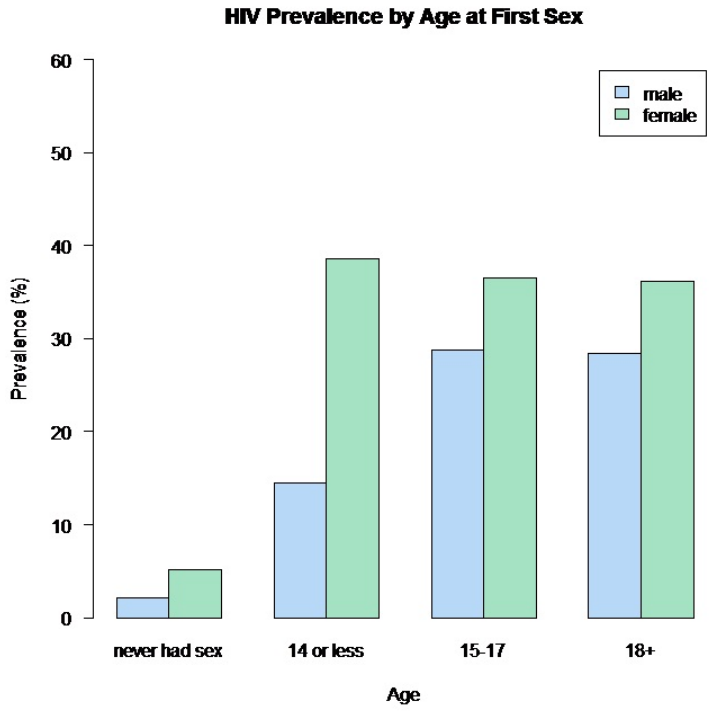


Figure 2.2: HIV Prevalence by Age at First Sex

Now coming to current marital status, for both males and females the highest prevalence rate (68.5% and 55.9% respectively) is in the widowed category. Most probably their spouses died of HIV/AIDS. Note worthy is the result in the statuses; never married and married which have the lowest HIV prevalence rates compared to the rest.

HIV prevalence is higher among women with no education (39.0%) and with primary education (33.6%). On the other hand, men in the higher education (23.1%) category have the second highest prevalence rate. In both

men and women the first category, no education, has the highest prevalence rate. For both males and females there seems to be no much variation in the prevalence of HIV over the different economic categories.

The prevalence rate for men who reported condom use every time they had sex with the last sex partner (last 12 months) was 26.7%. This is lower than the prevalence rate for those who reported non-condom use (31.5%). Similar results were obtained from the women, where those who reported non-condon use had high HIV prevalence rate (45.7%) compared to those who were consistently using condoms. In both males and females, those who reported to have had an STI in the last 12 months had the highest HIV prevalence. Women who had a birth have HIV prevalence rate at 37.8% compared to 25.8% for those who had no birth in the last 5 years.

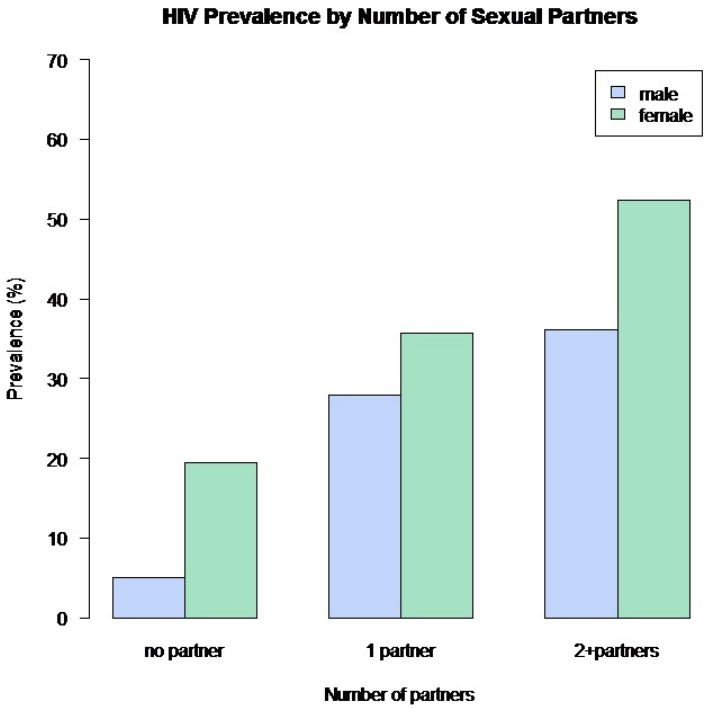


Figure 2.3: HIV Prevalence by Number of Sexual partners

From Figure 2.3, it is evident that the prevalence rate of HIV increases with an increase in the number of sexual partners in the last 12 months in both males and females. The HIV prevalence rate for men and women who had one or no sexual partner is low compared to those had two or more sexual partners.

2.3 Test of Association

In this section we present the results of the χ^2 test performed on the male and the female data sets. The results presented in Table 2.2 show that there is a positive association between HIV status and age (χ^2: 621.3, p-value: < 0.001 and χ^2: 451.0, p-value: < 0.001) for both men and women. Association between HIV and region (χ^2: 15.3, p-value: 0.033) was found in male population. There was however no association found between HIV status and region(χ^2: 5.6, p-value: 0.291) in women. No significant association was found between HIV status and wealth in both men and women whilst significant association was established between HIV status and type of place of residence and highest educational level. Circumcision, a medical practice done in men, was not associated with HIV status (χ^2: 0.33, p-value: 0.556). In this test, HIV status and number of births (χ^2: 76.0, p-value: < 0.001) were found to be related.

For both males and females, association between HIV status and the number of partners in the last 12 months was found, but no relationship was found between HIV status and condom use. Table 2.2 summarises the results of the crosstabulation.

Table 2.2: Crosstabulation of HIV Status with Selected Covariates

	Males			Females		
Variable	χ^2	df	p-value	χ^2	df	p-value
Age	621.3	6	< 0.001	451.0	6	< 0.001
Region	15.3	3	0.033	5.6	3	0.291
Type of place residence	29.5	1	< 0.001	23.3	1	< 0.001
Highest educational level	29.4	3	< 0.001	34.0	3	< 0.001
Wealth index	4.4	4	0.526	2.15	4	0.764
Current marital status	425.2	5	< 0.001	191.7	5	< 0.001
Age at first sex	345.4	3	< 0.001	308.5	3	< 0.001
Religion	6.7	4	0.154	1.4	4	0.843
Circumcision status	0.33	1	0.566	-	-	-
Had any STI in last 12 months	121.6	2	< 0.001	82.8	2	< 0.001
Condom use	1.6	1	0.203	1.3	1	0.549
Alcohol consumption	7.0	4	0.135	9.9	4	0.143
Number of sexual partners	357.5	3	< 0.001	145.9	3	< 0.001
Frequency of times away from home	46.2	4	< 0.001	16.1	4	0.03
Any births	-	-	-	76.0	1	< 0.001

2.4 Summary

In this chapter, we have introduced the data set used in the study. Exploratory data analysis was also done on the data. The HIV prevalence rate between males and females differs significantly hence in further analysis the two data sets (males and females) were analysed separately. Covariates found to be associated with HIV status were used in further analysis. The variable 'region' was found to be statistically significantly associated with HIV status in males but was statistically insignificant in females. Since it was found to be significant in males, it was used in further analysis.

Chapter 3

Bayesian Generalised Linear Mixed Models

In this chapter, we discuss the Bayesian genealised mixed model (GLMM). The SDHS data is clustered according to enumerating areas and to account for this heterogeneity we use GLMMs which are an extension of generalised linear models (GLMs). The chapter is divided into six sections. The first section is basically a brief review of the Bayesian inference. The second section focuses on the Bayesian generalised linear mixed model and prior distributions commonly used in the Bayesian GLMMs. The subsequent sections deals with estimation of parameters and the application of the model on the SDHS data. Two estimation methods are considered; the Monte Carlo Markov Chain (MCMC) and Integrated Laplace Approximation (INLA) methods,

and both methods are applied on the data in the last section.

3.1 Introduction

In this section we introduce general concepts behind Bayesian inference. These general concepts are the basis for Bayesian inference and are extended to more specific Bayesian models.

Let $\boldsymbol{y}$ be the data following an assumed parametric distribution with probability density function $f(\boldsymbol{y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. In the Bayesian paradigm, parameters are regarded as random variables, that is, the parameters have a probability density function $f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\boldsymbol{\theta}_0)$ known as a prior distribution. In the prior distribution, the parameters $\boldsymbol{\theta}_0$ are referred to as hyperparameters, and are assumed known. These hyperparameters maybe chosen based on expert opinion, similar previous studies or even non-informative (in the absence of prior information).

In Bayesian modeling, inference is based on the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$. Given a prior distribution $f(\boldsymbol{\theta})$, the posterior distribution is obtained via the Bayes Theorem:

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}). \tag{3.1}$$

In words we say that the *Posterior* is proportional to the *Likelihood* times the *Prior*.

3.2 The Model

Bayesian generalised linear mixed models are an extention of the Bayesian linear mixed model (LME). The Bayesian GLMMs differ from the Bayesian LMEs in the following ways:

- There is a non linear relationship between the response, the parameters and the random effects.
- The response does not follow a normal distribution but instead a distribution in the exponential family.

Suppose that the responses $\{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$ in the ith cluster are conditionally independent given the mean parameters $\boldsymbol{\beta}$ and the random effects $\boldsymbol{u}_i$. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$. A full Bayesian generalised linear mixed model (GLMM) can be stated as;

$$E(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{u}_i) = g(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i), i = 1, \ldots n_i, \tag{3.2}$$

$$\boldsymbol{u}_i \sim N(0, \boldsymbol{\Omega}),$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0),$$

$$\boldsymbol{\Omega} \sim W_q^{-1}(\eta, \Psi),$$

where $g(.)$ is a known inverse link function (e.g. the inverse logit link) and $\boldsymbol{X}_i$ $(n_i \times p)$ and $\boldsymbol{Z}_i$ $(n_i \times q)$ are known design matrices, $W_q^{-1}(\eta, \Psi)$ is the

inverse Wishart distribution with parameters η and Ψ, and $\boldsymbol{u}_i(1 \times q)$ are random effects. We assume that the hyperparameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ are known. Let $\boldsymbol{u}_i = (u_1, u_2, \ldots, u_q)$. Let us further assume that the prior distributions are independent, that is;

$$f(\boldsymbol{\beta}, \boldsymbol{\Omega}) = f(\boldsymbol{\beta}) f(\boldsymbol{\Omega}).$$

Then it follows that the posterior distribution of all parameters can be written as

$$f(\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{u}|\boldsymbol{y}) \propto \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i) f(\boldsymbol{\beta})\right] \left[\prod_{i=1}^{n} f(\boldsymbol{u}_i|\boldsymbol{\Omega}) f(\boldsymbol{\Omega})\right]. \tag{3.3}$$

The full conditionals for Bayesian inference are then given by

$$f(\boldsymbol{\beta}|\boldsymbol{\Omega}, \boldsymbol{u}, \boldsymbol{y}) \propto \prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i) f(\boldsymbol{\beta}),$$

$$f(\boldsymbol{u}|\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{y}) \propto \prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i) f(\boldsymbol{u}_i|\boldsymbol{\Omega}),$$

$$f(\boldsymbol{\Omega}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{y}) \propto \prod_{i=1}^{n} f(\boldsymbol{u}_i|\boldsymbol{\Omega}) f(\boldsymbol{\Omega}),$$

where

$$[\boldsymbol{\Omega}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{y}] \sim W_q^{-1}(\eta + n/2, \Psi + \sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i^T/2)$$

3.2.1 Prior Distributions

In Bayesian inference the choice of prior distributions is of paramount importance. Much of the controversy surrounding the use of Bayesian methods

revolves around prior distributions. The choice of a prior $f(\boldsymbol{\theta})$ may affect Bayesian estimation. A strong prior may have an influence on Bayesian inference. In the absence of any prior information, one may use a non-informative prior: $f(\boldsymbol{\theta}) \propto 1$. Recall that the likelihood $L(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta})$. So we have

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto L(\boldsymbol{\theta}|\boldsymbol{y}) f(\boldsymbol{\theta}).$$

This therefore, shows the link between Bayesian and likelihood based methods, in particular when a non-informative prior is used, Bayesian inference reduces to likelihood inference.

In choosing a prior distribution, if there is no inherent reason to prefer one prior distribution over another, for simplicity a conjugate prior is chosen. A conjugate prior is a prior distribution that when used the resulting posterior distribution also belongs to the same family of distributions. Since Bayesian inference depends on the posterior distribution this is important. The prior distribution $f(\boldsymbol{\theta})$ is said to be conjugate to $f(\boldsymbol{y}|\boldsymbol{\theta})$ if the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ is in the same family as the prior distribution $f(\boldsymbol{\theta})$. For instance, the normal (Gaussian family) is conjugate to itself, that is to say if a normal prior distribution is used, the resulting posterior distribution is also normal. All members of the exponential family have conjugate priors.

In regression models, the multivariate normal distribution is typically chosen as a prior distribution for the mean parameters $\boldsymbol{\beta}$, that is we assume that

$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ are hyper-parameters. Hyper-parameters may also be assumed unknown and may lead to adding another hierarchy in the model specification, however, in practice this is often the last stage. If a non-informative prior is desired, we can choose $\boldsymbol{\Sigma}_0^{-1} \approx \mathbf{0}$ or $\boldsymbol{\beta}_0 \sim U(-\infty, \infty)$. Typically, for the variance-covariance matrices the Wishart distribution is used as a prior distribution.

The Wishart distribution is a generalisation of two distributions, namely the χ^2 distribution and the gamma distribution to multiple dimensions. The Wishart distribution is useful in the estimation of covariance matrices. Let $\boldsymbol{Z}$ be a $n \times p$ matrix, with the ith row $\boldsymbol{z}_i \sim N_p(\mathbf{0}, \boldsymbol{V})$ independently, where $\boldsymbol{V}$ is a $p \times p$ covariance matrix which is positive definite. Then, the probability distribution of

$$\boldsymbol{W} = \boldsymbol{Z}^T \boldsymbol{Z} \tag{3.4}$$

has a Wishart distribution with degrees of freedom n, denoted by $W_p(\boldsymbol{V}, n)$ and whose density function is given by

$$f(\boldsymbol{W}) = \frac{|\boldsymbol{W}|^{(n-p-1)/2}}{2^{np/2}|\boldsymbol{V}|^{n/2}\Gamma_p(\frac{n}{2})} \exp\left(-\frac{1}{2} tr(\boldsymbol{V}^{-1}\boldsymbol{W})\right), \tag{3.5}$$

where $\boldsymbol{W} > 0$ (positive definite) and $\Gamma_p(.)$ is the multivariate gamma function defined as

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma((n+1-j)/2). \tag{3.6}$$

The Wishart distribution $W_p(\boldsymbol{V}, n)$ has mean $n\boldsymbol{V}$ and mode $(n-p-1)\boldsymbol{V}$ for $n \geq p+1$. In the case when $p = 1$ and $\boldsymbol{V} = 1$, the Wishart distribution $W_p(\boldsymbol{V}, n)$ coincide with the univariate χ^2 distribution. The Wishart distribution is the distribution of the maximum likelihood estimate (MLE) for the covariance matrix in a multivariate normal distribution.

A conjugate prior for the covariance matrix of a multivariate normal distribution is the inverse Wishart distribution which is defined below. Suppose that a $p \times p$ random matrix $\boldsymbol{A} \sim W_p(\boldsymbol{V}, n)$, then $\boldsymbol{B} = \boldsymbol{A}^{-1}$ has an inverse Wishart distribution denoted by $W_p^{-1}(\boldsymbol{V}^{-1}, n)$ or $IW_p(\boldsymbol{V}^{-1}, n)$ whose probability density function is give as;

$$f(\boldsymbol{B}) = \frac{|\boldsymbol{V}|^{-n/2}|\boldsymbol{B}|^{-(n+p+1)/2} exp\left(-tr(\boldsymbol{V}^{-1}\boldsymbol{B}^{-1})/2\right)}{2^{np/2}\Gamma_p(\frac{n}{2})}. \tag{3.7}$$

The mean of $W_p^{-1}(\boldsymbol{V}^{-1}, n)$ is given by

$$E(\boldsymbol{B}) = \boldsymbol{V}^{-1}/(n-p-1).$$

Let $\boldsymbol{X} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)$, with $\boldsymbol{x}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. If we assume a prior distribution $\boldsymbol{\Sigma} \sim W_p^{-1}(\Phi, m)$, then the posterior distribution is given by

$$\boldsymbol{\Sigma}|\boldsymbol{X} \sim W_p^{-1}(\boldsymbol{X}\boldsymbol{X}^T + \Phi, m+n).$$

When $p = 1$, the inverse Wishart distribution reduces to an inverse gamma distribution.

3.3 Estimation of Parameters: MCMC Methods

The two most popular MCMC algorithms are the Metropolis Hastings and the Gibbs samplers. These two algorithms have many variants and extensions that have been developed. These forms and extensions are more advanced and sometimes more specific to some problems. In this section we discuss in detail the Metropolis Hastings and the Gibbs samplers together with their invariants and extensions.

3.3.1 The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm was developed from the Metropolis algorithm by Hastings in 1970. Let $\boldsymbol{x}$ be the vector of parameters and also let $f(\boldsymbol{x})$ be the target distribution from which we wish to generate a sample size T ($\boldsymbol{x}$ can be replaced with the parameter of interest say θ). We can describe the metropolis-Hastings algorithm in the following iterative steps:

- Set initial values of $\boldsymbol{x}^{(0)}$.
- For $t = 1, 2, \ldots, T$ repeat the following steps;
  a. Set $\boldsymbol{x} = \boldsymbol{x}^{(t-1)}$.
  b. Generate new candidate values $\boldsymbol{x}'$ from a proposal distribution

$q(\boldsymbol{x} \rightarrow \boldsymbol{x}') = q(\boldsymbol{x}'|\boldsymbol{x})$

c. Calculate

$$\alpha = min\left(1, \frac{f(\boldsymbol{x}')q(\boldsymbol{x}|\boldsymbol{x}')}{f(\boldsymbol{x})q(\boldsymbol{x}'|\boldsymbol{x})}\right) \tag{3.8}$$

d. Update $\boldsymbol{x}^{(t)} = \boldsymbol{x}'$ with probability α and $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)}$ with probability $1 - \alpha$.

Regardless of the proposal distribution chosen, the Metropolis-Hastings algorithm will converge to the target distribution. In practice, however, the proposal distribution should be chosen with care as poor choices may delay convergence to the target distribution.

In the Bayesian framework the above algorithm can be implemented by replacing $\boldsymbol{x}$ with the parameters of interest $\boldsymbol{\theta}$ and the target distribution by the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y}.)$ In terms of $\boldsymbol{\theta}$ and $f(\boldsymbol{\theta}|\boldsymbol{y})$ the above algorithm can be given as;

- Set initial values of $\boldsymbol{\theta}^{(0)}$.

- For $t = 1, 2, \ldots, T$ repeat the following steps;

  a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$.

  b. Generate new candidate values $\boldsymbol{\theta}'$ from a proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$

  c. Calculate

$$\alpha = min\left(1, \frac{f(\boldsymbol{\theta}'|\boldsymbol{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right) \tag{3.9}$$

d. Update $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$ with probability α and $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ with probability $1 - \alpha$.

A notable characteristics of this algorithm is that there is no need to evaluate the normalising constant $f(\boldsymbol{y})$ involved in $f(\boldsymbol{\theta}|\boldsymbol{y})$ since it cancels out in the α.

3.3.1.1 The Random-Walk Metropolis

The original Metropolis algorithm considered only symmetric proposals of the type $q(\boldsymbol{\theta}|\boldsymbol{\theta}')$. The random walk Metropolis is a special case with $q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = q(|\boldsymbol{\theta}' - \boldsymbol{\theta}|)$. Both the original Metropolis algorithm and the random walk Metropolis result in an acceptance probability that depends only on the posterior or target distribution

$$\alpha = min\left(1, \frac{f(\boldsymbol{\theta}'|\boldsymbol{y})}{f(\boldsymbol{\theta}|\boldsymbol{y})}\right) = min\left(1, \frac{f(y|\boldsymbol{\theta}')f(\boldsymbol{\theta}')}{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}\right) \tag{3.10}$$

A usual proposal of this type is a multivariate normal $q(\boldsymbol{\theta}'|\boldsymbol{\theta}) \sim N_d(\boldsymbol{\theta}, \boldsymbol{C})$, where d is the dimension of $\boldsymbol{\theta}$. The covariance matrix $\boldsymbol{C}$ controls the convergence speed of the algorithm.

3.3.2 The Gibbs Sampler

This sampler is a special case of the Metropolis-Hastings algorithm using as proposal density $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t)})$ the full conditional posterior distribution $f(\theta_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y})$,

where $\boldsymbol{\theta}_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d)^T$. These proposal distributions result in acceptance probability $\alpha = 1$, making the proposed move to be accepted in all iterations. Despite being a special case of the Metropolis-Hastings algorithm, the Gibbs sampler is usually cited as a separate simulation technique because of its popularity and convenience. In each step in a Gibbs sampler, random values must be generated from unidimensional distributions for which a variety of computational tools exist. In most cases these conditional distributions have a known form hence random numbers can easily be simulated using standard functions in statistical and computing software such as BUGS. The Gibbs sampler always move to new values and does not require specification of proposal distribution. Its disadvantage is its ineffectiveness when the parameter space is complicated or when the parameters are highly correlated.

The Gibbs algorithm can be summarised as follows:

1. Set initial values $\theta^{(0)}$.

2. For $t = 1, 2, \ldots, T$ repeat the following steps;

- Set $\theta = \theta^{(t-1)}$.
- For $j = 1, 2, \ldots, d$, update θ_j from $\theta_j \sim f(\theta_j | \boldsymbol{\theta}_{-j}, \boldsymbol{y})$.
- Set $\theta^{(t)} = \theta$ and save it as the generated set of values at $t + 1$ iteration

of the algorithm.

Therefore, given a particular state of the chain $\theta^{(t)}$, we generate the parameter values by

$\theta_1^{(t)}$ from $f(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_p^{(t-1)}, \boldsymbol{y})$

$\theta_2^{(t)}$ from $f(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_p^{(t-1)}, \boldsymbol{y})$

$\theta_3^{(t)}$ from $f(\theta_3|\theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)}, \ldots, \theta_p^{(t-1)}, \boldsymbol{y})$

$\vdots$

$\theta_j^{(t)}$ from $f(\theta_j|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)} \ldots, \theta_p^{(t-1)}, \boldsymbol{y})$

$\vdots$

$\theta_p^{(t)}$ from $f(\theta_p|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{p-1}^{(t)}, \boldsymbol{y})$

Generating values from $\theta_j^{(t)}$ from $f(\theta_j|\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)} \ldots, \theta_p^{(t-1)}, \boldsymbol{y})$ is not complicated since it is a univariate distribution and can be written as $f(\theta_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y}) \propto f(\boldsymbol{\theta}|\boldsymbol{y})$, where all the variables except θ_j are held constant at their given values [34].

The Slice Gibbs Sampler

This sampler is based on Gibbs sampling. When the conditional posterior distribution do not have a convenient form the slice sampler is used. This method augments the parameter space by adding a set of convenient random

variables known as auxiliary variables that retain the marginal posterior distribution of interest unchanged but convert all conditionals to distributions of standard form. After the augmentation has been performed, then the Gibbs sampler can be applied. Due to the fact that this method uses auxiliary variables, it is often referred to as auxiliary variables method. The idea behind this method is summarised as follows. Consider a target distribution $g(x)$ that is difficult to generate from. Introduce a new variable u with the conditional $f(u|x)$, then the joint distribution can be written as

$$f(u, x) = f(u|x)g(x), \tag{3.11}$$

while the marginal distribution is equal to the marginal target distribution $g(x)$ since

$$f(x) = \int f(u, x)du = \int f(u|x)g(x)du = g(x). \tag{3.12}$$

The Gibbs sampler can be used which generates values from the joint distribution $f(u, x)$ and the corresponding marginals $f(u)$ and f(x):

1. Generate $u \sim f(u|x)$.

2. Generate $x \sim f(u|x)g(x)$.

Since $f(u|x)$ is involved in both steps, then it must be specified in a way that both $f(u|x)$ and $f(u|x)g(x)$ are convenient in terms of simulation. Common practice is to use the uniform distribution $U(0, g(x))$ for $f(u|x)$

$$f(u, x) = \frac{1}{g(x)}g(x)I(0 < u < g(x)) = I(0 < u < g(x)), \tag{3.13}$$

and

$$f(x) = \int I(0 < u < g(x))du = \int_0^{g(x)} du = [u]_0^{g(x)} = g(x), \tag{3.14}$$

where $I(x)$ is the indicator function taking value equal to one if x is true and zero otherwise. Then the Gibbs steps above becomes

1. Generate $u^{(t)} \sim U(0, g(x^{(t-1)}))$.

2. Generate $x^{(t)} \sim U(x : 0 \leq u \leq g(x))$.

It usual practice in the Bayesian context to facilitate $\boldsymbol{u} = (u_1, \ldots, u_n)$ auxiliary variables coming from the uniform distribution defined within the interval zero to the likelihood ordinate $f(y_i|\boldsymbol{\theta})$. Hence the joint distribution will be given by

$$f(\boldsymbol{\theta}, \boldsymbol{u}|\boldsymbol{y}) \propto \left\{ \prod_{i=1}^{n} I(0 \leq u_i \leq f(y_i|\boldsymbol{\theta})) \right\} f(\boldsymbol{\theta}),$$

resulting in a Gibbs sampler of the type;

1. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$.

2. For $i = 1, \ldots, n$, generate $u^{(t)} \sim U(0, f(y_i|\boldsymbol{\theta}))$.

3. For $j = 1, \ldots, d$, update $\theta_j \sim f(\theta_j) \prod_{i=1}^{n} I(0 \leq u_i^{(t)} \leq f(y_i|\boldsymbol{\theta})))$.

4. Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$.

The above scheme can easily be implemented in a wide range of popular statistical models including generalised linear models. This sampler avoids

the specification of the proposal densities embedded in Metropolis-Hastings algorithm. The main advantage of this scheme is that once the appropriate augmenting scheme has been found, the algorithm becomes directly applicable in all data sets without computational difficulties. Its setback is that the resulting chain is usually highly autocorrelated.

Metropolis Within Gibbs

Many available computational tools for generating random values from univariate distributions allow us to implement Gibbs sampling in a variety of cases, even when the resulting posterior is cumbersome. It is convenient, on some instances, to use the Metropolis-Hastings steps to generate from these univariate conditional posterior distributions. This approach is known as the Metropolis within Gibbs algorithm. This approach, however, is equivalent to a componentwise Metropolis-Hastings algorithm where some components of the parameter vector are directly generated from the corresponding full conditional posterior distribution.

The Componentwise Metropolis Algorithm

This sampler involves dividing the parameter vector $\boldsymbol{\theta}$ into subvectors which are updated sequencially using the Metroplis-Hastings algorithm. The Gibbs sampler is a special case of this algorithm hence this algorithm is often re-

ferred to as the Metropolis within Gibbs. When univariate components are updated sequencially, the method is called a single component Metropolis-Hastings algorithm.

We can summarise the algorithm as follows;

a. Set initial values $\boldsymbol{\theta}^{(0)}$.

b. For $t = 1, \ldots, T$ repeat the following steps:

- Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$.
- For $j = 1, \ldots, d$,

1. Generate new candidate values for $\theta_j^{'}$ for the *jth* component of vector $\boldsymbol{\theta}_j^{'}$ from the proposal distribution $q(\theta_j^{'}|\boldsymbol{\theta})$.

2. Calculate

$$\alpha = min\left(1, \frac{f(\theta_j^{'}|\boldsymbol{\theta}_{-j}, \boldsymbol{y})q(\theta_j|\theta_j^{'}, \boldsymbol{\theta}_{-j})}{f(\theta_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y})q(\theta_j|\theta_j, \boldsymbol{\theta}_{-j})}\right) = min\left(1, \frac{f(\boldsymbol{y}|\theta_j^{'}, \boldsymbol{\theta}_{-j})f(\theta_j^{'}, \boldsymbol{\theta}_{-j})q(\theta_j|\theta_j^{'}, \boldsymbol{\theta}_{-j})}{f(\boldsymbol{y}|\theta_j^{'}, \boldsymbol{\theta}_{-j})f(\theta_j^{'}, \boldsymbol{\theta}_{-j})q(\theta_j^{'}|\theta_j, \boldsymbol{\theta}_{-j})}\right), \tag{3.15}$$

where $\boldsymbol{\theta}_{-j}$ is the vector θ excluding the *jth* component.

3. update $\theta_j = \theta_j^{'}$ with probability α.

- Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$.

One generated observation $\boldsymbol{\theta}^{(t)}$ is obtained after updating all components of the parameter vector. The sequential updating of the elements of $\boldsymbol{\theta}$ does

not influence the convergence of the algorithm. To ensure randomness, random selection of the updating sequence may also be used, this is referred to as a random scan.

The main advantage of this MCMC scheme is that the sampler is decomposed into several univariate steps in which random number generation is usually straightforward. Its disadvantage is that convergence cannot be accelerated by using multivariate densities with appropriate correlation structures. Parameter blocking can be used to speed up convergence. In parameter blocking, the parameter vector is divided into subvectors with correlated elements called blocks and each block is updated in a separate Metropolis step.

3.3.3 Convergence Diagnostics

These are checks used to determine whether the algorithm has reached its equilibrium or target distribution. There are several ways used to monitor convergence. The simplest way is to monitor the Monte Carlo (MC) error. Small values of the MC error indicate that the quantity of interest has been calculated with precision.

There are two common ways used to estimate the MC error; the batch mean method and the window estimator method. The first is simple to

implement and is widely used, the second is hard to implement but precise.

Suppose that we have T iterations and after discarding (burn in) some we remain with $T^{'}$. The MC error using the batch mean method is calculated by first partitioning the resulting output into K batches, usually $K = 30$ or $K = 50$. Both the number of batches and the sample size $v = T^{'}/K$ must be sufficiently large so as to enable consistence in estimation and also to eliminate autocorrelations. In calculating the MC error of the posterior mean $\overline{G(\mathbf{x})}$ we begin by calculating batch means denoted by $\overline{G(\mathbf{x})}_b$;

$$\overline{G(\mathbf{x})}_b = \frac{1}{v} \sum_{t=(b-1)v+1}^{bv} G(\mathbf{x}^{(t)}) \tag{3.16}$$

for each batch $b = 1, 2, \ldots, K$ and the overall sample mean given by

$$\overline{G(\mathbf{x})} = \frac{1}{K} \sum_{b=1}^{K} \overline{G(\mathbf{x})}_b, \tag{3.17}$$

assuming that we keep $\mathbf{x}^1, \ldots, \mathbf{x}^{T^{'}}$ observations. Then the MC error is given by

$$MCE[G(\mathbf{x})] = SE[\overline{G(\mathbf{x})}] = \sqrt{\frac{1}{K(K-1)} \sum_{b=1}^{K} (G(\mathbf{x}) - \overline{G(\mathbf{x})_b})^2}. \tag{3.18}$$

The window estimator method is based on the expression of variance in autocorrelated samples

$$MCE[G(\mathbf{x})] = \frac{SD[G(\mathbf{x})]}{\sqrt{T'}} \sqrt{1 + 2 \sum_{k=1}^{\infty} \widehat{\rho}_k[G(\mathbf{x})]}, \tag{3.19}$$

where $\widehat{\rho}_k[G(\mathbf{x})]$ is the estimated autocorrelation of lag k. For large values of k, the autocorrelation cannot be reliably estimated due to the small number

of remaining observations. In practice, however, the autocorrelations will be close to zero for sufficiently large k. In this method we identify a window w after which autocorrelations are considerably low and discard ρ_k with $k > w$ from the preceding MC error estimate. Hence (3.19) is modified as

$$MCE[G(\mathbf{x})] = \frac{SD[G(\mathbf{x})]}{\sqrt{T'}} \sqrt{1 + 2\sum_{k=1}^{w} \widehat{\rho}_k[G(\mathbf{x})]}. \tag{3.20}$$

The Raftery and Lewis Diagnostic

Suppose we want to measure some posterior quantile of interest. Let q denote the quantile of interest we want to measure, r be the acceptance tolerance and s be the probability of being within the acceptance tolerance, then the Raftery and Lewis diagnostic calculates the number of iterations N necessary to satisfy the specified conditions. This diagnostic tool was designed to test the number of iterations and burn-in needed by first running and taking a shorter pilot chain. In practice, we can also just test our normal chain to see if it satisfies the results that the diagnostic suggests.

This diagnostic involves the following steps;

- Select a posterior quantile of interest q (eg. the 0.025 quantile).
- Select an acceptable tolerance r for this quantile, for example, $r = 0.005$ which means we want to measure the 0.025 quantile with an accuracy of ± 0.005.

- Select a probability s, which is the desired probability of being within $(q-r, q+r)$.

- Run a pilot sample to generate a Markov chain of minimum length by rounding up

$$n_{min} = \left[\Phi^{-1} \left(\frac{s+1}{2} \right) \frac{\sqrt{q(1-q)}}{r} \right]^2,$$

where Φ^{-1} is the inverse of the normal CDF.

Other than using the MC error and the Raftery and Lewis diagnostics, graphical diagnostic methods (virtual inspection) can be used to monitor convergence. Trace plots can be used. These are plots of the iterations against the generated values. If all values are within a zone without strong periodicities and tendencies then we assume convergence. Other graphical methods include the plot of autocorrelations and ergodic mean[34].

3.4 Integrated Laplace Approximation (INLA)

INLA is a computational tool for Bayesian inference which has been introduced by Rue and Martino [37] and Rue et al [38]. This recently introduced approach does Bayesian inference in the broad class of latent Gaussian models. The Gaussian models are models of an outcome variable y_i that assume independence conditional on some undelying latent field $\boldsymbol{\xi}$ and a vector of

hyperparameters $\boldsymbol{\theta}$. INLA is an alternative estimation method to the widely used Markov Chain Monte Carlo (MCMC) methods. This approach directly approximates the posteriors of interest with a close form expression unlike the MCMC methods which samples from the posterior distribution.

The main aim of INLA approach is to approximate the marginal posteriors for the latent variables as well as for the hyperparameters of the Gaussian latent model given by

$$f(\xi_i|\boldsymbol{y}) = \int f(\xi_i|\boldsymbol{\theta}, \boldsymbol{y}) f(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{3.21}$$

$$f(\boldsymbol{\theta}_j|\boldsymbol{y}) = \int f(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{3.22}$$

The basis of this approximation is on the combination of Laplace approximation to the full conditional $f(\boldsymbol{\theta}|\boldsymbol{y})$ and $f(\xi_i|\boldsymbol{\theta}, \boldsymbol{y})$, for $i = 1, 2, \ldots, n$ and numerical integration routines to integrate out the hyperparameters $\boldsymbol{\theta}$. To obtain the marginal posteriors in (3.21) and (3.22), the approximation involves three steps. The first step is to approximate the full posterior $f(\boldsymbol{\theta}|\boldsymbol{y})$. To do this, we approximate $f(\boldsymbol{\xi}|\boldsymbol{y}, \boldsymbol{\theta})$ by a multivariate Gaussian density $\tilde{f}_G(\boldsymbol{\xi}|\boldsymbol{y}, \boldsymbol{\theta})$ evaluated at its mode. Then the posterior density of $\boldsymbol{\theta}$ is approximated by using the Laplace approximation

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{f(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{f}_G(\boldsymbol{\xi}|\boldsymbol{\theta}, \boldsymbol{y})}|_{\boldsymbol{\xi}=\boldsymbol{\xi}^*(\boldsymbol{\theta})}, \tag{3.23}$$

where $\boldsymbol{\xi}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\boldsymbol{\xi}$ for a given $\boldsymbol{\theta}$. There

is no close form solution available of $\boldsymbol{\xi}^*(\boldsymbol{\theta})$, a numerical procedure like the Newton-Raphson algorithm can be used.

In the second step, we compute the Laplace approximation of the full conditionals $f(\xi_i|\boldsymbol{y},\boldsymbol{\theta})$ for selected values of $\boldsymbol{\theta}$. These selected values of $\boldsymbol{\theta}$ must be chosen carefully as they will be used as evaluation points in the numerical integration applied to obtain the posterior marginals of ξ_i in (3.21) and (3.22). The density $f(\xi_i|\boldsymbol{y},\boldsymbol{\theta})$ is approximated using Laplace approximation defined by;

$$f_{LA}(\xi_i|\boldsymbol{\theta},\boldsymbol{y}) \propto \frac{f(\boldsymbol{\xi},\boldsymbol{\theta},\boldsymbol{y})}{\tilde{f}_G(\boldsymbol{\xi}_{-i}|\xi_i,\boldsymbol{\theta},\boldsymbol{y})}|_{\boldsymbol{\xi}_{-i}=\boldsymbol{\xi}^*_{-i}(\xi_i,\boldsymbol{\theta})}, \tag{3.24}$$

where ξ_{-1} denotes the vector $\boldsymbol{\xi}$ with the ith component omitted, $\tilde{f}_G(\boldsymbol{\xi}_{-i}|\xi_i,\boldsymbol{\theta},\boldsymbol{y})$ is the Gaussian approximation of $f(\boldsymbol{\xi}_{-i}|\xi_i,\boldsymbol{\theta},\boldsymbol{y})$, treating ξ_i as fixed and $\boldsymbol{\xi}^*_{-i}(\xi_i,\boldsymbol{\theta})$ is the mode of $f(\boldsymbol{\xi}_{-1}|\xi_i,\boldsymbol{\theta},\boldsymbol{y})$.

The third step involves combining the two full posteriors obtained in the previous steps and the marginal densities of ξ_i and θ_j obtained by integrating out the relevant terms. The approximation for the marginal of the latent variables can be obtained by the expression

$$f(\xi_i|\boldsymbol{y}) = \int f(\xi_i|\boldsymbol{y},\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{y}) d\theta \approx \sum_k \tilde{f}(\xi_i|\boldsymbol{\theta}_k,\boldsymbol{y}) \tilde{f}(\boldsymbol{\theta}_k|\boldsymbol{y}) \Delta_k \tag{3.25}$$

which is evaluated using numerical integration on a set of grid points for $\boldsymbol{\theta}$, with area weights Δ_k for $k = 1, 2, \ldots, K$.

3.5 Application

3.5.1 Results - MCMC

In analysing the SDHS data using the Bayesian GLMMs, separate analyses were done for the males and females data sets. This was done since the HIV prevalence rates in the two groups differ significantly. In analysing the data R for windows software was used. In R, the MCMCglmm package [17] was used. This package uses the Gibbs sampler. A total number of 60000 iterations were run in both groups. Before the 60000 iterations were used, the Raftery and Lewis diagnostic was performed on a shorter pilot chain. The results of the diagnostic were that about 20000 iterations will be needed for convergence for both males and females. In both males and females, the specified conditions were $p = 0.025$, $r = 0.005$ and $s = 0.95$. Several values far bigger than the one given by the Raftery and Lewis diagnostic were tried while monitoring the trace trace plots and finally 60000 iterations were chosen.

As discussed in previous sections, in Bayesian inference we use prior distributions for parameters. In this analysis, the multivariate prior distribution was used for the mean parameters (fixed effects) and an inverse Wishart prior distribution with a Cauchy parameter expansion was used for the variances of random effects. Since in GLMMs, the residual variance cannot be estimated

from the data, we fixed the residual variance at some value (we used 1). In fitting this model, the logit link function was used for the response variable and random effects or errors at cluster level were fitted. Model (MCMC) convergence was monitored using trace plots. The results of the analysis are shown in Table 3.1.

Table 3.1: Bayesian GLMM results for HIV prevalence

	Males		Females	
Variable	OR (SE)	95%CI	OR (SE)	95%CI
Age				
15-19	1.00	ref	1.00	ref
20-24	4.51 (0.30)	2.51 -8.13	4.75 (0.16)	3.53 -6.51
25-29	10.43 (0.30)	5.72 -18.74	8.21 (0.17)	5.87 -11.51
30-34	25.01 (0.34)	12.81 -48.40	6.13 (0.19)	4.34 -8.92
35-39	29.23 (0.34)	15.30 -56.08	3.82 (0.18)	2.59 -5.60
40-44	21.30 (0.36)	10.30 -42.27	2.10 (0.21)	1.42 -3.21
44-49	11.02 (0.36)	5.38 -22.22	1.05 (0.21)	0.67 -1.66
Region				
Hhohho	1.00	ref	1.00	ref
Manzini	0.76 (0.17)	0.54 -1.05	0.88 (0.12)	0.68 -1.12
Shiselweni	0.78 (0.19)	0.54 -1.14	0.94 (0.13)	0.73 -1.22
Lubombo	0.76 (0.17)	0.54 -1.06	0.89 (0.13)	0.69 -1.15
Residence type				
urban	1.00	ref	1.00	ref
rural	0.87 (0.13)	0.67 -1.13	0.75 (0.10)	0.61 -0.92
Marital status				
never married	1.00	ref	1.00	ref
married	1.18 (0.18)	0.83 -1.66	0.54 (0.12)	0.43 -0.68
living together	2.80 (0.22)	1.82 -4.27	1.22 (0.14)	0.91 -1.64
widowed	8.78 (0.42)	3.92 -20.02	3.50 (0.20)	2.34 -5.20
divorced	2.10 (0.50)	0.78 -5.56	2.85 (0.66)	0.79 -10.46
not living together	3.45 (0.29)	1.96 -6.20	2.08 (0.24)	1.30 -3.37
Educational level				
no education	1.00	ref	1.00	ref
primary	1.10 (0.21)	0.73 -1.66	0.98 (0.15)	0.73 -1.33
secondary	0.89 (0.20)	0.60 -1.32	0.64 (0.16)	0.47 -0.88
Higher	0.48 (0.27)	0.29 -0.82	0.36 (0.23)	0.23 -0.56
Age at first sex				
never had sex	1.00	ref	1.00	ref
14 or less	1.06 (0.45)	0.43 -2.58	6.96 (0.27)	4.05 -11.73
15-17	2.18 (0.35)	1.10 -4.34	6.16 (0.24)	3.85 -9.80
18+	1.45 (0.35)	0.74 -2.95	5.45 (0.24)	3.35 -8.70
Had an STI in last 12 months				
no	1.00	ref	1.00	ref
yes	2.88 (0.20)	1.98 -4.28	2.34 (0.18)	1.66 -3.32
dont know	1.04 (0.71)	0.25 -4.15	0.22 (0.77)	0.04 -0.85
Number of sexual partners				
no partner	1.00	ref	1.00	ref
1 partner	1.58 (0.23)	1.02 -2.53	1.16 (0.37)	0.88 -1.52
2+ partners	2.89 (0.25)	1.78 -4.66	1.85 (0.35)	0.93 -3.58
Number of times away from home				
none	1.00	ref	1.00	ref
1-5 times	1.12 (0.16)	0.82 -1.55	0.83 (0.11)	0.67 -1.03
6-10 times	1.54 (0.18)	1.07 -2.23	0.94 (0.14)	0.72 -1.22
11-15 times	1.16 (0.21)	0.76 -1.74	0.87 (0.15)	0.65 -1.18
16+ times	0.82 (0.18)	0.58 -1.17	1.12 (0.14)	0.85 -1.47
Any births				
no			1.00	ref
Yes			0.80(0.10)	0.66-0.96
Random effects				
$\sigma_{cluster}$	0.39(0.10)		0.30(0.05)	

In both men and women, significant association was found between HIV

serostatus and age. Women aged 20-24 years are at a significantly higher probability (OR: 8.21; 95% CI: 5.87-11.51) of being HIV positive than women in the reference category 15-19 years. Men at higher probability (OR: 29.23; 95% CI: 15.30-56.08) of being HIV positive are those in the age group 35-39. For both men and women, there is an increase in prevalence odds as the age increases, then followed by a decline. No significant association was found between region and HIV serostatus for the two gender groups. There were no significant differences found between men in the rural and urban areas as regards to HIV serostatus. Women in the rural areas are less likely to be HIV positive (OR: 0.75; 95% CI: 0.61-0.92) than those in urban areas.

Significant association between HIV and highest educational level was found in both males and females for those in the higher education category. Also, for women, significant association was found between HIV serostatus and those classified as having secondary level of education (OR: 0.64; 95% CI: 0.47-0.88). Although the prevalence odds of HIV for those with primary level education is higher in both males and females, they are not significantly different than those with no education.

Men who are widowed were found to be almost 9 times more likely to be HIV infected than those who were never married, similarly, women who are widowed were found to be 3.5 times more likely to be HIV positive. Being widowed, in both gender groups increases the odds of HIV infection. Being

married reduces the odds of HIV infection in females. Also, the category "not living together" was found to be associated with a higher probability of being HIV positive in both groups compared to those never married. There is a higher probability of HIV infection (OR: 6.96; 95% CI: 4.05-11.73) for women who started having sex at the age of 14 or less than those who had never had sex. The association between HIV serostatus and age at first sex was found to be strong in women compared to men.

The number of sexual partners in the past year have a significant relationship with HIV serostatus in men. The odds of infection increases with an increase in the number of partners. Those who had 1 partner in the past year were 1.58 times more likely to be HIV positive than those with none while those who had two or more partners in the past year were 2.89 times more likely to be infected with HIV than those who had no sexual partners in the past year. Men who had an STI in the past 12 months were more likely (OR: 2.88; 95% CI: 1.98-4.28) to be HIV postive than those who had no STI. Similar results were obtained for the females. Not knowing if one had an STI in the past 12 months was found to be significanltly associated with HIV risk in women but it was not the case with men. The number of sexual partners and exposure to STIs play a significant role in the generation of new infections or incidences.

Weak association between the number of times away from home in the last

12 months and HIV serostatus was found in both men and women. Among women, reporting a birth in the past 5 years was found to be significantly related to ones HIV serostatus (OR: 0.80; 95% CI: 0.66-0.96). In interpreting the estimates of the random effects we use the divide by 4 rule from Gelman and Hill [10]. For males, the cluster level has an estimated standard error of 0.15 on the logit scale. The differences between clusters with regards to HIV serostatus is approximately $\pm 3.7\%$ on the probability scale, that is, over and above the differences explained by the other covariates. Similar interpretation holds for the females. The differences among clusters are approximately $\pm 2.3\%$ on the probability scale for the females.

3.5.2 Results - INLA

Table 3.2 shows the results produced by the INLA approach to full Bayesian inference. In this analysis, we used similar priors to the ones used in the MCMCglmm.

The two Bayesian estimation approaches lead to the same inference. Variables that are not statistically significant in the MCMC approach are also not significant with the INLA approach. However, estimates from these two methods of estimation are not exactly the same. The two approaches differ in the following ways. The first outstanding difference between the two ap-

proaches is computational speed. In running the model in INLA, the results were obtained in less than a minute whilst with the MCMC the results were obtained after more than 20 minutes. Negligible differences were observed in standard errors.

Table 3.2: Bayesian GLMM results using INLA for HIV prevalence

	Males		Females	
Variable	OR (SE)	95%CI	OR (SE)	95%CI
Age				
15-19	1.00	ref	1.00	ref
20-24	4.21 (0.29)	2.44-7.59	3.91 (0.13)	3.01-5.11
25-29	8.85 (0.30)	5.02-16.32	6.22 (0.15)	4.66-8.35
30-34	19.29 (0.31)	10.70-36.31	4.86 (0.16)	3.59-6.62
35-39	22.20 (0.32)	12.10-42.41	3.20 (0.17)	2.32-4.45
40-44	16.53 (0.33)	8.72-32.52	1.94 (0.18)	1.37-2.76
45-49	9.53 (0.34)	4.95-18.96	1.06 (0.20)	0.72-1.57
Region				
Hhohho	1.00	ref	1.00	ref
Manzini	0.77 (0.16)	0.56-1.05	0.89 (0.12)	0.70-1.11
Shiselweni	0.81 (0.18)	0.57-1.14	0.95 (0.13)	0.74-1.22
Lubombo	0.78 (0.16)	0.57-1.08	0.90 (0.12)	0.71-1.15
Residence type				
urban	1.00	ref	1.00	ref
rural	0.90 (0.13)	0.70-1.15	0.78 (0.10)	0.65-0.95
Marital status				
never married	1.00	ref	1.00	ref
married	1.12 (0.15)	0.84-1.49	0.59 (0.10)	0.49-0.72
living together	2.41 (0.18)	1.68-3.46	1.19 (0.12)	0.93-1.52
widowed	6.28 (0.36)	3.16-13.10	2.93 (0.17)	2.11-4.10
divorced	1.96 (0.43)	0.84-4.53	2.38 (0.55)	0.81-7.00
not living together	2.82 (0.25)	1.73-4.61	1.85 (0.21)	1.24-2.78
Educational level				
no education	1.00	ref	1.00	ref
primary	1.10 (0.17)	0.79-1.54	0.98 (0.13)	0.76-1.27
secondary	0.93 (0.17)	0.66-1.30	0.69 (0.13)	0.53-0.90
higher	0.56 (0.23)	0.36-0.87	0.42 (0.19)	0.29-0.61
Age at first sex				
never had sex	1.00	ref	1.00	ref
14 yrs or less	1.14 (0.41)	0.50-2.53	5.60 (0.23)	3.56-8.91
15-17 yrs	2.08 (0.32)	1.11-3.97	5.07 (0.21)	3.38-7.72
18+ yrs	1.45 (0.32)	0.78-2.77	4.53 (0.21)	3.01-6.94
Had an STI in last 12 months				
no	1.00	ref	1.00	ref
yes	2.50 (0.17)	1.79-3.50	2.04 (0.15)	1.53-2.74
don't know	1.14 (0.59)	0.33-3.40	0.27 (0.65)	0.06-0.84
Number of sexual partners				
no partner	1.00	ref	1.00	ref
1 partner	1.52 (0.20)	1.03-2.30	1.13 (0.12)	0.90-1.41
2+partners	2.50 (0.22)	1.63-3.91	1.69 (0.28)	0.97-2.96
Number of times away from home				
none	1.00	ref	1.00	ref
1-5 times	1.09 (0.14)	0.83-1.43	0.86 (0.09)	0.71-1.03
6-10 times	1.44 (0.15)	1.06-1.95	0.94 (0.11)	0.75-1.18
11-15 times	1.13 (0.18)	0.79-1.61	0.89 (0.13)	0.68-1.15
16+ times	0.84 (0.15)	0.62-1.13	1.09 (0.12)	0.86-1.37
Any births				
no			1.00	ref
Yes			0.83 (0.08)	0.70-0.97
Random effects				
$\sigma_{cluster}$	0.37(0.09)		0.29(0.06)	

3.6 Summary and Discussion

In this chapter, we have investigated the relationship between HIV serostatus and various explanatory variables using the SDHS data. In doing this, a Bayesian generalised linear mixed model was fitted to the data. Two estimation techniques are considered in this chapter; the MCMC methods and INLA. The results obtained in this analysis are discussed below.

Age proved to be a significant risk factor in both males and females. The odds of HIV infection increases with age in both groups, and later decreases after reaching the peak. This result is consistent with findings from other studies [14, 19, 27]. Contrary to the findings of [27, 19], this study found no relationship between region and the probability of HIV infection. Being widowed in both men and women was found to increase the odds of HIV infection. In this category the odds of HIV infection was higher in males than in females. This finding is in line with findings of previous studies [27, 33, 41]. Weak association was found between type of place of residence and HIV serostatus.

Though some studies [28, 46] have found some association between wealth and HIV risk, this study found wealth to be insignificant in explaining HIV serostatus. In both males and females, no significant relationship was found between wealth and the risk of HIV infection at the cross tabulation. Men

and women in the higher level of education are less likely to be HIV positive. This result is in line with the results of Johnson and Way [19].

Strong association between age at first sex and HIV risk was found in both males and females. The odds of HIV infection are higher in females than in males for the age category 14 years or less. This suggest that females start having sex at an earlier age than males. Starting having sexual intercourse early in females increases their vulnerability to HIV infection since in earlier ages their reproductive organs are not yet fully developed. Relationship between number of partners and the HIV risk was found only in males. The Swazi culture encourages or allows men to have multiple relationships, but on the contrary discourages women to have multiple relationships [40] hence the result is not surprising.

Among the biological factors, presence of an STI in the last 12 months was found to be significantly associated with the risk of HIV infection. This finding is in line with findings from previous studies [19, 3].

The two estimation methods namely the MCMC and INLA produced similar results. The differences between parameter estimates are minimal. In fact both methods lead to the same inference. The main advantage of INLA over the MCMC method is computational speed. As the INLA approach is still being developed, it promises to provide a good alternative to the widely used MCMC methods.

Chapter 4

Multilevel Models

In this chapter, we discuss multilevel models. Multilevel models are powerful and flexible extension to conventional regression frameworks. They extend the linear model and the generalised linear model by incorporating levels directlty into the model statement and by so doing account for aggregation present in the data. All of the familiar model forms for linear, dichotomous, count, restricted range, ordered categorical, and unordered outcomes are supplemented by adding a structural component. This structure classifies cases into known groups which may have their own set of explanatory variables at group level. Multilevel models allow the researcher to have group level explanatory variables and as well as individual level covariates in the same model.

4.1 Multilevel Data Structures

In the medical, social and biological sciences multilevel data structures are the norm and they are also encountered in many other application areas. A clear example of a multilevel data structure is school education where individuals (pupils) are subject to influences of grouping. In this example, pupils are nested in classes, classes are nested in schools, and schools are nested within school boards. A typical multilevel model in this case will assign pupils to level 1, classes to level 2, schools to level 3 and lastly school boards to level 4.

In a household survey, the level 1 units are individuals, level 2 units are households and the level 3 units are clusters. Such a hierarchy is often described in terms of clusters of level 1 units within each level 2 unit and also within each level 3 unit and this is commonly known as a clustered multilevel population. Other examples of multilevel structures include repeated measurement studies and medical trials. In medical trials, medical centres can be regarded as level 2 units and the individuals studied in each centre as level 1 units. In longitudinal studies, the individual from which the measurements are taken is regarded as the level 2 unit, and the measurements as level 1 units. Multilevel also permits estimation of effects at each level of the hierarchy.

4.1.1 Consequences of Ignoring Multilevel Structure

In all the above examples, elements in a cluster share common characteristics, that is observations from the same cluster tend to be more alike than observations from a different cluster. Observations from the same cluster are correlated. The heterogeneity between clusters introduces an additional source of variation, and as a result complicates the analysis. If the analysis is carried out without accounting for the heterogeneity induced by clustering, erroneous estimation of the variability of parameter estimates may occur. The point of multilevel modeling is that a statistical model should explicitly recognize a hierarchical structure when such a feature is present in the data [35].

Non-multilevel methods of analysing nested data are available. A common procedure with a two level data structure is to aggregate the micro level data to the macro level. The simplest way to do this is to work with averages for each single macro level unit. In cases where the research focuses on macro level propositions, there is nothing wrong with aggregation, but it should be borne in mind that the reliability of an aggregated variable depends, among other things, on the number of micro level units. In cases where the researcher is interested in micro level propositions aggregation may result in gross error. In other words, if the analysis is done at the macro level unit

then inferences cannot be made at the micro level unit. The major problem with aggregation is that it does not exploit all the within group information.

Potential errors arise in aggregation. The first potential error is the shift of meaning, that is, a variable that is aggregated to the second level refers to the level 2 units and not directly to the level 1 units. The second potential error is known as ecological fallacy. Correlation between level 2 variables cannot be used to make assertions about level 1 relations. The third potential error is the neglect of the original data structure, especially when some kind of covariance analysis is to be made.

The other non-multilevel analysis is to disaggregate all higher order variables to level 1. The drawback with disaggregation is that we cannot use the assumption of independence of observations [36]. We have to account for the correlation of lower level units within a higher level unit.

4.1.2 Complete Pooling, Partial Pooling and No Pooling

Multilevel models are thought as sitting between two extremes that are available to researchers dealing with data with some groupings. These extremes are complete pooling and no pooling. The fully pooled model ignores the groupings in the data and it fits a model treating the group level variables

as individual level covariates. For a model with one explanatory variable at the individual level (X_1) and one measured at the group level (X_2) with complete pooling the model can be written as;

$$Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + \epsilon_i. \tag{4.1}$$

In this model, the assertion is that group distinctions do not matter and that cases should be treated homogeneously, ignoring possible variations between groups. On the other hand is a set of models in which we treat each group as a separate dataset and model them separately. Such a model can be written as;

$$Y_{ij} = \beta_{0j} + X_{ij}\beta_{1j} + \epsilon_{ij}; \tag{4.2}$$

for $j = 1, \ldots, J$.

In this model, the group level predictor X_2 is not included since $X_{2i}\beta_2$ is constant within a group and therefore subsumed into the intercept term. The no pooling approach is the opposite of the fully pooled approach as it asserts that the groups are so completely different that it does not make sense to associate them in the same model. Such separate regression models clearly overstate the variation between groups, making them look more different than they really should be. Between these two extremes lies multilevel models. A multilevel model compromises between full distinction of groups and full ignoring of groups. This approach (multilevel model) can be regarded as

partial pooling or semi pooling in the sense that the groups are collected together into a single model, but their distinctiveness is preserved.

4.2 Frequentist Multilevel Model

In this section we discuss the non Bayesian multilevel model. To gain a deeper understanding of the non-linear multilevel model which is later applied to the SDHS data, we begin by reviewing the linear multilevel model.

4.2.1 The Multilevel Linear Model

In this section, we briefly discuss multilevel models for normally distributed response variables. Keeping an eye on the SDHS data as described in Chapter 2, we consider a three level hierarchical model. From the data, we assign clusters to level 3, households to level 2 and individuals to level 1.

Suppose we have a sample consisting of K clusters, with J_k households within the kth cluster ($k = 1, \ldots, K$) and N_{jk} individuals within the jth household from the kth cluster ($j = 1, \ldots, J_k, k = 1, \ldots, K$). Let y_{ijk} denote the value of the response variable from the ith individual within the jth household from the kth cluster. We now define the standard linear three level model as follows;

$$y_{ijk} = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + \boldsymbol{z}_{3,ijk}^T \mathbf{u}_k^{(3)} + \boldsymbol{z}_{2,ijk}^T \mathbf{u}_{jk}^{(2)} + \boldsymbol{\epsilon}_{ijk}, \tag{4.3}$$

where $\boldsymbol{x}_{ijk}$ is a vector of covariates having fixed effects $\boldsymbol{\beta}$, $\boldsymbol{z}_{3,ijk}$ is a vector of covariates having random effects $\mathbf{u}_k^{(3)}$ at the cluster level, $\boldsymbol{z}_{2,ijk}$ is a vector of covariates having random effects $\mathbf{u}_{jk}^{(2)}$ at the household level and $\boldsymbol{\epsilon}_{ijk}$ is vector of error term. The random terms are assumed to be mutually independent and distributed, that is,

$$\mathbf{u}_k^{(3)} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}^{(3)})$$

$$\mathbf{u}_{jk}^{(2)} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}^{(2)})$$

$$\boldsymbol{\epsilon}_{ijk} \sim \mathbf{N}(0, \sigma_e^2)$$

The vectors of covariates $\boldsymbol{z}_{3,ijk}$ and $\boldsymbol{z}_{2,ijk}$ will usually be a subset of the fixed effect covariates $\boldsymbol{x}_{ijk}$, although they need not be. The associated random effects are used to account for variation in the data that is attributable to clustering at the corresponding levels of hierarchy. The vector $\mathbf{u}_{jk}^{(2)}$ represent the effect of the jth household in the kth cluster on the covariates $\boldsymbol{z}_{2,ijk}$ and is characteristic of between household variability. Similarly $\mathbf{u}_k^{(3)}$ represents the effect of the kth cluster on the covariates $\boldsymbol{z}_{3,ijk}$, and is characteristic of between cluster variability.

In model (4.3), we assume that the residual variance is constant (homoscedasticity). This assumption can be relaxed by allowing dependence on specific covariates, that is replacing $\boldsymbol{\epsilon}_{ijk}$ by $\boldsymbol{z}_{1,ijk}^T \boldsymbol{\epsilon}_{ijk}$, where $\boldsymbol{z}_{1,ijk}$ are specific covariates of interest. This results in in complex variability in level one that

includes subgroup variability and heterogeneity.

Model (4.3) can be written in matrix form as a special case of the general linear mixed model as follows;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \tag{4.4}$$

where $\mathbf{y}$ is a vector of responses, $\mathbf{X}$ is a design matrix for the fixed effects, $\mathbf{u}$ is a vector of random effects obtained by stacking $\mathbf{u}_{jk}^{(2)}$ on top of the cluster effects $\mathbf{u}_{k}^{(3)}$, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{Z}$ is a design matrix for the random effects, and $\boldsymbol{\epsilon}$ are error terms obtained by stacking $\boldsymbol{\epsilon}_{ijk}$.

4.2.1.1 Parameter Estimation

In the multilevel model, parameter estimation can be done by maximising the likelihood function. Direct maximazation using the Newton Raphson or the expectation maximisation algorithm can be performed. Equivalently, an Iterative Generalised Least Squares (IGLS) procedure can be performed, and this was proposed by Golstein [13]. IGLS iterates between the estimation of the fixed and the random parameters using standardised least squares principles. Its advantage over the direct maximisation is its computational efficiency. It can also be modified to obtain residual or restricted maximum likelihood (REML) which are unbiased for random parameters, and in this case the procedure is referred to as RIGLS.

4.2.2 Non-Linear Multilevel Models

In this section, we restrict our attention to binary response cases, although the discussion can generally be applied to models of binomial and Poisson data. Using the same notation as in the previous section, we define a non-linear multilevel as;

$$g(\pi_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + \boldsymbol{z}_{3,ijk}^T\mathbf{u}_k^{(3)} + \boldsymbol{z}_{2,ijk}^T\mathbf{u}_{jk}^{(2)}, \tag{4.5}$$

where $\pi_{ijk} = P[y_{ijk} = 1|\mathbf{u}_k^{(3)}, \mathbf{u}_{jk}^{(2)}]$ and $g(.)$ is a link function which may be the logit, probit or log-log functions. In matrix form model (4.5) can be written as;

$$g(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \tag{4.6}$$

where $\boldsymbol{\pi}$ is the vector of response probabilities π_{ijk}. As in the linear model, we shall assume that all components of vector $\mathbf{u}$ are mutually independent and normally distributed. A further assumption is that conditional on $\mathbf{u}$, the binary response $\mathbf{y}_{ijk}$ are independent which is known as the conditional independence assumption.

The difference between model (4.6) and the linear multilevel model is that level one variability is not directly comparable to the variability at higher levels. This is because the link function $g(.)$ is in general different from the identity link, random disturbances from level 2 and above appear on a transformed scale such as the logit, whereas the level 1 variance characterises

binomial variation.

Estimation: Maximum Marginal Likelihood

In this section, for notational simplicity, we consider a two level model;

$$g(\boldsymbol{\pi}_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\mathbf{u}_j, \tag{4.7}$$

where $\boldsymbol{\pi}_{ij} = P[\mathbf{y}_{ij} = 1|\mathbf{u}_j]$,$(j = 1, \ldots, J; i = 1, \ldots, n_j)$ and $\mathbf{u}_j \sim N(\mathbf{0}, \boldsymbol{\Omega})$.

By the local independence assumption, the conditional likelihood of unit j (level 2) takes the binomial form, its contribution to the log marginal likelihood, obtained by integrating over the random effects can be written as;

$$\ell_j(\boldsymbol{\beta}, \boldsymbol{\Omega}) = log \int \prod_{i=1}^{n_j} \boldsymbol{\pi}_{ij}^{y_{ij}}(1 - \boldsymbol{\pi}_{ij})^{1-y_{ij}}\phi(\mathbf{u}_j, \boldsymbol{\Omega})d\mathbf{u}_j, \tag{4.8}$$

with $\phi(\mathbf{u}_j, \boldsymbol{\Omega})$ being the normal density function $N(\mathbf{0}, \boldsymbol{\Omega})$. The log marginal likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \sum_{j=1}^{n} \ell_j(\boldsymbol{\beta}, \boldsymbol{\Omega}) \tag{4.9}$$

can be maximised to obtain the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ using any standard optimisation methods. Unfortunately (4.8) is intractable hence the need to do numerical integration. The Gauss-Hermite quadrature can be used to evaluate the integral. This technique works well when the dimension of integration is small.

Several approaches have been proposed to ease the computational burden caused by the need for numerical integration. These techniques include the penalised quasi likelihood (PQL) and the marginal quasi likelihood (MQL). The PQL uses Laplace's integral approximation. PQL and MQL can be regarded as iterative procedures that requires fitting linear multilevel models based on the first order Tylor expansion of the mean function about the current fixed part predictor (MQL) or the current predicted value (PQL). Although these approximate procedures are computationally efficient than ML methods they are not without defect. When the number of simulations is quite large, these procedures may be seriously biased [35].

The Laplace Approximation

Consider the model given in (4.6) with $\boldsymbol{u} \sim N(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}))$, where $\boldsymbol{\Omega}$ is the $q \times q$ variance-covariance matrix. Although $\boldsymbol{\Omega}$ is a very large matrix, it is determined by a parameter vector, $\boldsymbol{\theta}$, whose dimension is typically small. Interest here is on finding the estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ that maximises the likelihood of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, given the data $\boldsymbol{y}$. This likelihood is numerically equivalent to the marginal density of $\boldsymbol{y}$, given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$;

$$f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_{\boldsymbol{u}} p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}) f(\boldsymbol{u}|\boldsymbol{\Omega}(\boldsymbol{\theta})) d\boldsymbol{\theta}, \tag{4.10}$$

where $f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\theta})$ is the probability mass function of $\boldsymbol{y}$, given $\boldsymbol{\beta}$ and $\boldsymbol{u}$, and $f(\boldsymbol{u}|\boldsymbol{\Omega})$ is the Gaussian probability density at $\boldsymbol{u}$ and $p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u})$ is the conditional density of $\boldsymbol{y}$.

When $p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u})$ is binomial, the above integral does not have a close form solution and one way to find its solution is to approximate it using a Laplace approximation. To obtain a Laplace's approximation to the likelihood $L(\boldsymbol{\beta},\boldsymbol{\theta}|\boldsymbol{y})$ we replace the logarithm of the integrand in (4.10) by its second-order Taylor series at conditional maximum, $u(\beta,\theta)$. On the scale of the deviance, that is, negative twice the log-likelihood, the approximation is given by;

$$-2\ell(\boldsymbol{\beta},\boldsymbol{\theta}|y) = -2log\left\{\int_{\boldsymbol{u}} p(\boldsymbol{y}|\boldsymbol{\beta},u)f(\boldsymbol{u}|\boldsymbol{\Omega}(\boldsymbol{\theta}))du\right\} \tag{4.11}$$

$$\approx 2log\int_{\boldsymbol{u}} exp\left\{-\frac{1}{2}[d(\boldsymbol{\beta},\tilde{\boldsymbol{u}},\boldsymbol{y}) + \tilde{\boldsymbol{u}}^T\Omega_{-1}\tilde{\boldsymbol{u}} + log|\boldsymbol{\Omega}| + \boldsymbol{u}^T\boldsymbol{D}^{-1}\boldsymbol{u}]\right\}$$

$$= d(\boldsymbol{\beta},\tilde{\boldsymbol{u}},\boldsymbol{y}) + \tilde{\boldsymbol{u}}^T\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{u}} + log|\boldsymbol{\Omega}| + log|\boldsymbol{D}|;$$

where $d(\boldsymbol{\beta},\tilde{\boldsymbol{u}},\boldsymbol{y})$ is the deviance function from the linear predictor only; $d(\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{y}) = -\log\, p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u})$. This quantity can be evaluated as the sum of deviance residuals.

4.2.3 Model Diagnostics and Selection

4.2.3.1 The Deviance

This is one of the statistics used in assessing goodness of fit and is defined as follows;

$$D^2 = 2\left\{ln[L_s(\hat{\boldsymbol{\beta}})] - ln[L_m(\hat{\boldsymbol{\beta}})]\right\}, \tag{4.12}$$

where $lnL_s(\hat{\boldsymbol{\beta}})$ is the maximised log-likelihood of the fitted model and $ln[L_m(\hat{\boldsymbol{\beta}})]$ is the maximised log-likelihood of the saturated model.

This quantity compares the values predicted by the fitted model and those predicted by the most complete model we could possibly fit. Evidence of lack of fit is shown by large value of D^2. Under specific regularity conditions D^2 converges asymptotically to a χ^2 distribution with p degrees of freedom: $D^2 \sim \chi^2_{(p)}$, where p is the difference between the number of parameters in the saturated model and the number of parameters in the model being considered. The saturated model represents the largest possible model we can fit and leads to perfect prediction of the outcome of interest.

4.2.3.2 Akaike's Infomation Criterion (AIC)

This is the widely used method of model selection. The idea behind the AIC is to select the model that minimises the negative likelihood penalised by the

number of parameters. Mathematically, the AIC can be written as;

$$AIC = -2log(L) + 2p, \tag{4.13}$$

where L is the likelihood under the fitted model and p is the number of parameters in the model. The AIC aims at finding the best approximating model to the unknown true data generating process [2].

4.2.4 Application

In this section, we present the results of the multilevel model as fitted in R using the Laplace approximation in the lme4 package. For initial analysis, we ignored the nested structure that exist in the data. We analysed the data using a logistic regression model. This analysis is referred to as complete pooling, that is, cluster level variables were treated as individual level covariates and a logistic regression model was run. The following section presents the results of the fully pooled model which is immediately followed by a section presenting results of the multilevel analysis.

The Fully Pooled Model

In fitting this model, the lmer function in lme4 in R was used. The cluster level variables included in the analysis are cluster average wealth, cluster average age and cluster proportion HIV positive . These were selected after

having run the model with all the cluster level variables, and the best model was chosen using the information criteria. The response variable is HIV status and its binary, that is, its either 1 for positive and 0 for negative. The outcome variable can be modeled as $Y_i|p_i \sim Bern(p_i)$ and;

$$log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X}_i\boldsymbol{\beta}, \tag{4.14}$$

where $\boldsymbol{X}_i$ is the vector of covariates for the ith individual, and $\boldsymbol{\beta}$ is a vector of coefficients to be estimated.

Table 4.1 shows the results of the initial analysis using a logistic regression where the multilevel structure that exist in the data set is ignored. The results of the analysis show age in both males and females to be a significant predictor for HIV serostatus, with the exception of the last category in females. No association was found between HIV serostatus and region in both males and females. The results also show that there is no significant association between HIV serostatus and type of place of residence in both males (OR: 1.22, 95% CI: 0.89–1.67) and females (OR: 1.24, 95% CI: 0.98–1.57). Being widowed in men is associated with a high probability of being HIV infected (OR: 4.93, 95% CI: 2.43–10.46). Similar results were also found in women (OR: 2.93, 95%CI: 1.98–3.81). Women who are married (OR: 0.62, 95% CI: 0.51–0.76) are less likely to be HIV infected than those who are never married

Women (OR: 0.50, 95% CI: 0.34–0.74) with higher level of education are less likely to be HIV infected than their counterparts with no education whilst no significant association was found between higher level of education (OR: 0.80, 95% CI: 0.51–1.27) in males and HIV status. Strong association between HIV risk and age at first sex was found in women, with those who started having sex at 14 years or less being 5.14 times more likely to be HIV positive. Men who reported to have had an STI in the last 12 months are more likely to be HIV positive than those who had no STI in the last 12 months (OR: 2.41, 95% CI: 1.70–3.40). Strong association between HIV serostatus and number of sexual partners was found in males, where both categories are statistically significant at $\alpha = 5\%$.

No significant association was found between number of sexual partners in the last 12 months and HIV serostatus in women. Also, weak association was found between the number of times one is away from home and HIV serostatus in both males and females. Women who had had a birth in the last 5 years are less likely to be HIV positive than those who had no birth (OR: 0.83, 95% CI: 0.71–0.98).

Among the cluster level variables, in the males cluster average age and cluster proportion HIV positive were found to be statistically significant, and for the females, only cluster proportion HIV was found to be statistically significant.

Table 4.1: Fully Pooled Model Results

	Males		Females	
Variable	OR (SE)	95%CI	OR (SE)	95%CI
Age				
15-19	1.00	ref	1.00	ref
20-24	4.59 (0.30)	2.62 -8.38	3.79 (0.14)	2.90 -4.97
25-29	10.10 (0.31)	5.62 -18.88	6.06 (0.15)	4.51 -8.17
30-34	24.06 (0.32)	13.05 -46.06	4.79 (0.16)	3.51 -6.57
35-39	28.39 (0.33)	15.10 -55.30	3.11 (0.17)	2.23 -4.35
40-44	20.70 (0.35)	10.63 -41.61	1.98 (0.18)	1.38 -2.83
45-49	13.74 (0.35)	6.94 -27.99	1.09 (0.20)	0.73 -1.63
Region				
Hhohho	1.00	ref	1.00	ref
Manzini	0.96 (0.14)	0.72 -1.27	0.97 (0.10)	0.80 -1.18
Shiselweni	0.94 (0.16)	0.68 -1.29	0.98 (0.11)	0.79 -1.22
Lubombo	0.95 (0.14)	0.72 -1.26	0.93 (0.11)	0.76 -1.15
Residence type				
urban	1.00	ref	1.00	ref
rural	1.22 (0.16)	0.89 -1.67	1.24 (0.12)	0.98 -1.57
Marital status				
never married	1.00	ref	1.00	ref
married	1.00 (0.15)	0.74 -1.35	0.62 (0.10)	0.51 -0.76
living together	2.17 (0.19)	1.48 -3.17	1.17 (0.13)	0.91 -1.50
widowed	4.93 (0.37)	2.43 -10.46	2.93 (0.17)	2.09 -4.12
divorced	1.69 (0.45)	0.69 -4.07	1.89 (0.57)	0.61 -5.93
not living together	2.49 (0.26)	1.49 -4.15	1.84 (0.21)	1.22 -2.78
Educational level				
no education	1.00	ref	1.00	ref
primary	1.15 (0.18)	0.81 -1.63	0.99 (0.13)	0.77 -1.29
secondary	1.04 (0.18)	0.73 -1.47	0.73 (0.14)	0.56 -0.96
higher	0.80 (0.23)	0.51 -1.27	0.50 (0.20)	0.34 -0.74
Age at first sex				
never had sex	1.00	ref	1.00	ref
14 yrs or less	1.18 (0.43)	0.50 -2.70	5.14 (0.24)	3.25 -8.23
15-17	1.99 (0.33)	1.05 -3.85	4.76 (0.21)	3.15 -7.29
18+yrs	1.36 (0.33)	0.71 -2.63	4.33 (0.22)	2.85 -6.67
Had an STI in last 12 months				
no	1.00	ref	1.00	ref
yes	2.41 (0.18)	1.70 -3.40	1.95 (0.15)	1.45 -2.65
don't know	1.59 (0.61)	0.44 -4.87	0.23 (0.67)	0.05 -0.76
Number of sexual partners				
no partner	1.00	ref	1.00	ref
1 partner	1.38 (0.21)	0.92 -2.11	1.07 (0.12)	0.85 -1.35
2+ partners	2.13 (0.23)	1.36 -3.37	1.62 (0.29)	0.92 -2.87
Number of times away from home				
none	1.00	ref	1.00	ref
1-5 times	1.10 (0.14)	0.83 -1.45	0.86 (0.10)	0.71 -1.04
6-10 times	1.36 (0.16)	0.99 -1.86	0.93 (0.12)	0.74 -1.16
11-15 times	1.06 (0.19)	0.73 -1.54	0.89 (0.14)	0.68 -1.16
16+times	0.86 (0.16)	0.64 -1.17	1.07 (0.12)	0.85 -1.36
Any births				
no			1.00	ref
Yes			0.83 (0.08)	0.71 -0.98
Cluster level variables (not odds ratios)				
cwealth.full	0.03 (0.07)	(-0.11) -0.16	0.02 (0.05)	(-0.07) -0.12
chiv.full	6.60 (0.44)	5.75 -7.48	4.88 (0.31)	4.29 -5.49
cage.full	-0.47 (0.09)	(-0.65) -(-0.28)	-0.11 (0.07)	(-0.26) -0.03

4.2.4.1 Multilevel Model

In this section we present the results of the multilevel model fitted using frequentist methods. The model was fitted in R using the Laplace approximation in the package lme4. The way the SDHS data was collected is such that it has three levels; individuals are nested within households and households are nested within clusters. In this analysis, households were ignored. This is because in some of the households there are very few individuals, in some there is only one individual making it difficult to get good estimates of household variability. Also in presenting these results, cluster level variables (though in the same table) are not given as odds ratios but just estimates. The results of the multilevel model are shown in Table 4.2.

The results show that age is an important predictor of HIV risk. The odds of infection are at peak in the age group 35-39 in males and in the age group 25-29 in females. Men in the age group 35-39 are 20 times more likely to be HIV infected than those in the age group 15-19. Women in the age group 25-29 are almost 6 times more likely to HIV positive than those at the age group 15-19. No significant association was found between region and HIV serostatus in both males and females. Residence type was found to be statistically significant in females. Men (OR: 0.90, 95% CI: 0.72–1.12) and women (OR: 0.78, 95% CI: 0.66–0.93) in the rural places are less likely to be

HIV positive than those in urban areas.

Strong association between being widowed persons and HIV serostatus in both males and females was found. Men who are widowed are 6.11 times more likely to be HIV positive than those who are never married. Women who are widowed are almost 3 times more likely to be HIV positive than those who are never married. Being married (OR: 0.59, 95% CI: 0.49–0.72) in women reduces the probability of being HIV positive. Women who are divorced are 2.41 times more likely to be HIV infected than those who are in the never married category. On the other hand, men who are divorced are almost 2 times more likely to be HIV positive than those who are never married.

Men having primary level of education (OR: 1.08, 95% CI: 0.77–1.51) were found to be more likely to be HIV positive than those in the no education category. In both males (OR: 0.54, 95% CI: 0.35–0.83) and females (OR: 0.42, 95% CI: 0.29–0.62) having higher level of education is associated with lower probability being HIV positive. Concerning age at first sex and HIV serostatus, strong association was found among the females. Females who begun having sex encounters at age 14 or less (OR: 5.49, 95% CI: 3.48–8.68) are more likely to be HIV positive than those who had never had sex. Also, women who started having sex between the age of 15 and 17 are 4.99 times more likely to be HIV positive than those who never had sex. Weak

association between age at first sex and HIV serostatus was found among males.

Among males, an increase in the number of sexual partners one had in the past 12 months increases the probability of HIV infection. No significant associations were found between the number of sexual partners and HIV serostatus in females. Men who were away from home for 6-10 times (OR: 1.45, 95% CI: 1.07–1.95) are more likely to be HIV positive than those who never slept away from home in the past 12 months. No significant association was found between the number of times away from home and HIV serostatus in females.

The cluster level errors for males have an estimated standard deviation of 0.29 on the logit scale. Dividing this figure by 4 tells us that the clusters differed by approximately ±7% on the probability scale over and above the differences explained by the other covariates. A similar interpretation holds for the females. Dividing 0.25 by 4 tells that with regards to HIV serostatus,the clusters differed by approximately ±6% on the probability scale over and above the differences explained by the other covariates.

Table 4.2: Multilevel Model Results

	Males		Females	
Variable	OR (SE)	95%CI	OR (SE)	95%CI
Age				
15-19	1.00	ref	1.00	ref
20-24	4.09 (0.29)	2.31 -7.26	3.82 (0.13)	2.94 -4.97
25-29	8.42 (0.30)	4.66 -15.23	5.99 (0.15)	4.49 -8.01
30-34	17.79 (0.31)	9.64 -32.85	4.73 (0.16)	3.49 -6.41
35-39	20.35 (0.32)	10.86 -38.11	3.17 (0.17)	2.29 -4.39
40-44	15.53 (0.34)	8.03 -30.04	1.92 (0.18)	1.35 -2.72
45-49	8.91 (0.34)	4.54 -17.46	1.07 (0.20)	0.73 -1.58
Region				
Hhohho	1.00	ref	1.00	ref
Manzini	0.80 (0.14)	0.60 -1.06	0.89 (0.10)	0.72 -1.09
Shiselweni	0.80 (0.16)	0.58 -1.11	0.95 (0.12)	0.76 -1.19
Lubombo	0.81 (0.14)	0.61 -1.08	0.91 (0.11)	0.73 -1.13
Residence type				
urban	1.00	ref	1.00	ref
rural	0.90 (0.11)	0.72 -1.12	0.78 (0.09)	0.66 -0.93
Marital status				
never married	1.00	ref	1.00	ref
married	1.13 (0.14)	0.85 -1.51	0.59 (0.10)	0.49 -0.72
living together	2.29 (0.18)	1.61 -3.27	1.17 (0.12)	0.92 -1.49
widowed	6.11 (0.36)	3.02 -12.37	2.84 (0.17)	2.04 -3.95
divorced	1.79 (0.42)	0.78 -4.10	2.41 (0.55)	0.83 -7.04
not living together	2.75 (0.25)	1.69 -4.47	1.82 (0.20)	1.22 -2.71
Educational level				
no education	1.00	ref	1.00	ref
primary	1.08 (0.17)	0.77 -1.51	0.99 (0.13)	0.76 -1.28
secondary	0.90 (0.17)	0.64 -1.25	0.69 (0.13)	0.53 -0.90
higher	0.54 (0.22)	0.35 -0.83	0.42 (0.19)	0.29 -0.62
Age at first sex				
never had sex	1.00	ref	1.00	ref
14 yrs or less	1.19 (0.42)	0.52-2.68	5.49 (0.23)	3.48-8.68
15-17	2.10 (0.33)	1.11-3.97	4.99 (0.21)	3.30-7.54
18+yrs	1.49 (0.33)	0.78-2.76	4.49 (0.21)	2.96-6.83
Had an STI in last 12 months				
no	1.00	ref	1.00	ref
yes	2.47 (0.17)	1.77 -3.44	2.03 (0.15)	1.52 -2.71
don't know	1.04 (0.58)	0.33 -3.26	0.28 (0.66)	0.08 -1.00
Number of sexual partners				
no partner	1.00	ref	1.00	ref
1 partner	1.50 (0.20)	1.01 -2.24	1.13 (0.11)	0.91 -1.42
2+ partners	2.47 (0.22)	1.59 -3.81	1.68 (0.28)	0.96 -2.92
Number of times away from home				
none	1.00	ref	1.00	ref
1-5 times	1.09 (0.14)	0.83 -1.42	0.86 (0.09)	0.72 -1.04
6-10 times	1.45 (0.15)	1.07 -1.95	0.95 (0.11)	0.76 -1.18
11-15 times	1.14 (0.18)	0.80 -1.63	0.89 (0.13)	0.69 -1.15
16+times	0.85 (0.15)	0.64 -1.14	1.10 (0.12)	0.87 -1.38
Any births				
no			1.00	ref
yes			0.83 (0.08)	0.71 -0.97
Cluster level variables (not odds ratios)				
cwealth.full	0.15 (0.05)	0.04 -0.26	0.02 (0.03)	(-0.03) -0.08
chiv.full	0.02 (0.35)	(-0.66) -0.70	0.04 (0.26)	(-0.47) -0.55
cage.full	-0.14 (0.08)	(-0.29) -0.02	0.06 (0.10)	(-0.13) -0.24
$\sigma_{cluster}$	0.29		0.25	

4.3 Bayesian Multilevel Models

In this section we consider Bayesian Multilevel models. The discussion begins with the normal linear multilevel model followed by multilevel generalised linear models. This approach is followed so that the distinction between the two types of models is clarified.

4.3.1 The Linear Multilevel Model

A multilevel model typically assumes observations to be independent conditional on fixed regression and random effects defined at one or more levels of the data hierarchy. With continuous outcomes, the two level random effects model have been widely used for nested data and is given as;

$$\boldsymbol{y}_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{u}_j + \boldsymbol{\epsilon}_{ij}, \tag{4.15}$$

where y_{ij} is the response for the ith observation ($i = 1, 2, \ldots, n_j$) in cluster j ($j = 1, 2, \ldots, m$), $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of covariates associated with response variable, $\boldsymbol{\beta}$ is a column vector of regression coefficients that are of scientific interest, $\boldsymbol{z}_{ij}$ is a $q \times 1$ vector of random coefficients, $\boldsymbol{u}_j$ and $\boldsymbol{\epsilon}_{ij}$ denote cluster and observation level random effects. The vector $\boldsymbol{x}_{ij}$ includes the intercept. With $N = \sum_{j=1}^{m} n_j$, the total number of observations, the nested form of the model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}, \tag{4.16}$$

where $\boldsymbol{y}$ is a $N \times 1$ vector, $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \dots \\ \boldsymbol{X}_m \end{bmatrix}$ is $N \times p$ with $\boldsymbol{X}_j = (\boldsymbol{x}_{1j}, \dots, \boldsymbol{x}_{n_j})^T$ of dimension $n_j \times p$, $\boldsymbol{Z}$ is $N \times mq$ block diagonal matrix with m diagonal blocks, $\boldsymbol{Z}_i = (\boldsymbol{z}_{1j}, \dots, \boldsymbol{z}_{jn_j})^T$ of dimension $n_j \times q$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population parameters and $\boldsymbol{u}_j = (u_{1j}, \dots, u_{qj})^T$ is $q \times 1$ vector of zero mean cluster specific deviations around those population parameters, with $\boldsymbol{u}_j$ assumed random. Thus $\boldsymbol{u} = (\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \dots, \boldsymbol{u}_m^T)$ is a $mq \times 1$ vector.

The conjugate linear normal model with random cluster effects assumes multivariate normality for these effects as well as the observational level errors that is

$$\boldsymbol{u}_j = (u_{1j}, \dots, u_{qj})^T \sim N_q(\boldsymbol{0}, \boldsymbol{\Omega}).$$

Assuming a prior $\boldsymbol{\epsilon}_j \sim N_{n_j}(\boldsymbol{0}, \boldsymbol{H}_j)$, where $\boldsymbol{H}_j$ is the within-cluster dispersion matrix. The stacked form of the linear mixed model at cluster level, namely

$$\begin{bmatrix} \boldsymbol{y}_j \\ \boldsymbol{u}_j \end{bmatrix} \sim N_{n_j+q} \left(\begin{bmatrix} \boldsymbol{X}_j\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_j\boldsymbol{\Omega}\boldsymbol{Z}_j^T + \boldsymbol{H}_j & \boldsymbol{Z}_j\boldsymbol{\Omega} \\ \boldsymbol{\Omega}\boldsymbol{Z}_j & \boldsymbol{\Omega} \end{bmatrix} \right)$$

or as

$$\boldsymbol{y}_j \sim N_{n_j} \left(\boldsymbol{X}_j\boldsymbol{\beta}, \boldsymbol{Z}_j\boldsymbol{\Omega}\boldsymbol{Z}_j^T + \boldsymbol{H}_j \right)$$

in marginal form.

For priors, the conjugate model takes inverse gamma and the inverse Wishart for σ^2 and $\boldsymbol{\Omega}$ respectively. Common practice is to assign proper priors such as $\sigma^2 \sim IG(\epsilon, \epsilon)$ with ϵ small. Alternative priors for the observation level variance is to use the uniform prior or the half t prior [9]. For $\boldsymbol{\beta}$ we use the multivariate normal distribution as a prior, that is, $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$.

4.3.1.1 The Lindley-Smith Model Format

This is an alternative way of representing the normal linear multilevel model and is based on the scheme of Lindley and Smith [24]. In this scheme, it is assumed that all the effects of a level 1 predictor vary randomly over clusters with their variability explained by cluster predictors

$$\boldsymbol{W}_j = (w_{1j}, w_{2j}, \ldots, w_{rj})^T.$$

Based on this scheme, the two level model can be written as;

$$\boldsymbol{y}_j = \boldsymbol{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \tag{4.17}$$

$$\boldsymbol{\beta}_j = \boldsymbol{\kappa} \boldsymbol{W}_j + \boldsymbol{u}_j;$$

where $\boldsymbol{y}_{ij} = (y_{j1}, \ldots, y_{jn_j})^T$ is an $n_j \times 1$, $\boldsymbol{\kappa}$ is $q \times r$, $\boldsymbol{Z}_j$ is $n_j \times q$, $\boldsymbol{\beta}_j$ is a vector of random cluster regression parameters, and the errors $\boldsymbol{\epsilon}_j = (\epsilon_{1j}, \ldots, \epsilon_{n_j j})^T$ have a prior $\epsilon_{ij} \sim N(0, \sigma^2)$. The level 2 regression for $\boldsymbol{\beta}_j$ involves a fixed effect parameters, $\boldsymbol{\kappa}$, and the errors $\boldsymbol{u}_j = (u_{1j}, \ldots, u_{qj})^T$ having a prior $\boldsymbol{u}_j \sim$

$N(\mathbf{0}, \boldsymbol{T}_u)$. When the second equation is substituted in the first equation we get

$$\boldsymbol{y}_j = \boldsymbol{Z}_j \boldsymbol{\kappa} \boldsymbol{W}_j + \boldsymbol{Z}_j \boldsymbol{u}_j + \boldsymbol{\epsilon}_j. \tag{4.18}$$

The model can be reformulated as a mixed model in order to constrain the effect of one or more level 1 predictors to have an identical effect across clusters.

For this model one may assume a uniform prior for $\boldsymbol{\kappa}$, gamma and Wishart priors for σ and $\boldsymbol{T}_u$, that is, $1/\sigma^2 \sim G(a_\epsilon, b_\epsilon)$ and $\boldsymbol{T}_u \sim W(S_u, \upsilon_u)$. Let $\boldsymbol{r}_{ij} = \boldsymbol{y}_{ij} - \boldsymbol{Z}_{ij}\boldsymbol{\beta}_j$, $\hat{\boldsymbol{\beta}}_j = (\boldsymbol{Z}_j^T \boldsymbol{Z}_j)^{-1} \boldsymbol{Z}_j \boldsymbol{y}_j$, $\tilde{\boldsymbol{V}}_j = (\sigma^{-2} \boldsymbol{Z}_j^T \boldsymbol{Z}_j + \boldsymbol{T}_u)^{-1}$, $\boldsymbol{V}_j = \sigma^2 \boldsymbol{Z}_j^T \boldsymbol{Z}_j$, $\boldsymbol{\Lambda}_j = (\boldsymbol{V}_j^{-1} + \boldsymbol{T}_u)\boldsymbol{V}_j^{-1}$, $\boldsymbol{U}_j = (\boldsymbol{\beta}_j - \boldsymbol{\kappa} W_j)$ and $G = [\sum \boldsymbol{W}_j^T \boldsymbol{T}_u W_j]^{-1}$, then the full conditionals for Gibbs sampling are

$$\begin{aligned}
1/\sigma^2 &\sim G\left(0.5(a_\epsilon, m), 0.5(b_\epsilon + \sum_{j=1}^{m}\sum_{i=1}^{n_j} r_{ij}^2)\right), \\
\boldsymbol{\beta}_j &\sim N_q(\boldsymbol{\Lambda}_j \hat{\boldsymbol{\beta}}_j + (\boldsymbol{I} - \boldsymbol{\Lambda}_j)\boldsymbol{\kappa}\boldsymbol{W}_j, \tilde{\boldsymbol{V}}_j), \\
\boldsymbol{T}_u &\sim W\left(S_u + \sum_{j=1}^{m} \boldsymbol{U}_j \boldsymbol{U}_j^T, m + \upsilon_u\right), \\
\boldsymbol{\kappa} &\sim N_r\left(\boldsymbol{G}\sum_{j=1}^{m} \boldsymbol{W}_j \boldsymbol{T}_u \beta_j, \boldsymbol{G}\right).
\end{aligned}$$

4.3.2 Multilevel Model for Discrete Data

The linear normal model discussed thus far can be extended to discrete outcomes. Consider a univariate, y_{ij}, with repetitions i nested in cluster j, that

is conditional on cluster effects u_j, follow an exponential density,

$$f(y_{ij}|u_j) \propto exp\left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{\phi_{ij}} + c(y_{ij}, \phi_{ij})\right\}; \tag{4.19}$$

where θ_{ij} is the canonical parameter, and ϕ_{ij} is the scale parameter. Furthermore, $E(y_{ij}|\theta_{ij}) = d'(\theta_{ij})$ and $\text{var}(y_{ij}|\theta_{ij}, \phi_{ij}) = d''(\theta_{ij})\phi_{ij}$. In the Poisson model, for instance, $d(u) = exp(u)$ and for the binomial case $d(u) = log(1 + e^u)$.

Let $\eta_{ij} = g(\theta_{ij})$, where $g(.)$ is a link function such that $\boldsymbol{\eta}_j = (\eta_{1j}, \ldots, \eta_{n_j j})^T$,now the observation level model is given by

$$\boldsymbol{\eta}_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{u}_j, \tag{4.20}$$

$$\boldsymbol{u}_j = \boldsymbol{\kappa}\boldsymbol{W}_j + \boldsymbol{\epsilon};$$

where $\boldsymbol{\beta}$ and $\boldsymbol{u}_j$ are of dimension p and q, respectively.

To model overdispersion, for example in the Poisson model, it is common to include observation level residual term so that

$$\boldsymbol{\eta}_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{u}_j + \boldsymbol{\epsilon}_{ij}. \tag{4.21}$$

It is also possible to have observational level predictors, say g_{ij} of dimension s with varying effects at observational level. In such a case (4.21) becomes

$$\boldsymbol{\eta}_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{u}_j + \boldsymbol{g}_{ij}^T\epsilon_{ij}. \tag{4.22}$$

For priors, we assume $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \boldsymbol{R})$, $\boldsymbol{u}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_u)$, $\boldsymbol{\epsilon} \sim N_r(\mathbf{0}, \boldsymbol{\Omega}_\epsilon)$, $\boldsymbol{\Omega}_u \sim IW(\upsilon_u, S_u)$, $\boldsymbol{\Omega}_\epsilon \sim IW(\upsilon_\epsilon, S_\epsilon)$, then the full conditional for each u_j vector is given by

$$p(\boldsymbol{u}_j|\boldsymbol{u}_{[j]}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\Omega}_u, \boldsymbol{\Omega}_\epsilon) \propto exp\left\{-0.5\boldsymbol{u}_j^T\boldsymbol{\Omega}_u^{-1}\boldsymbol{u}_j + \sum_{i=1}^{n_j}\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{\phi_{ij}}\right\},$$

while the full conditional for each ϵ vector is

$$p(\boldsymbol{\epsilon}_{ij}|\boldsymbol{\epsilon}_{ij}, \boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{\epsilon}, \boldsymbol{\Omega}_u, \boldsymbol{\Omega}_\epsilon) \propto exp\left\{-0.5\boldsymbol{\epsilon}_{ij}^T\boldsymbol{\Omega}_\epsilon^{-1}\boldsymbol{\epsilon}_{ij} + \frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{\phi_{ij}}\right\}.$$

Also, the covariance matrices have inverse Wishart full conditionals, namely,

$$\boldsymbol{\Omega}_u \sim IW\left(\upsilon_u + m, S_u + \sum_{j=1}^{m}\boldsymbol{u}_j\boldsymbol{u}_j^T\right)$$

$$\boldsymbol{\Omega}_\epsilon \sim IW\left(\upsilon_\epsilon + \sum_{j=1}^{m}n_j, S_u + \sum_{i,j}\boldsymbol{\epsilon}_{ij}\boldsymbol{\epsilon}_{ij}^T\right).$$

This approach can be extended to multinomial observations.

4.3.3 Application

Table 4.3 shows the results of fitting the Bayesian multilevel model to the SDHS data. The model was fitted using the following prior distributions. For the unmodeled parameters the multivariate normal distribution (essentially uninformative) was used, and for the variance-covariabce matrix we used the inverse Wishart prior distibution (weakily informative). In the multilevel logistic model the level 1 error varriance is fixed and cannot be estimated

from the data, therefore in fitting this model we fixed the level 1 varriance at 1. The model was fitted in MCMCglmm using 60 000 iterations with a thinning rate of 10. The Raftery and Lewis diagnostic was used in coming up with the number of iterations to be run, that is a shorter pilot chain was first run. Convergence of the chain was monitored using diagnostic plots which are found in the appendix.

The results of the Bayesian multilevel model shows that age is a strong predictor of HIV infection, with males having the highest odds ratios. Men in the age categories 30-34 (OR: 26.67 , 95% CI: 13.73–53.26), 35-39 (OR: 31.55, 95% CI: 15.89–64.74) and 40-44 (OR: 22.51, 95% CI: 11.12–49.36) are more likely to be HIV positive than those aged 15-19. Women aged 25-29 (OR: 8.29, 95% CI: 5.97–11.54) are more likely to be HIV infected than those in the reference category. Also, women in the age group 30-34 are 6.19 times more likely to be HIV positive than those aged 15-19. The last age category in the females is not statistically significant as the confidence interval does contain 1.

In both males and females, region was found to be statistically insignificant. It it worth noting that there is no variability in the odds ratios for the different regions in both males and females. The probability of being HIV infected does not seem to vary from region to region in both groups. Women (OR: 0.74, 95% CI: 0.61–0.91) in the rural areas are less likely to

be HIV infected than those in the urban areas. For Men (OR: 0.88, 95% CI: 0.68–1.16), no significant association was found between HIV serostatus and the type of place of residence. Strong association between marital status and HIV serostatus was found. Women who are married (OR: 0.54, 95% CI: 0.43–0.68) are less likely to be HIV positive while those who are widowed (OR: 3.54, 95% CI: 2.39–5.18) and divorced (OR: 2.87, 95% CI: 0.78–10.45) are more likely to be HIV positive than those in the never married category. Similarly, men who are divorcrd (OR: 8.87, 95% CI: 3.92–20.30) and those not living together (OR: 3.43, 95% CI: 1.89–6.14) with their partners are more likely to be HIV positive than those who are never married.

Men with higher level (OR: 0.47, 95% CI: 0.28–0.77) of education are less likely to be HIV positive compared those with no education. Similarly, women with secondary level (OR: 0.65, 95% CI: 0.47–0.88) and higher level (OR: 0.36, 95% CI: 0.23–0.56) of education have a lower probability of being HIV infected. Age at first sex proved to be stronglty related to HIV serostatus in women. Women who started having sexual intercourse at the age of 14 or less (OR: 6.99, 95% CI: 4.17–11.64) and those who started having sex in the age group 15-17 (OR: 6.15, 95% CI: 3.82–9.84) are more likely to be HIV positive than those who never had sex. Weak association between age at first sex was found among the males. On the other hand, however, strong association was found between the number of sexual partners and HIV

serostatus in males. Men with two or more sexual partners (OR: 2.85, 95% CI: 1.75–4.73) are more likely to be HIV positive than those who had no sexual partner in the last 12 months. Weak association between the number of sexual partners and HIV serostatus in women.

Weak association was found between the number of times one is away from home and HIV serostatus in both men and women. Women who had a birth (OR: 0.80, 95% CI: 0.66–0.97) in the last 5 years are less likely to be HIV positive compared to those had a birth. Among the cluster level variables, only cluster average wealth was stistically significant in males.

The cluster level errors for men have an estimated standard deviation of 0.30 on the logit scale. Upon diving by 4 tells us that the clusters differed by approximately $\pm 7.5\%$ on the probability scale over and above the differences explained by the other covariates. For the females, the cluster level errors are estimated at 0.29. Over and above the differences explained by the other covariates, the clusters differed by approximately $\pm 7.3\%$ with regards to HIV serostatus on the probability scale.

Table 4.3: Bayesian Multilevel Model Results

	Males		Females	
Variable	OR (SE)	95%CI	OR (SE)	95%CI
Age				
15-19	1.00	ref	1.00	ref
20-24	4.79(0.32)	2.54 -8.95	4.84 (0.15)	3.59 -6.54
25-29	10.97(0.33)	6.05 -21.53	8.29 (0.17)	5.97 -11.54
30-34	26.67(0.35)	13.73 -53.26	6.19 (0.18)	4.37 -8.94
35-39	31.55(0.36)	15.89 -64.74	3.87 (0.19)	2.69 -5.62
40-44	22.51(0.38)	11.12 -49.36	2.15 (0.21)	1.43 -3.27
45-49	11.53(0.38)	5.62 -25.37	1.07 (0.23)	0.68 -1.65
Region				
Hhohho	1.00	ref	1.00	ref
Manzini	0.78(0.17)	0.55 -1.07	0.88 (0.13)	0.68 -1.12
Shiselweni	0.78(0.19)	0.53 -1.13	0.95 (0.14)	0.73 -1.24
Lubombo	0.79(0.17)	0.57 -1.10	0.90 (0.13)	0.69 -1.16
Residence type				
urban	1.00	ref	1.00	ref
rural	0.88(0.14)	0.68 -1.16	0.74 (0.10)	0.61 -0.91
Marital status				
never married	1.00	ref	1.00	ref
married	1.17(0.17)	0.84 -1.62	0.54 (0.12)	0.43 -0.68
living together	2.72(0.22)	1.79 -4.14	1.21 (0.15)	0.90 -1.63
widowed	8.87(0.43)	3.92 -20.30	3.54 (0.20)	2.39 -5.18
divorced	1.99(0.52)	0.71 -5.34	2.87 (0.65)	0.78 -10.45
not living together	3.43(0.30)	1.89 -6.14	2.09 (0.24)	1.28 -3.32
Educational level				
no education	1.00	ref	1.00	ref
primary	1.10(0.20)	0.75 -1.61	0.99 (0.16)	0.72 -1.34
secondary	0.87(0.20)	0.59 -1.28	0.65 (0.16)	0.47 -0.88
higher	0.47(0.26)	0.28 -0.77	0.36 (0.23)	0.23 -0.56
Age at first sex				
never had sex	1.00	ref	1.00	ref
14 yrs or less	1.06(0.45)	0.44 -2.53	6.99 (0.27)	4.17 -11.64
15-17yrs	2.13(0.35)	1.04 -4.19	6.15 (0.24)	3.82 -9.84
18+yrs	1.43(0.35)	0.71 -2.82	5.42 (0.24)	3.37 -8.74
Had an STI in last 12 months				
no	1.00	ref	1.00	ref
yes	2.92(0.20)	1.99 -4.35	2.35 (0.17)	1.69 -3.30
don't know	1.02(0.70)	0.24 -3.82	0.21 (0.77)	0.04 -0.84
Number of sexual partners				
no partner	1.00	ref	1.00	ref
1 partner	1.57(0.23)	1.00 -2.50	1.16 (0.14)	0.88 -1.53
2+ partners	2.85(0.25)	1.75 -4.73	1.85 (0.34)	0.95 -3.57
Number of times away from home				
none	1.00	ref	1.00	ref
1-5 times	1.13(0.16)	0.82 -1.55	0.84 (0.11)	0.67 -1.05
6-10 times	1.58(0.18)	1.12 -2.25	0.94 (0.14)	0.71 -1.21
11-15 times	1.17(0.22)	0.78 -1.80	0.87 (0.16)	0.64 -1.19
16+times	0.84(0.17)	0.59 -1.17	1.12 (0.14)	0.84 -1.47
Any births				
no			1.00	ref
yes			0.80 (0.10)	0.66 0.97
Cluster level variables (not odds ratios)				
cwealth.full	0.18(0.06)	0.05 -0.30	0.03 (0.03)	(-0.04) -0.10
chiv.full	0.00(0.40)	(-0.80) -0.80	0.05 (0.31)	(-0.55) -0.68
cage.full	-0.16(0.10)	(-0.35) -0.03	0.07 (0.11)	(-0.16) -0.30
$\sigma_{cluster}$	0.30(0.10)		0.29(0.06)	

4.3.4 Summary

In this chapter, we have analysed the SDHS data using multilevel models. The models fitted in this chapter are the Bayesian and the frequentist multilevel models. Before fitting the frequentist multilevel model, we first fitted a complete pooling model to the data. In this section, we shall first discuss the performance of the models on the data and then the results obtained by fitting these models.

The first two models fitted to the data are the complete pooling model and the frequentist multilevel model. Both models were fitted in R. The fully pooled model was fitted using the glm function whilst for the frequentist multilevel model lmer function was used. There are notable differences between the results of the two methods. One of the differences is that of standard errors. The standard errors obtained from the fully pooled model are bigger compared to the ones obtained from the multilevel model. The second difference concerns confidence intervals. Confidence intervals from the fully pooled model are generally wider than the ones from the frequentist multilevel model. There are cases where these two methods lead to different inferences. In the males, the higher level of education is not statistically significant in the fully pooled model but it is statistically significant in the multilevel model. Also, in the variable "number of sexual partners" one

category is not statistically significant in the males whilst it is statistically significant in the frequentist multilevel model. The complete pooling analysis ignores any variation in the HIV serostatus between the clusters and hence this can lead to misleading inference [10].

Now in comparing the results of the Bayesian multilevel model and the frequentist multilevel model, the following differences were observed. The parameter estimates are similar for both models with the exception for the variable age. For the variable age, the Bayesian multilevel model produced larger odds ratios than the ones estimated using likelihood based methods. In both methods same inference can be reached, that is to say, variables that are statistically significant in the Bayesian multilevel model are also statistically significant in the frequentist multilevel model. The standard errors for the Bayesian multilevel model are larger than the ones from the frequentist multilevel model. Since the standard errors are used for the computation of estimates of confidence intervals, the confidence intervals from the Bayesian multilevel model are generally wider than the ones from the frequentist multilevel model though the difference is minimal.

Browne and Draper [7] in their paper "A comparison of Bayesian and likelihood-based methods for fitting multilevel model", in two examples that they did, find that the variance components estimates obtained using Bayesian methods (MCMC) are bigger than the variance components estimates ob-

tained using the likelihood based approaches. In this study, similar results were found. The estimates of the variance components $\sigma^2_{cluster}$ are larger for the Bayesian multilevel model compared to the ones from the frequentist multilevel model. As stated above that both methods of estimation lead to the same inference, in choosing which method to use one has to consider factors such as computational speed, flexibility, informativeness of the output. Bayesian outputs generally carry more information. From a Bayesian output one can choose between reporting the mean or the median. Flexibility in building the model by incorporating prior information is also an inherent feature embedded in Bayesian methods. The major drawback of Bayesian methods (MCMC) is computational speed.

Strong association was found between age and HIV serostatus. The odds of infection are quite higher for males than for females. Similar studies [14, 19] found similar results between age and serostatus. Contrary to the findings of Way and Johson [19] and Magadi and Desta [27], this study found no relationship between region and HIV serostatus in both males and females. Women in rural areas were found to be less likely to be HIV infected. This result is in line with the findings of Boerma et al [6] with regards to women but contrary to the findings of Solomon et al [42] who found that HIV prevalence was higher in rural areas than in urban areas.

Women classified as widowed or divorced were found to be more likely

to be HIV positive. For the widowed, it maybe that their husbands died of HIV. This result is in line with the findings of Johnson and Way [19]. This study also found that men and women with higher levels of education are less likely to be HIV positive. This result is in line with the findings of Johnson and Way [19] on the other hand contrary to the findings of Glynn et al [12].

Strong relationship between age at first sex and HIV serostatus was found among females but weak association was found among males. Among women, the odds of infection are higher for those who begun having sexual intercourse at the age of 14 or less or from 15–17 years. Starting having sexual intercourse at an early age in women increases thier vulnerability to HIV infection. Economic hardships and challenges sometimes force young girls and older women to engage in sexual practices for survival [44].

The number of sexual partners one had in the last 12 months was found to be a strong predictor of HIV risk among men but weak association was found among women. This result is in line with results of Morris and Kretzshmar [32] who found that concurrent parterships increase the number of infected individuals and the growth rate of the epidemic. Contrary to the findings of Gillespie et al [11] who found convincing evidence linking the spread of HIV and the number of times away from home, this study found non convincing evidence linking the two.

This study found that for both males and females, with regards to HIV

seropositivity, the clusters differed by less than 10%. The cluster HIV seropositivity variation accounted for less than 10% of the total variability over and above the differences explained by the other covariates.

Chapter 5

Conclusion

In this study, we have modeled HIV in Swaziland using Bayesian GLMMs and both frequentist and Bayesian multilevel models to ascertain the extent to which HIV serostatus is related to various explanatory variables. One of the objectives of this study was to compare performance of these models on the SDHS data. The SDHS data was collected through multistage sampling, hence the data has some groupings or levels. The nesting in the data is such that individuals are nested within households and households are nested in clusters. We restricted this study to two levels, level 1 being individuals and level two being clusters, that is, we ignored households as a level since in some households there are few individuals. Since the outcome variable in this study is binary, and due to the presence of random effects or levels, the GLMMs and multilevel models were deemed appropriate in analysing the

data. Multilevel models differ from ordinary GLMMs in the sense that they add predictors at individual and group level reducing unexplained variation in each level [10].

In chapter 3, we fitted the Bayesian generalised linear mixed model. In fitting this model, we used two estimation techniques namely MCMC method and INLA. The INLA approach is superior to the MCMC methods in terms of computational speed. Though INLA is still under development, it promises to be a better alternative to the widely used MCMC methods for full Bayesian inference.

In chapter 4, we fitted the multilevel models from a frequentist and the Bayesian perspectives. Before fitting the frequentist multilevel model, we fitted as a preliminary analysis a complete pooling model ignoring the groupings in the data. Differences between the frequentist multilevel model and the complete pooling model were observed. Multilevel models are regarded as a compromise between complete pooling and no pooling. In this study, we observed that the frequentist multilevel model produces estimates with smaller standard errors than complete pooling. When the between group variation is neglible that is $\sigma_{cluster} \rightarrow 0$, we expect that the results from complete pooling and multilevel model to be very similar.

Both the frequentist and the Bayesian multilevel models yielded similar results. There were some differences in standard errors and estimates of

variance components but both methods lead to same inference. Differences were observed between the results of the multilevel models and the Bayesian GLMMs of Chapter 3. One of the pronounced difference between between the results is that of differing estimates of cluster effects. In the Bayesian GLMM, the estimates of the cluster effect are larger than the ones from the multilevel models. The Bayesian GLMM was run without the inclusion of cluster level variables. Group level predictors play a crucial role in multilevel modeling by reducing unexplained group level variation and in the event reducing group level standard deviation [10]. The superiority of the multilevel models come from the fact they are able to account for the unexplained variability than the GLMMs.

In an attempt to identify key drivers of HIV in Swaziland, this study found that age, age at first sex, marital status and the number of sexual partners one had in the last 12 months are associated with HIV serostatus. In addition, exposure to STIs also seems to put one at higher risk of being HIV infected than an individual who is not exposed to STIs. Weak between cluster variations were found in both men an women.

5.1 Implications

In an attempt to curb the spread of HIV/AIDS in Swaziland, efforts should be directed at discouraging females from engaging in sexual intercourse early in their lives as this increases their chances of contracting HIV. The factors that make females to engage in sex early in their lives should be identified since it is possible that some are forced into sexual practices because of economic difficulties. Secondly, the issue of concurrent partnerships should be discouraged among men. Some of the cultural norms should be revisited, multiple sexual partnerships among males are encouraged by the Swazi culture. Young men with many girlfriends are regarded as heroes in the Swazi culture.

5.2 Further Work

As much as this study put into consideration the way the data was collected in the analysis, further work can be done on sensitivity of the Bayesian estimates to priors and also on missing data as this study used complete case analysis. Finally, further work should also be done on the contribution of culture in the spread of HIV/AIDS in Swaziland.

Bibliography

[1] L. Ackermann and G.W. de Klerk. Social factors that make South African women vulnerable to HIV infection. *Health Care for Women International*, 23(2):163–172, 2002.

[2] H. Acquah. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2(1):001–006, 2010.

[3] S. Ahmed, T. Lutalo, M. Wawer, D. Serwadda, N.K. Sewankambo, F. Nalugoda, F. Makumbi, F. Wabwire-Mangen, N. Kiwanuka, G. Kigozi, et al. HIV incidence and sexually transmitted disease prevalence associated with condom use: a population study in Rakai, Uganda. *Aids*, 15(16):2171, 2001.

[4] R.C. Bailey, S. Moses, C.B. Parker, K. Agot, I. Maclean, J.N. Krieger, C.F.M. Williams, R.T. Campbell, and J.O. Ndinya-Achola. Male cir-

cumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The Lancet*, 369(9562):643–656, 2007.

[5] T. Bärnighausen, V. Hosegood, I.M. Timaeus, and M.L. Newell. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS (London, England)*, 21(Suppl 7):S29, 2007.

[6] J. Boerma, M. Urassa, K. Senkoro, A. Klokke, et al. Spread of HIV infection in a rural area of Tanzania. *Aids*, 13(10):1233, 1999.

[7] W.J. Browne and D. Draper. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3):473–514, 2006.

[8] A.G. Buseh. Patterns of sexual behaviour among secondary school students in Swaziland, southern Africa. *Culture, health & sexuality*, 6(4):355–367, 2004.

[9] P. Congdon. *Applied Bayesian Hierarchical Methods*. Chapman & Hall, 2010.

[10] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

[11] S. Gillespie, S. Kadiyala, and R. Greener. Is poverty or wealth driving HIV transmission? *Aids*, 21:S5, 2007.

[12] J.R. Glynn, M. Carael, A. Buve, S. Anagonou, L. Zekeng, M. Kahindo, and R. Musonda. Does increased general schooling protect against HIV infection? a study in four African cities. *Tropical medicine & international health*, 9(1):4–14, 2004.

[13] H. Goldstein. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, pages 45–51, 1991.

[14] E. Gouws, K.A. Stanecki, R. Lyerla, and P.D. Ghys. The epidemiology of HIV infection among young people aged 15-24 years in southern Africa. *Aids*, 22:S5, 2008.

[15] R.H. Gray, G. Kigozi, D. Serwadda, F. Makumbi, S. Watya, F. Nalugoda, N. Kiwanuka, L.H. Moulton, M.A. Chaudhary, M.Z. Chen, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *The Lancet*, 369(9562):657–666, 2007.

[16] R.H. Gray, M.J. Wawer, D. Serwadda, N. Sewankambo, C. Li, F. Wabwire-Mangen, L. Paxton, N. Kiwanuka, G. Kigozi, J. Konde-Lule, et al. Population-based study of fertility in women with HIV-1 infection in Uganda. *The Lancet*, 351(9096):98–103, 1998.

[17] J.D. Hadfield. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22, 2010.

[18] J.R. Hargreaves, C.P. Bonell, L.A. Morison, J.C. Kim, G. Phetla, J.D.H. Porter, C. Watts, and P.M. Pronyk. Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001-2004. *Aids*, 21:S39, 2007.

[19] K. Johnson and A. Way. Risk factors for HIV infection in a national adult population: evidence from the 2003 Kenya Demographic and Health Survey. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 42(5):627, 2006.

[20] S.C. Kalichman, L.C. Simbayi, M. Kaufman, D. Cain, and S. Jooste. Alcohol use and sexual risks for HIV/AIDS in sub-Saharan Africa: systematic review of empirical findings. *Prevention Science*, 8(2):141–151, 2007.

[21] C. Kishamawe, D.C.J. Vissers, M. Urassa, R. Isingo, G. Mwaluko, G.J.J.M. Borsboom, H.A.C.M. Voeten, B. Zaba, J.D.F. Habbema, and

S.J. de Vlas. Mobility and HIV in Tanzanian couples: both mobile persons and their partners show increased risk. *Aids*, 20(4):601, 2006.

[22] I. Kleinschmidt, A. Pettifor, N. Morris, C. MacPhail, and H. Rees. Geographic distribution of human immunodeficiency virus in South Africa. *The American journal of tropical medicine and hygiene*, 77(6):1163–1169, 2007.

[23] T. Lawes. Evaluation of decentralised TB case finding for HIV positive patients: a mixed methods study in Lubombo, Swaziland. Master's thesis, Department of Health Sciences, 2010.

[24] D.V. Lindley and A.F.M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–41, 1972.

[25] B. Lopman, J. Lewis, C. Nyamukapa, P. Mushati, S. Chandiwana, and S. Gregson. HIV incidence and poverty in Manicaland, Zimbabwe: isHIV becoming a disease of the poor? *AIDS (London, England)*, 21(Suppl 7):S57, 2007.

[26] M.N. Lurie, B.G. Williams, K. Zuma, D. Mkaya-Mwamburi, G.P. Garnett, A.W. Sturm, M.D. Sweat, J. Gittelsohn, and S.S. Abdool Karim. The impact of migration on HIV-1 transmission in South africa: a study

of migrant and nonmigrant men and their partners. *Sexually Transmitted Diseases*, 30(2):149, 2003.

[27] M. Magadi and M. Desta. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 2011.

[28] W. Masanjala. The poverty-HIV/AIDS nexus in Africa: A livelihood approach. *Social science & medicine*, 64(5):1032–1041, 2007.

[29] K.H. Mayer and C. Beyrer. HIV epidemiology update and transmission factors: risks and risk contexts16th international AIDS conference epidemiology plenary. *Clinical Infectious Diseases*, 44(7):981, 2007.

[30] G.T. Mhlongo. Drug abuse in adolescents in Swaziland. *Unisa*, 2009.

[31] V. Mishra, S.B.V. Assche, R. Greener, M. Vaessen, R. Hong, P.D. Ghys, J. Boerma, A. Van Assche, S. Khan, and S. Rutstein. HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. *Aids*, 21:S17, 2007.

[32] M. Morris and M. Kretzschmar. Concurrent partnerships and the spread of HIV. *Aids*, 11(5):641, 1997.

[33] M. M. Ngigi. A geographical study on the HIV/AIDS pandemic in Kenya. Master's thesis, Graduate School of Life and Environmental Sciences, 2007.

[34] I. Ntzoufras. *Bayesian Modelling Using WinBUGS*. Wiley, 2009.

[35] D. Renard. *Topics in modeling multilevel and longitudinal data*. PhD thesis, Limburgs Universitair Centrum, 2002.

[36] G. Roli. *Hierarchical logistic regression in a multicentric study of multiple dietary effects on a disease outcome: a fully Bayesian approach*. PhD thesis, PHD thesis.(Available from http://www2. stat. unibo. it/Dottorato/MSRS/TesiDottoratoMSRS/2006/RoliGiulia. pdf), 2006.

[37] S. Rue and S. Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192, 2007.

[38] S. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

[39] SDHS. Swaziland health and demographic survey report. *Macro International Inc.*, 2008.

[40] SHDR. HIV /AIDS and culture. Technical report, Swaziland Human Development Report, 2007.

[41] O. Shisana, N. Zungu-Dirwayi, Y. Toefy, LC Simbayi, S. Malik, and K. Zuma. Marital status and risk of HIV infection in South Africa. *South African medical journal*, 94(7):537, 2008.

[42] S. Solomon, N. Kumarasamy, AK Ganesh, and R.E. Amalraj. Prevalence and risk factors of HIV-1 and HIV-2 infection in urban and rural areas in Tamil Nadu, India. *International journal of STD & AIDS*, 9(2):98–103, 1998.

[43] J. Trinitapoli. Religion and HIV risk behaviors among married men: Initial results from a study in rural sub-Saharan Africa. *Journal for the Scientific Study of religion*, Volume 45:505–528, 2006.

[44] S.D. Weiser, K. Leiter, D.R. Bangsberg, L.M. Butler, F. Percy-de Korte, Z. Hlanze, N. Phaladze, V. Iacopino, and M. Heisler. Food insufficiency is associated with high-risk sexual behavior among women in Botswana and Swaziland. *PLoS Medicine*, 4(10):e260, 2007.

[45] H.A. Weiss, M.A. Quigley, and R.J. Hayes. Male circumcision and risk of HIV infection in sub-SaharanAfrica: a systematic review and meta-analysis. *Aids*, 14(15):2361, 2000.

[46] A. Whiteside, A. Hickey, N. Ngcobo, and J. Tomlinson. What is driving the HIV/AIDS epidemic in Swaziland, and what more can we do about it. *Health Economics and HIV/AIDS Research Division (Durban, University of Natal)*, 2003.

Appendix

Code for Bayesian GLMM Using MCMCs

```
males=read.csv("males.csv")

attach(males)

males

library(MCMCglmm)

MC1 = MCMCglmm(hiv ~ 1+factor(age)+factor(region)+factor(residence)+factor(edulevel)

+factor(marital)+factor(asex)+factor(sti)+factor(partners)+factor(away),random=cluster,

data=males,family="categorical",prior=list(R=list(V=1,

fix=1), G=list(G1=list(V=1, nu=0.001))),nitt=60000,slice=F,thin=10)

summary(MC1$VCV) # prints the random effects

summary(MC1$Sol) # prints the fixed effects

plot(MC1)

summary(MC1)
```

```
    females=read.csv("females.csv")
attach(females)
females
library(MCMCglmm)
MC1 = MCMCglmm(hiv ~ 1+factor(age)+factor(region)+factor(residence)+factor(edulevel)
+factor(marital)+factor(asex)+factor(sti)+factor(partners)+factor(away)+factor(birth),
random = cluster,data=females,
family="categorical",prior=list(R=list(V=1,
fix=1), G=list(G1=list(V=1, nu=0.001))),nitt=60000,slice=F,thin=10)
summary(MC1$VCV) # prints the random effects
summary(MC1$Sol) # prints the fixed effects
plot(MC1)
summary(MC1)
```

Code for Bayesian GLMM Using INLA

```
males=read.csv("males.csv")
attach(males)
library(INLA)
```

```
names(males)
fit=inla(hiv ~ factor(age)+ factor(region)+factor(residence)+
factor(edulevel) + factor(marital) +factor(asex) + factor(sti) +
factor(partners) + factor(away)+factor(birth) +
f(cluster, model="iid",param=c(0.001,1)),data=males,
family="binomial",Ntrials=1)
summary(fit)
plot(fit)

    females=read.csv("females.csv")
attach(females)
library(INLA)
names(females)
fit=inla(hiv ~ factor(age)+ factor(region)+factor(residence)+
factor(edulevel) + factor(marital) +factor(asex) + factor(sti) +
factor(partners) + factor(away)+factor(birth)
+ f(cluster, model="iid",param=c(0.001,1)),
data=females, family="binomial",Ntrials=1)
summary(fit)
plot(fit)
```

Code for Complete Pooling Model

```
males=read.csv("males.csv")

attach(males)

males

names(males)

library(multilevel)

library(lattice)

library(lme4)

fit=glm(formula = hiv ~ factor(age)+factor(region)+

factor(residence)+factor(edulevel)+factor(marital)

+factor(asex)+factor(sti)+factor(partners)+factor(away)+ cwealth+

chiv+cage, family = binomial(link = "logit"))

summary(fit)

confint(fit)

BIC(fit)


    females=read.csv("females.csv")

attach(females)

females

names(females)
```

```
library(multilevel)

library(lattice)

library(lme4)

fit=glm(formula = hiv ~ factor(age)+factor(region)+

factor(residence)+factor(edulevel)+factor(marital)

+factor(asex)+factor(sti)+factor(partners)+factor(away)+

factor(birth) + cwealth+chiv+cage,

family = binomial(link = "logit"))

summary(fit)

confint(fit)
```

Code for the Frequentist Multilevel Model

```
males=read.csv("males.csv")

attach(males)

males

names(males)

library(multilevel)

library(lattice)

library(lme4)
```

```
cwealth.full=cwealth[cluster]

chiv.full=chiv[cluster]

cage.full=cage[cluster]

fit=lmer(formula = hiv ~ factor(age)+factor(region)+

factor(residence)+factor(edulevel)+factor(marital)

+factor(asex)+factor(sti)+factor(partners)+factor(away)+

cwealth.full+chiv.full+cage.full + (1—cluster)

,family = binomial(link = “logit”))

summary(fit)


    females=read.csv(“females.csv”)

attach(females)

females

names(males)

library(multilevel)

library(lattice)

library(lme4)

cwealth.full=cwealth[cluster]

chiv.full=chiv[cluster]

cage.full=cage[cluster]

fit11=lmer(formula = hiv ~ factor(age)+factor(region)+
```

```
factor(residence)+factor(edulevel)+factor(marital)
+factor(asex)+factor(sti)+factor(partners)+factor(away)+
factor(birth) + cwealth.full+chiv.full+cage.full + (1—cluster)
, family = binomial(link = “logit”), data= females)
summary(fit11)
```

Code for the Bayesian Multilevel Model

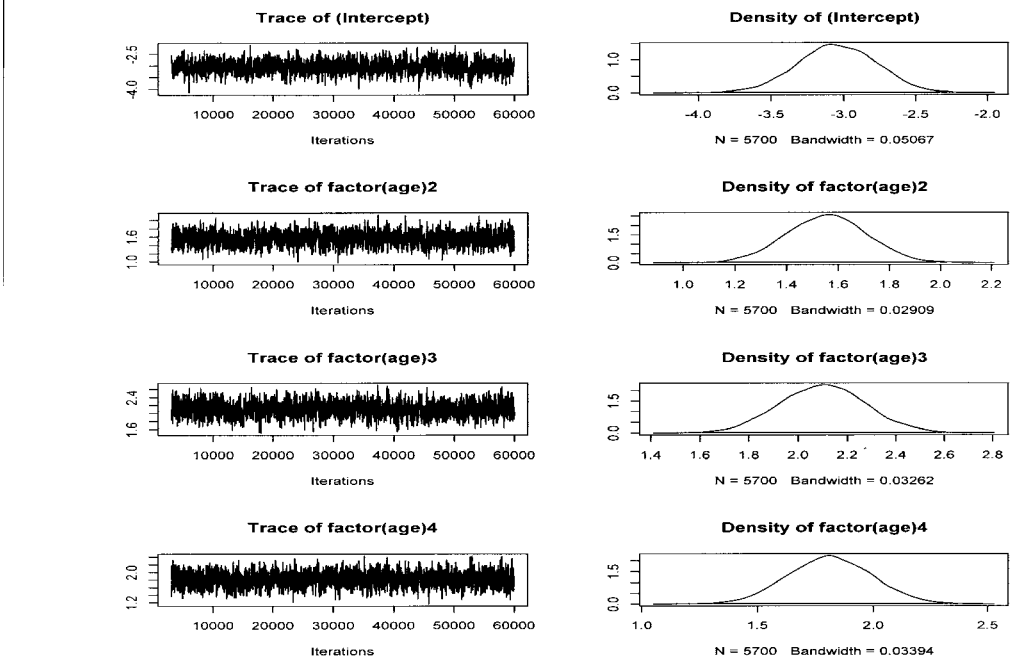
```
cwealth.full=cwealth[cluster]
chiv.full=chiv[cluster]
cage.full=cage[cluster]
M1=MCMCglmm(hiv ∼ 1+factor(age)+factor(region)+
factor(residence)+factor(edulevel)+factor(marital)
+factor(asex)+factor(sti)+factor(partners)+factor(away)+
factor(birth)+cwealth.full+
chiv.full+cage.full,random=cluster,data=females,
family=”categorical”,prior=list(R=list(V=1,fix=1),
G=list(G1=list(V=1, nu=0.001))),nitt=60000,slice=F,thin=10)
summary(M1$VCV)
summary(M1$Sol)
```

```
plot(M1)

summary(M1)


    males=read.csv("males.csv")

attach(males)

males

library(MCMCglmm)

cwealth.full=cwealth[cluster]

chiv.full=chiv[cluster]

cage.full=cage[cluster]

MC1=MCMCglmm(hiv ~1+factor(age)+factor(region)+

factor(residence)+factor(edulevel)+factor(marital)+factor(asex)+

factor(sti)+factor(partners)+factor(away)+cwealth.full+

chiv.full+cage.full ,random=cluster,data=males,

family="categorical",prior=list(R=list(V=1,fix=1),

G=list(G1=list(V=1, nu=0.001))),nitt=60000,slice=F,thin=10)

summary(MC1$VCV)

summary(MC1$Sol)

plot(MC1)

summary(MC1)
```

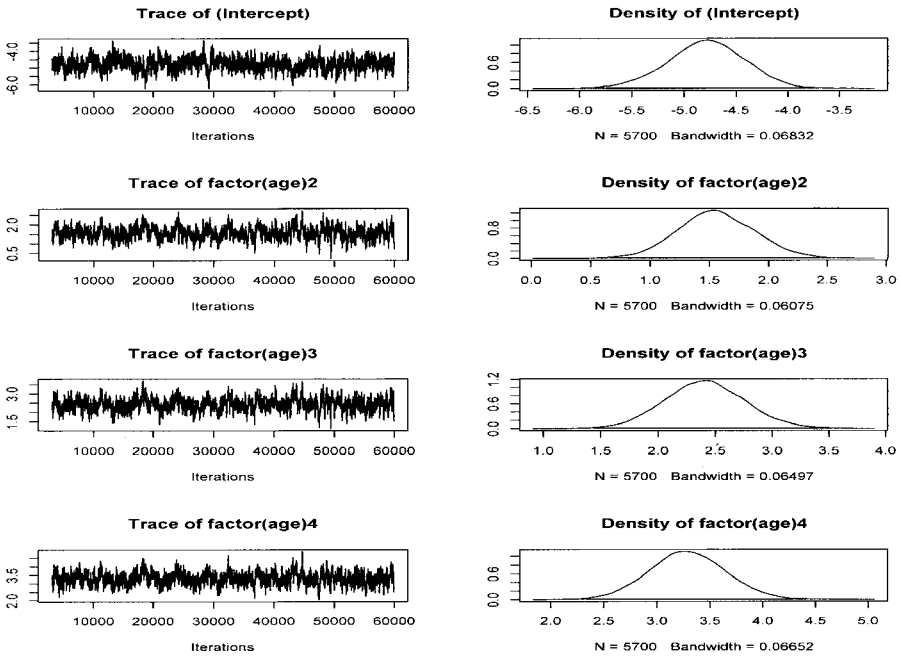
Trace Plots

In this section we present a sample of the trace plots for the two Bayesian models.

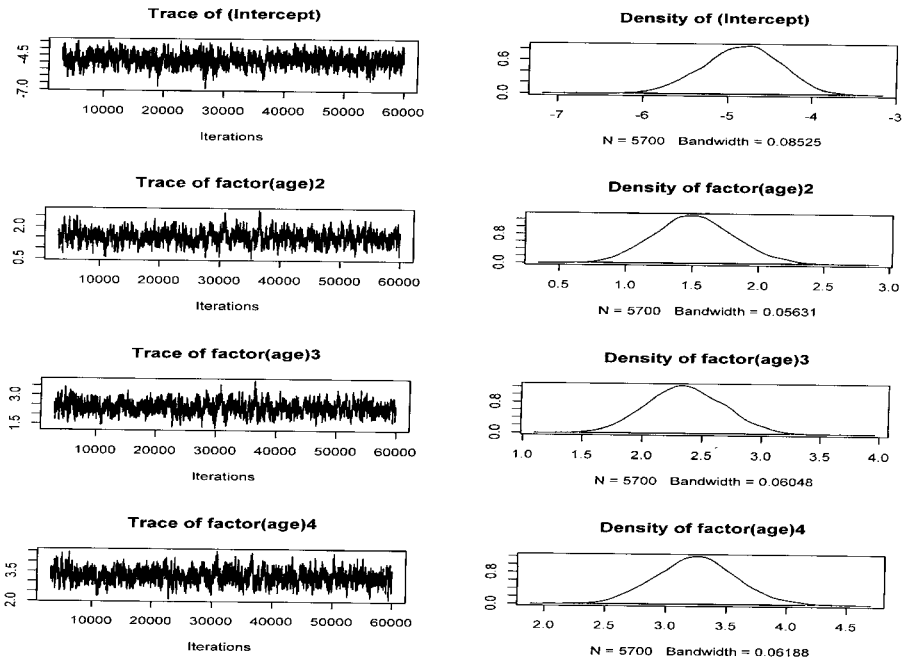
Bayesian GLMM Trace Plots for Females



Bayesian GLMM Trace Plots for Males



Bayesian Multilevel Model Trace Plots for Males



Bayesian Multilevel Model Trace Plots for Females