2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Investigating factors associated with HIV among 15 to 24 year old females

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Prof. KHANGELANI ZUMA

Thembile Mzolo

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Investigating factors associated with HIV among 15 to 24 year old females

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Prof. KHANGELANI ZUMA

Thembile Mzolo

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Acknowledgements

First of all I would like to extend my gratitude to VLIR who without them I would not be here. I would also like to thank my supervisor Prof. dr. Ziv Shkedy. I learnt a lot from you, and I am positive that we will continue working together in future, and also thank you for your time that you dedicated on my summer project. This work would not have been possible without HSRC who allowed me to use their data set, and I would also like to thank Dr Khangelani Zuma for the time you spent on my thesis.

I am grateful to my parents for the support and love they've given me throughout the years. To my niece and nephews I am sorry for spending most of the time away from you, and I am sorry for missing you birthdays, school farewells' and functions and all the events that I couldn't attend but always remember your aunt loves you so much! Lastly, thank you so much Wami for being there all the time when I needed you most!

Abstract

Background: Among the sub-saharan African countries South Africa has the highest number of people who are living with HIV and AIDS in the world. HIV prevalence have remain high among young females. In this study we focus on the factors associated with HIV among females in age group 15 to 24 years where the outcome variables of interest are HIV status and the perceived risk of being infected.

Methods: Data used is from a third surveillance survey conducted by Human Sciences Research Council in South Africa in 2008. The survey included a master sample of 1 000 enumeration areas (communities) and 15 households were selected per enumeration area. Within a household, at most 4 eligible individuals were selected. In this study we focus on females in age group 15 to 24 who participated in the survey totaling 2 815. Generalized Linear Mixed models are used to study this association while accounting for the survey design where the univariate and bivariate models are fitted where the communities are used as clusters.

Results: Africans, 20 to 24 year old, lack of condom use at sexual debut, having sexual relationship with older individuals are all associated with an increased HIV infection as well as being at a higher perceived risk. Furthermore, those with lower education with a low perceived risk are at higher risk of HIV. Whilst 20 to 24 year old who has been in their current relationship longer than a year are likely to be at higher perceived risk. Those having a sexual relationship with older individuals are more likely to have a higher perceived risk. The outcomes are jointly associated in a sense that when the perceived risk increase, there will be an increase in HIV infection too.

Conclusion: More work is still needed to be done at community level in order to win the fight against HIV. In addition, policy-makers must pay more attention on teaching the youth about preventive measures that are available to them.

Keywords: *Clustered data, Generalized Linear Mixed models, HIV, Joint modelling, Perceived risk, Survey data.*

Contents

Acknowledgements ii
Abstract iii
List of Figures vi
List of Tables vii

1 Introduction **1**
1.1 Research hypothesis 2
1.2 Thesis overview 2

2 Methodology **3**
2.1 The Data description 3
2.2 Statistical Methodology 4
2.2.1 Generalized Linear Mixed models for binary outcomes 5
2.2.1.1 Univariate Binary models 7
2.2.1.2 Other statistical models 8
2.2.2 Generalized Linear Mixed models for joint outcomes 8
2.2.2.1 Bivariate models 9
2.2.3 Multiple Imputation 10
2.2.4 Statistical packages 10

3 Results **11**
3.1 Exploratory Data Analysis 11
3.2 Statistical Results 14
3.2.1 Univariate Models 14
3.2.1.1 HIV model 14
3.2.1.2 Perceived Risk model 16
3.2.2 Other statistical models 18
3.2.2.1 Ignoring design weights 18
3.2.2.2 Cluster size used as weights 20
3.2.3 Bivariate model 21
3.2.4 Complete data analysis 22

4 Discussion and Conclusion **25**
4.1 Discussion 25
4.2 Conclusion 29

4.3 Future Research work 30

Bibliography **31**

Appendices:

A Descriptive results **A-1**

B Univariate results **B-1**

C Bivariate results **C-1**

List of Figures

2.1 *Master sample for SABSSM III* 4

3.1 *Participation and HIV prevalence by Age and Race groups* 12
3.2 *Participation and HIV prevalence by Geographical area and Highest Education qualification* 12
3.3 *Interaction of education and perceived risk for the HIV model* 16
3.4 *Interaction of Age and Duration of a relationship for the perceived RISK model* 18
3.5 *Empirical Bayes estimates for the bivariate model with correlated random effects* 22

B.1 *Empirical Bayes estimates for HIV model with design weights* B-1
B.2 *Empirical Bayes estimates for the perceived risk model with design weights* B-1

List of Tables

3.1 *Descriptive statistics for continuous variables* 11
3.2 *Results from the HIV model with design weights* 15
3.3 *Results from the perceived risk model with design weights* 17
3.4 *Results from the HIV model without weights* 19
3.5 *Results from the perceived risk model without weights* 19
3.6 *Results from the HIV model with cluster size as weights* 20
3.7 *Results from the perceived risk model with cluster size as weights* . . . 20
3.8 *Random effects parameter estimates for the Bivariate models under different covariance structures* 21
3.9 *Parameter estimates for the HIV model-Available cases and Multiple imputation* 23
3.10 *Parameter estimates for the risk model-Available cases and Multiple imputation* 23

A.1 *Socio-Demographic factors by HIV and perceived risk* A-1
A.2 *Behavioural factors by HIV and perceived risk* A-2

C.1 *Comparison of the random effects structures for the Bivariate models* C-1
C.2 *Parameter estimates for the Bivariate models* C-1

Chapter 1

Introduction

In the world 50% of people living with HIV (Human Immune Virus) and AIDS (Acquired Immunodeficiency Syndrome) are females (WHO, 2009). In sub-Saharan Africa, 61% of all people living with HIV are women. People living in this region are the most vulnerable population in the world Hodge and Roby (2010). Young women (15-24 years) are three times more likely to be infected than men in the same age group WHO (2009). Among countries in the sub-Saharan Africa, South Africa has the highest number of people who are infected with HIV in the world with more that 5 million people living with the disease whilst Swaziland has the highest HIV adult prevalence in the world (25.9%) UNAIDS (2010); Shisana et al. (2009).

HIV incidence has decreased, but the total number of people living with HIV continues to rise. AIDS related deaths has also decreased due to the introduction of the anti-retroviral therapy (ART) which can prolong an individuals life expectancy (Jahn et al., 2008). The most common modes of transmission are through heterosexual sex, men who have sexual encounters with other men and injecting drug users are also at higher risk of HIV transmission. However, in South Africa the most common mode of transmission is through heterosexual sex (Shisana et al., 2009). Indicators that are related to sexual behaviour risks for HIV infection are age at sexual debut, multiple sexual partnerships, unprotected sexual intercourse, and age mixing to mention a few (Shisana et al., 2009).

It has been reported that young males tend to have early sexual debut as compared to women and multiple partnerships are more common among those who had an early sexual debut (Zuma et al., 2010). There are several studies that have been studied to further understand the distribution of HIV in South Africa (Peltzer et al., 2010, 2009; Zuma et al., 2010; Mzolo, 2009). Women are at higher risk compared to men which may be due to gender-related factors that contribute to the spread of HIV; these include increased sexual violence, economic security, orphanage and poverty (Worth, 1990). Young women's ability to practice safe sex is inhibited by their partners' demands in those relationships where there is an imbalance of power (Eaton et al., 2003). Young girls are exposed to sexual abuse, rape and commercial sex activities for survival which exacerbates the risk of being infected with HIV.

In addition, young women tend to have sexual relationships with older men (who are at high risk of HIV) in which such relationship are mainly for material gain (Leclerc-Madlala, 2008). Transactional sex is a huge problem among this population where young girls will have sexual relations mostly for economic security, and thus putting themselves at high risk of being infected. As a result, this has a huge negative impact on them as in most cases they do not have a say as to whether protection (such as condoms) should be used or not (Pettifor et al., 2005).

The most commonly used method of prevention is by the use of condoms. These do not only guard against HIV infection but also unnecessary pregnancy as well as other sexually transmitted diseases (STIs). Infection with HIV is also a significant risk factor of STIs possibly due to reduced immunity (Zuma et al., 2005). In South Africa three national surveys for HIV have been conducted so far (2002, 2005 and 2008) which were all conducted by the Human Sciences Research Council (HSRC) (Shisana et al., 2009, 2005; Shisana and Simbayi, 2002). From all these surveys HIV prevalence among females in the 15 to 24 years age group has been higher than that for males (Shisana et al., 2009).

In the study the main focus is on young females in age group 15 to 24 years using the 2008 survey data. Before these surveys information about HIV was only obtained from the anti-natal clinic attendees. The main disadvantages about these studies is that they are only focused on child-bearing women and not the whole population.

1.1 Research hypothesis

Most intervention programmes have been targeting youth however no empirical analysis of their impact on HIV has been achieved. Thus, a comprehensive analysis and understanding of sexual behaviour and determinants of HIV among females in the age group 15 to 24 years is critical. The aim of this study is to identify factors that are associated with HIV among 15 to 24 year old females and further assess the impact of these factors. This is done by using the HIV status (positive or negative) as well as the perceived risk of HIV (high risk or low risk) as the outcomes of interest. In addition the association between the outcome variables is studied.

1.2 Thesis overview

The thesis is structured as follows: The data description as well as the statistical methodology are described in Chapter 2. Results which include exploratory data analysis and results from the statistical methods are presented in Chapter 3. Discussion of the results is in Chapter 4 as well as conclusions drawn from the analysis. Furthermore, possible future research are explained in this chapter also.

Chapter 2

Methodology

2.1 The Data description

Data that will be used for this study is from a cross-sectional population-based household survey which was conducted using a multi-stage stratified sampling approach by the HSRC in 2008. The 2008 survey is the third survey conducted by HSRC, the first was in 2002 and the second in 2005. This survey included individuals of all ages living in South Africa. All persons living in the selected household were eligible to participate including those living in hostels but excluding individuals staying in educational institutions, old-age homes, hospitals, homeless people, and uniformed-service barracks.

A multi-stage disproportionate, stratified sampling approach was used. A total of 1 000 census enumeration areas[1] (EAs) from the 2001 population census were selected from a database of 86 000 EAs. The selection of EAs was stratified by province and locality type where locality types were identified as urban formal, urban informal, rural formal (including commercial farms), and rural informal. In formal urban areas, race was also used as a third stratification variable based on the predominant race group in the selected EA.

The selected 1 000 EAs formed the primary sampling units. These EAs are indicated in Figure 2.1 by red dots and from this map it can been seen that more EAs are found in highly populated cities, for example Johannesburg, Durban and Cape Town. Within an EA, a random sample of 15 households or visiting points were selected as the secondary sampling units. The ultimate sampling units were individuals who were eligible to be selected within a household. Thus in total 15 households were selected from each EA yielding a total of 15 000 households.
From each household, only one person within each age group was selected subject to there being at least one eligible person in the specified age group. Four mutually

[1]Enumeration area (EA) is a spatial area used by Statistics South Africa (StatsSA) to collect census information on the South African population. It consists of approximately 180 households in urban areas and 80 - 120 households in a deep rural areas and is considered to be a small enough sample size for one person to collect census information for StatsSA.

exclusive age groups were used for sampling respondents: under 2 years, 2 - 14 years, 15 - 24 years and 25 years and above. Among the valid households that agreed to participate in the survey, 23 369 individuals were eligible to be interviewed and among these, 20 826 participated in the study. Among individuals who participated in the study 15 031 agreed to provide blood specimen for HIV testing. This study focuses on females aged 15 to 24 years who participated in the survey (n=2 815).

Owing to sampling design of the survey, some individuals have a greater or lesser probability of selection than others. To correct for this problem, sample weights are introduced to correct for bias at the EA, household, and individual levels and also adjust for non-response. More about the sampling weights can be found in (Shisana et al., 2009). In this study variables of interest include the demographic factors, socio-economic factors as well as behavioural factors.

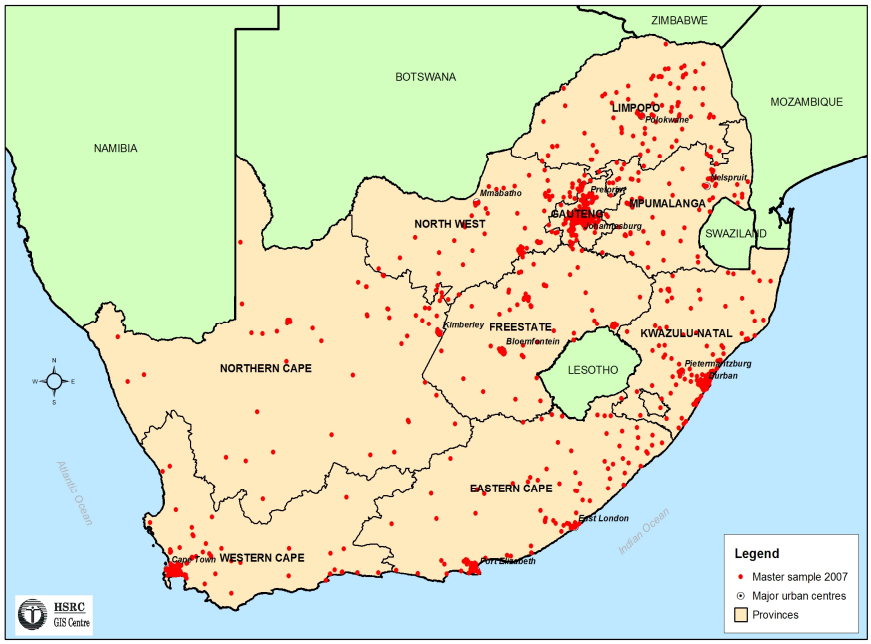

Figure 2.1: *Master sample for SABSSM III*

2.2 Statistical Methodology

Most epidemiological studies involve complex designs, which makes it easy to conduct such studies (Barros and Hirakata, 2003). These include cross-sectional studies, longitudinal studies and others. However these complex designs are expensive to conduct and standard statistical methodology is no longer straightforward. For binary outcomes logistic regression is a simple method that is usually used, but for such studies this is no longer possible (McCullagh and Nelder, 1989). This is due to the fact that a hierarchical design is used for these studies and this induces correlation between and within clusters.

Multi-level modeling is the key statistical technique of relevance for this hierarchical design. This modeling approach is desirable because it allows relationships across and within hierarchical levels of a multi-stage design to be explored, taking account of the variability at different levels (Goldstein, 2003; Snijders and Bosker, 1999; Kreft and de Leew, 1998) where in this case these levels are the communities, i.e EAs. Several models are available that can be used to account for this correlation, these include marginal and subject-specific models (Carriere and Bouyer, 2002). In this study Generalized Linear Mixed Models (GLMMs) methodology will be applied.

In surveys several information is missing, which could depend on the sensitivity of a question (e.g. question about income), refusal, non-response. Survey non-response arises from a census or sample survey whenever the population consists of units such as individuals or households. As a result data values intended by the survey design to be observed are missing and this leads to less efficient estimates due to the reduced size of the data and standard data methods cannot be immediately used to analyze the data (Rubin, 1987).

The GLMMs fall under the direct likelihood method, which is valid when there is missing data (Molenberghs and Verbeke, 2005). However, Multiple imputation (MI) will be used to take into account missingness and results will be compared to the GLMMs results.

2.2.1 Generalized Linear Mixed models for binary outcomes

The most frequently used random effects model for discrete outcomes is the generalized linear mixed model (Molenberghs and Verbeke, 2005). It is a straightforward extension on the generalized linear model (GLM) to the context of clustered measurements where the random effects are added in the mean structure. The random effects incorporate correlation between the repeated observations within each cluster and variation between clusters, resulting in GLMMs (Wu, 2010). It is assumed that correlation arises among repeated observations within a given cluster because of the shared random effects, but these repeated observations are assumed to be conditionally independent given the random effects (Wu, 2010).

Let $\mathbf{y_i} = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ be the n_i repeated observations of the response within cluster i, $i = 1, 2, \ldots, n$. We assume that, conditioning on the random effects $\mathbf{b_i}$, the repeated measurements $\mathbf{y_i} = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ are independent and each follows a distribution in the exponential family. A general GLMM can be written as

$$g(\mu_{ij}) = \mathbf{x}_{ij}^T\beta + \mathbf{z}_{ij}^T\mathbf{b}_i \tag{2.1}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}) \tag{2.2}$$

where $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. $\mu_{ij} = E(y_{ij}|\beta, \mathbf{b_i})$ is the conditional mean. $\mathbf{x_{ij}}$ and $\mathbf{z_{ij}}$ are vectors containing covariates. β is a vector containing fixed effects, $\mathbf{b}$ is a vector of random effects which are assumed to be normally distribution with mean zero and variance-covariance matrix $\mathbf{D}$. $g(\cdot)$ is a known function which links the mean and the linear form of predictors called the link function (McCullogh and Searle, 2001). GLMMs are sometimes called conditional models or subject-specific models, since the model is specified based on the conditional mean. (Wu, 2010)

Statistical inference for a GLMM is typically based on the likelihood method. In GLMM (2.1), the marginal distribution for $\mathbf{y_i}$ is

$$f(\mathbf{y}_i|\beta, \mathbf{D}) = \int \prod_{j=1}^{n_i} \big[f(y_{ij}|x_{ij}, z_{ij}, \beta, \phi, \mathbf{b}_i) f(\mathbf{b}_i|\mathbf{D}) \big] d\mathbf{b}_i \tag{2.3}$$

which usually does not have an analytic or closed-form expression since the model is nonlinear in the random effects $\mathbf{b_i}$. The likelihood for all observed data is given by

$$L(\beta, \mathbf{D}|y) = \prod_{i=1}^{n} \Big[\int \prod_{j=1}^{n_i} \big[f(y_{ij}|x_{ij}, z_{ij}, \beta, \phi, \mathbf{b}_i) f(\mathbf{b}_i|D) \big] d\mathbf{b}_i \Big] \tag{2.4}$$

The above likelihood involves an intractable multi-dimensional integral with respect to the random effects which is due to the presence of n integrals over a q-dimensional random effects (Molenberghs and Verbeke, 2005). There are various numerical approximations that can be used to maximize the likelihood. These numerical approximations include those that are based on approximating the integrand, approximating the data, and integrating the integral itself (Wu, 2010; Molenberghs and Verbeke, 2005).

The most commonly used inference for GLMMs include methods based on Gauss-Hermite quadrature or Monte Carlo integration techniques, EM algorithms, and approximate methods based on Taylor approximations or Laplace approximations (Lee et al., 2006; Molenberghs and Verbeke, 2005; Breslow and Clayton, 1993). The penalized quasi-likelihood (PQL) is an approximate method which tends to be biased for non-Gaussian responses, especially for binary responses (Wu, 2010; Joe, 2008; Breslow and Clayton, 1993).

Approximate methods which avoid integrations are computationally much more efficient, however these methods can be computationally intensive when the dimension of the random effects is large (Wu, 2010; Molenberghs and Verbeke, 2005). These numerical methods include Gaussian Quadrature and Adaptive Gaussian Quadrature where the former is less precise but less time consuming and the latter is precise but much more time consuming (Molenberghs and Verbeke, 2005).

It is common that interest is on estimating parameters in the marginal distribution of $\mathbf{Y_i}$, however it is also necessary to obtain estimates for the random effects.

These reflect between-cluster specific variability, which makes them more helpful for detecting special cases, such as outlying observations or a group of individuals evolving differently. These estimates are needed when interest is in prediction of subject-specific evolutions. Estimation of the random effects will be based on their posterior distributions and obtained estimates are called the Empirical Bayes (EB) estimates.

Inference on random effects is also of interest and classical methods can be used to test the significance of these effects. However, these classical methods can only be used if the hypotheses to be tested are not on the boundary of the parameter space, i.e. restricted to positive values only. Therefore, under the null hypothesis, the test statistic follows the positive normal distribution in 50% of the cases and this gives the mixture of chi-square distribution. (Molenberghs and Verbeke, 2005; Verbeke and Molenberghs, 2000)

2.2.1.1 Univariate Binary models

Models are fitted taking into account of the design weights for both HIV and perceived risk outcomes. Some modifications are required for models which account for design weights as a result that using raw design weights tends to lead to numerical difficulties. Several approaches have been proposed to account for such problems. These include normalizing or re-scaling of the design weights. Normalizing weights means that each sample weight is divided by the mean of the final weight for the entire sample (Heeringa et al., 2010). These normalized weights will have a mean value of one and the normalized weights for all sample cases should add up to the sample size. This is the approach that will be followed in this report.

Another method that can be used is to scale design weights so that the new weights sum to the total cluster size (Carle, 2009). A second approach involves scaling design weights so that the new weights sum to the effective cluster size (Carle, 2009). However, it is important to note that there is no fixed method that can be applied. When the cluster size increases the estimates become less biased (Asparouhov, 2006; Pfeffermann et al., 1998) thus, scaling of weights may not be that important.

Models that are fitted are formulated as follow: Let Y_{ij} be the binary outcome (0/1) for individual j in community i. It is assumed that

$$Y_{ij}|\mathbf{b}_i \sim \text{Bernoulli}(\pi_{ij})$$

where $\pi_{ij} = Pr(Y_{ij} = 1)$ is the probability of being infected for the j^{th} individual in the i^{th} community. The general model is given by equation (2.5), where $g(\cdot)$ is the link function in this case a logit link.

$$\pi_{ij} = g^{-1}(\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{b}_i) \tag{2.5}$$

The parsimonious random-intercept models are presented in Chapter 3. In all these models the random effects are assumed to be normally distributed. Adaptive Gaussian Quadrature with 100 quadrature points will be used to numerically approximate the likelihood and the Newton-Raphson method will be used as an optimization technique. The combination of these methods produce the most reliable results (Molenberghs and Verbeke, 2005).

2.2.1.2 Other statistical models

In addition to models which account for weights, two additional models will be fitted. The first model is fitted ignoring design weights from the survey. The second model uses the cluster size as weights. This approach is used when the cluster size is informative; that is the response among observations in a cluster is associated with the cluster size. This approach was proposed by (Williamson et al., 2003) for marginal models where he stressed that if the cluster size is informative, standard marginal models will provide parameter estimates that are weighted for clusters.

In this study a similar approach is followed but in the random-effects model, that is the GLMM. These results will be compared to the results when accounting for the design weights.

2.2.2 Generalized Linear Mixed models for joint outcomes

From the exploratory analysis it was found that the two outcome variables are significantly associated. Therefore, univariate analyses may not be enough and bivariate analysis might be necessary to further understand the distribution of the two outcomes. This will enables us to study the association and draw joint inferences about different outcomes.

Joint modelling in most cases is required because the association structure between the outcomes is of interest, or the researcher may be interested in studying how the association between outcomes evolves over time or how outcome specific evolutions are related to each other (for a longitudinal study) (Feuws and Verbeke, 2004). Fitting these models can become very cumbersome, unless under unrealistically strong assumptions (Molenberghs and Verbeke, 2005).

Different joint modelling approaches exist such as analyzing one response while conditioning on other response. The disadvantage of this approach is that one has to choose a response to condition on it. Random effects approach for the joint modelling of multivariate longitudinal profiles received a lot of attention in recent publications and it is a flexible solution to model the association between the different responses using random effects (Feuws and Verbeke, 2004). Models for multivariate binary data are a currently developing area with still limited literature.

A special case of the joint model is a shared-parameter model where the same set of random effects is assumed for all outcomes. This model is advantageous since it has a low dimension of random effects compared to the high dimensional model. In high dimension model, the dimension of random effects increases with the number of outcomes, whilst in the shared-parameter this is not the case. Although the shared-parameter is superior to high dimension models, it also has some demerits as it is based on much stronger assumptions about the association between the outcomes, which may not be valid. (Molenberghs and Verbeke, 2005)

2.2.2.1 Bivariate models

A joint model is fitted under three different assumptions about the random effects. However it is important to note that models can have different mean structures as was the case in the current study. This is highlighted by different labels for fixed effects, where α is the vector of fixed effects for the HIV model, β the vector of fixed effects for the perceived risk model and $g(\cdot)$ is the logit link function.

Case 1: Common random effects

In this section the two models are assumed to be perfectly correlated. This is done by assuming that HIV and perceived risk share the same random effects. This in turn implies that when there is a change in one response the other response will change towards the same direction as the first one. Using the similar formulation as equation (2.5) the model is now written as follows:

$$g(\pi_{ijk}) = \begin{cases} \mathbf{X}\alpha + \mathbf{b}_i & \text{HIV if k} = 1 \\ \mathbf{Z}\beta + \mathbf{b}_i & \text{Risk if k} = 2 \end{cases}$$

Case 2: Common random effects with a scale parameter

The assumption that the two responses are perfectly correlated maybe be misleading, thus in this section this assumption is relaxed. This is done by still assuming that the responses share the same random effects but with different variances, where a scaling parameter (ν) is used. In general this means that $\sigma^2_{HIV} = \nu^2\sigma^2_{Risk}$. Under this assumption the model is given by:

$$g(\pi_{ijk}) = \begin{cases} \mathbf{X}\alpha + \mathbf{b}_i & \text{HIV if k} = 1 \\ \mathbf{Z}\beta + \nu\mathbf{b}_i & \text{Risk if k} = 2 \end{cases}$$

Case 3: Different random effects

So far it was assumed that the models share similar random effects (Case 1 and 2), in this section the assumption is now relaxed to explore other available cases.

The responses are assumed to have different random effects $\mathbf{a_i}$ and $\mathbf{b_i}$ for HIV and perceived risk, respectively.

$$g(\pi_{ijk}) = \begin{cases} \mathbf{X}\alpha + \mathbf{a}_i & \text{HIV if k} = 1 \\ \mathbf{Z}\beta + \mathbf{b}_i & \text{Risk if k} = 2 \end{cases}$$

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \sim N\begin{pmatrix} \mathbf{0} & , & \mathbf{D} \end{pmatrix}$$

Different assumptions are made about the random effects. The first assumption is that the two random effects are independent, the second assumption is that the two are correlated. This approach was applied by (Del Fava et al., 2011) where in addition to the above mentioned structures, a toeplitz structure was used. In the current study this was not done since this structure is only meaningful on studies where time points are equally spaced which is not the case for this study (Verbeke and Molenberghs, 2000). These variance-covariance matrices are given by $\mathbf{D_1}$ and $\mathbf{D_2}$ as indicated

$$\mathbf{D_1} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}, \quad \mathbf{D_2} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

2.2.3 Multiple Imputation

MI is relevant to all problems of missing data and the broadest definition of survey non-response is accepted. Missing values are filled in m times to generate m complete data sets. These are generated from a plausible model which is based on a plausible set of parameters drawn from a sampling distribution of the parameter estimates. Results from the analyses are combined for inferences and this results in valid statistical inferences that properly reflect the uncertainty due to missingness, that is, valid confidence intervals for parameters. (Longford, 2005; Rubin, 1987)

2.2.4 Statistical packages

The analyses will be done using SAS 9.2. Graphical outputs will be done using both R 2.13.1 and Microsoft Excel® 2007.

Chapter 3

Results

In this section different analyses are presented. The first being the exploratory data analysis (EDAs) where the data is explored thoroughly to understand it better. Secondly statistical results will be presented that were obtained by applying statistical methods as mentioned in Chapter 2.

3.1 Exploratory Data Analysis

The data used contains 2 815 young females who participated in the survey. The average age of females included in the study is 19 years old (Table 3.1). Among these females 63% of them were Africans, a big proportion of them are from urban areas where 58% reside in urban formal, 13.6% from urban informal, 23% from tribal areas and 5.4% from rural informal areas (Figure 3.2). Among individuals who agreed to be interviewed 71% (1 990) of them gave blood specimens for HIV testing and from these individuals 11.78% of the blood specimens tested positive for HIV. The average age at sexual debut was found to be 17 years and the minimum being 12 years of age. In addition, the average age of the male partner at sexual debut is 20 years old and the maximum being 52 years old as indicated in Table 3.1.

Table 3.1: *Descriptive statistics for continuous variables*

Variable	N	Mean	STD DEV	Min	Max
Age	2815	19.53	2.85	15	24
Age at sexual debut	1345	17.30	1.93	12	23
Age of the partners at sexual debut	1334	20.40	3.44	14	52
Number of partners in the past 12 months	1107	1.15	1.34	0	29

Figure 3.1 and Figure 3.2 show the descriptive statistics for the socio-demographic factors. Cross-tabulations were used to study the association of the variables with outcome variables and the results are tabulated in Table A.1 in Appendix A. The overall HIV prevalence was estimated to be 13.9% and 40.3% for perceived risk (those with a high perceived risk). These statistics vary among different potential factors as indicated in the descriptive statistics. HIV and preceived risk prevalence

was found to be higher among Africans compared to non-Africans (Figure 3.1). Age of the individuals was categorized to two levels due to the fact that a big difference was observed among those who are younger than 20 years as compared to the older ones. From these levels, it was observed that the HIV prevalence is higher among 20 to 24 year olds (21% vs 6.8%). Similarly perceived risk was higher among older females compared to the younger ones.

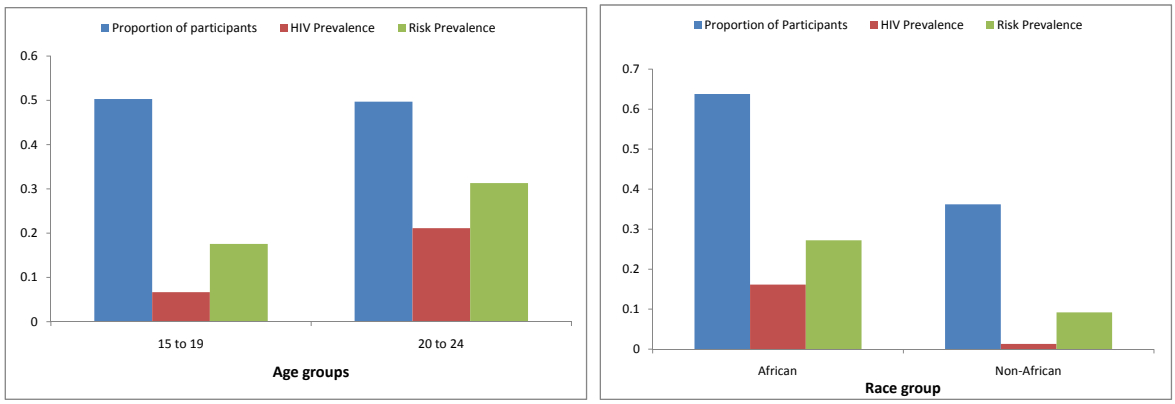

Figure 3.1: *Participation and HIV prevalence by Age and Race groups*

Looking at the geographical nature of the EA it was observed that HIV prevalence is much higher in urban informal, rural formal, tribal area and urban formal, respectively as also can be seen in Figure 3.2. However, for the risk prevalence a similar trend was observed as for HIV except that the prevalence for those residing in tribal areas is higher than those in rural informal. HIV prevalence was found to be higher among individuals who have no or primary education (31.7%), and it is lower among those with tertiary education as their highest qualification (Figure 3.2). A similar behaviour was observed for the perceived risk prevalence.

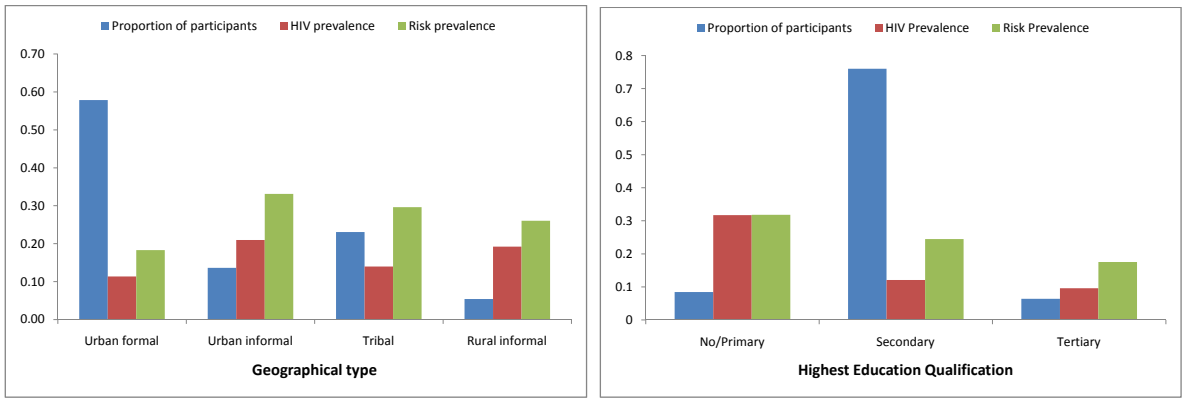

Figure 3.2: *Participation and HIV prevalence by Geographical area and Highest Education qualification*

A higher proportion of individuals who participated in the survey are unemployed.

HIV prevalence was found to be small among unemployed individuals as compare to the employed ones, whilst risk was higher among unemployed than employed individuals. Those who have ever been married had a higher prevalence of both HIV and risk (14.4% and 27.5%). More individuals reported that their health status is good and HIV prevalence among this group was found to be 12.8% compared to 24.5% for those who reported to being in a poor health condition. Health status is an important variable in studying sexually transmitted diseases (including HIV) since the immune system plays a huge role is the probability of infection.

Descriptive statistics for the behavioural variables are presented in Table A.2 in the Appendix. It was observed that more individuals have had an HIV test before and HIV prevalence in this group is 18.7% which is much higher than that for those who have never taken an HIV test (8%). Similarly those who have had an HIV test had a higher perceived risk. HIV prevalence was estimated to be higher among those who are sexually active (19.3%) and 3% among those who are not sexually active. This is an indication that there are other possible modes of HIV transmission rather than sexually transmitted HIV infection. However, perceived risk prevalence was higher among those who are not yet sexually active. This emphasizes the need of further statistical analysis to clearly understanding the distribution of the outcomes.

Individuals who used a condom at their sexual debut have a lower HIV prevalence than those who did not use a condom (14.8% vs 25%). HIV prevalence for those who were sexually active in the past 12 months and those who were not was estimated to be 19%. Those individuals who reported to have had STIs in the past months had a high HIV prevalence. Consequently this was the case for the perceived risk prevalence. Individuals who had multiple sexual partners had a higher HIV prevalence. It was observed that HIV prevalence is higher for those who have been in their current relationship for longer than a year. HIV prevalence for those individuals with whom the age difference of them and their current partners is 5 years or more was found to be much higher.

Individuals who have ever been pregnant had a higher prevalence of HIV, however it was observed that it is lower among those who were pregnant in the last 12 months. Similar interpretations for the perceived risk response are retained with some exceptions. It is important to note that the risk response was self-reported while HIV status was clinically tested.

The nature of the design of the study indicates that there might exist some clustering in the data, thus it is important to check this prior to statistical analysis. Preliminary analyses indicated a significant intra-cluster correlations for the two outcome variables and these are was estimated to be $\hat{\rho} = 0.149$ for HIV and $\hat{\rho} = 0.169$ for perceived risk, thus this has to be accounted in the analyses. Furthermore, it is

important to check whether the two responses are associated or not. This was done by using the likelihood ratio test (LRT) which indicated evidence against independence with a p-value (< 0.0001) less than 5% level of significance.

In addition, the correlation between the two responses was found to be significant where Spearman's correlation coefficient was found to be $\hat{\rho} = 0.157$ with p-value<.0001. Thus, this can be further confirmed by using much broader statistical methodologies mentioned in Chapter 2. The total proportion of missingness for the HIV response is 29% and 20% for the perceived risk response. Preliminary analysis indicated missing values depend on other observed variables, which gave evidence against missing completely at random (MCAR) in favor of missing completely at random (MAR). Multiple imputation will used as a supporting analysis to check how the results vary under the available cases analysis (without imputation) and multiply-imputed data analysis. This will be done for the univariate models with design weights only.

3.2 Statistical Results

3.2.1 Univariate Models

3.2.1.1 HIV model

Using the method for scaling design weights as was mentioned in Chapter 2, variables found to be the important risk factors are: Race, Age, Condom use at sexual debut, Age difference and the interaction between education and perceived risk. The estimated random intercept variance is $\hat{\sigma}^2 = 1.9663$ (SE = 0.6403). This random intercept was found to be significant with LRT = 40.83 and the mixture of chi-square p-value< 0.0001 which is highly significant at 5% level of significance. Results showed that the odds of being HIV infected for Africans is almost 10 times that of non-Africans (p-value = 0.0016).

Odds of being HIV positive for individuals in age group 15 to 19 years are at least 66% less than those who are in the 20 to 24 years age group while the odds for those who used a condom at their sexual debut are 43% less likely to be infected with HIV than those who did not use a condom. Individuals who reported having a sexual partner who is less that five years their senior are 57% less likely to be infected than those having partners who are at least 5 years their senior, and this was highly significant (p-value = 0.0011).

An interaction between education and perceived risk indicated that individuals with a lower education and perceive themselves as being at a lower risk are 15 times more likely to be infected with HIV, while those with a secondary education and a low

perceived risk are 13 times more likely to be infected with HIV. The resulted log odds estimates, standard errors, odds ratios and the p-values are presented in Table 3.2.

Table 3.2: *Results from the HIV model with design weights*

Effect	Estimate	Standard Error	Odds Ratio	p-value
Intercept	-1.839	1.027	0.16	0.074
Race (ref: non-African)				
African	2.265	0.712	9.64	0.002
Age (ref: 20 to 24)				
15 to 19	-1.086	0.298	0.34	0.001
Highest educational qualification (ref: Tertiary)				
No/primary	-0.178	0.876	0.84	0.839
Secondary	-1.479	0.763	0.23	0.054
Condom use at sexual debut (ref: No)				
Yes	-0.565	0.255	0.57	0.027
Perceived risk of HIV (ref: High)				
Low	-2.436	1.000	0.09	0.015
Age difference (ref: $\geq$ 5 years)				
$<$ 5 years	-0.851	0.259	0.43	0.001
Education and Perceived risk				
No/primary with a low perceived HIV risk	2.769	1.210	15.948	0.023
Secondary education with a low perceived HIV risk	2.595	1.051	13.395	0.014

An interaction between education and perceived risk is depicted in Figure 3.3. It is observed that those with lower qualification and also having low perceived risk are in fact more likely to be infected with HIV than those who perceived themselves as being at higher risk. Similarly those with secondary school qualification with a low perceived risk are significantly more likely to be infected with HIV. Among those with tertiary qualification, it was observed that those with a high perceived risk were 11 times more likely to be infected with HIV than those having a low perceived risk. Empirical Bayes estimates for this model are plotted in Figure B.1 in Appendix B. From this histogram it was observed that there seem not to be an indication of outliers.

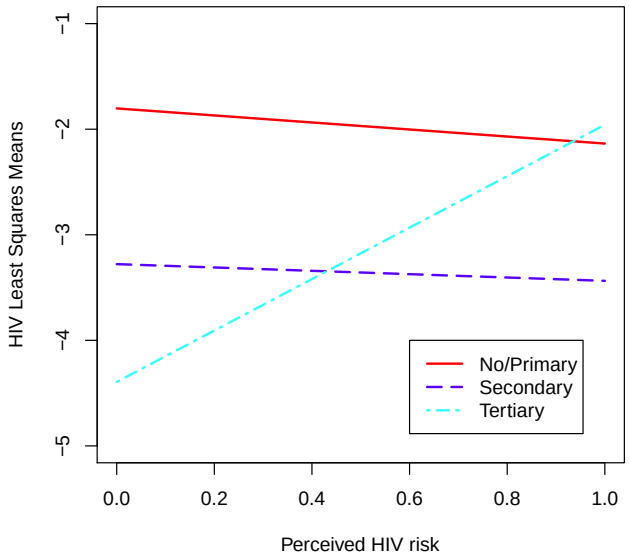

Figure 3.3: *Interaction of education and perceived risk for the HIV model*

3.2.1.2 Perceived Risk model

In this section the focus is shifted to the second outcome of interest which is the perceived risk of HIV, however it should be noted that the two outcomes are different. In the previous section the response was HIV status which was clinically tested and now we focus on the perceived risk of HIV, that is individuals were asked to rate themselves as to whether they think they are at high or low risk of being infected which is self-reported.

The fitted model included the design weights and the final model that was obtained contained: Race, Health, Condom use at sexual debut, Condom use at last sexual encounter, Age at sexual debut and the interaction of age and duration as risk factors. The variance for the random intercept was estimated to be $\hat{\sigma}^2 = 4.2567$ with SE = 1.0125, which is significant at 5% level where the p-value is obtained from a mixture of χ^2-distribution with equal weights (LRT=107.94, p-value<0.0001). Results from this model are tabulated in Table 3.3, that is the log odds estimates, standard errors, odds ratios and the p-values. The odds of being at high perceived risk for Africans are 14 times more likely than those for non-Africans, and this is highly significant at 5% level of significance (p-value<.0001).

Those who reported being in a good health condition are at least 80% less likely to have a high perceived risk than those with poor health. The odds of being in a high perceived risk are 60% less likely for those who used a condom at their first sexual encounter than those who did not use a condom, similarly the odds for those who used a condom at their last sexual encounter are at least 60% less likely to have a

high perceived risk. The odds of being at high perceived risk are 14% less likely for a unit increase in age at sexual debut. The interaction of age and duration of a relationship indicated that, 15 to 19 years females who are currently in a relationship for less than a year are less likely to be at a high perceived risk than those in a long term relationship.

Table 3.3: *Results from the perceived risk model with design weights*

Effect	Estimate	Standard Error	Odds Ratio	p-value
Intercept	2.046	0.949	7.73	0.032
Race (ref: Non-African)				
African	2.654	0.543	14.21	<.001
Age (ref: 20 to 24)				
15 to 19	-0.675	0.309	0.51	0.029
Health (ref: Poor)				
Good	-1.693	0.459	0.18	<0.001
Condom use at last sexual encounter (ref: No)				
Yes	-0.889	0.257	0.41	0.001
Duration of current relationship (ref: ≥ 1 year)				
< 1 year	0.320	0.415	1.38	0.441
Condom use at last sexual encounter (ref: No)				
Yes	-1.003	0.289	0.37	0.001
Age at sexual debut	-0.149	0.042	0.86	<0.001
Age and Duration of a relationship				
15 to 19 year old in a short term relationship	-2.2499	0.7160	0.105	0.002

An interaction of age and duration of a relationship is shown in Figure 3.4. From this figure it was observed that there was a big difference between those in a longer duration and those in a shorter duration among 15 to 19 years. However, those in a long term duration were found to be significantly at high perceived risk than those in a short term relationship within this age group (15 to 19 years). A small difference was observed among those in age group 20 to 24, where those in a short term relationship were found to be at higher risk (perceived), but this was not statistically significant. Empirical Bayes estimates are plotted in Figure B.2 in Appendix B and from this histogram there is no evidence of outlying cases.

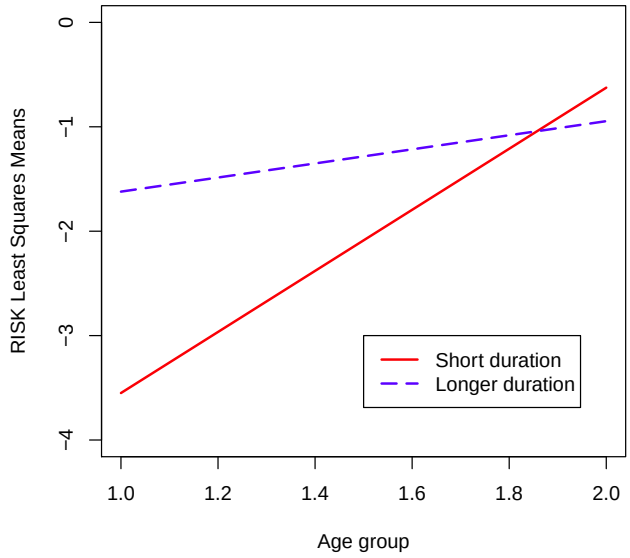

Figure 3.4: *Interaction of Age and Duration of a relationship for the perceived RISK model*

3.2.2 Other statistical models

3.2.2.1 Ignoring design weights

In this section models obtained in the previous sections are fitted but now without design weights. This is done to highlight the impact of ignoring the design when analysis data for a complicated design study. Results for HIV model are presented in Table 3.4 and from these results it was observed that variables that were significant when design weights when included in the analysis are no longer significant. Thus, this indicates that one will end up with different conclusions for analyses with/without design weights.

Table 3.4: *Results from the HIV model without weights*

Effect	Estimate	Standard Error	Odds ratio	p-value
Intercept	-1.832	0.736	0.16	0.013
Race (ref: Non-African)				
African	1.717	0.387	5.57	<.001
Age (ref: 20 to 24)				
15 to 19	-0.339	0.228	0.71	0.138
Education (ref: Tertiary)				
Primary	0.186	0.714	1.20	0.795
Secondary	-0.167	0.641	0.85	0.795
Condom use at sexual debut (ref: No)				
Yes	-0.568	0.202	0.57	0.005
Perceived Risk (ref: High)				
Low	-1.596	0.961	0.20	0.0977
Age difference (ref: ≥ 5 years)				
< 5 years	-0.734	0.201	0.48	0.0003
Education and Perceived risk				
No/primary with a low perceived HIV risk	1.705	1.078	5.50	0.115
Secondary education with a low perceived HIV risk	1.090	0.985	2.97	0.269

Results for the perceived risk response when ignoring design weights are presented in Table 3.5. In this analysis all variables are significant but they are different from the results obtained when accounting for design weights.

Table 3.5: *Results from the perceived risk model without weights*

Effect	Estimate	Standard Error	Odds Ratio	p-value
Intercept	0.763	0.643	2.15	0.2360
Race (ref: Non-African)				
African	1.820	0.285	6.17	<.0001
Age (ref: 20 to 24)				
15 to 19	-0.399	0.225	0.67	0.0763
Health (ref: Poor)				
Good	-0.849	0.322	0.43	0.0088
Condom use at sexual debut (ref: No)				
Yes	-0.402	0.181	0.67	0.0267
Duration of current relationship (ref: ≥ 1year)				
< 1 year	0.391	0.276	1.48	0.1566
Condom use at last sexual encounter (ref: No)				
Yes	-0.609	0.201	0.54	0.0026
Age at sexual debut	-0.093	0.0301	0.91	0.0021
Age and Duration of a relationship				
15 to 19 year old in a < 1 year relationship	-1.451	0.489	0.23	0.0032

3.2.2.2 Cluster size used as weights

In this section, the cluster size was used as weights. This method weights the results but it is important to note that this is different from the design weights analysis. It was observed that for the HIV model the interaction term is no longer significant.

Table 3.6: *Results from the HIV model with cluster size as weights*

Effect	Estimate	Standard Error	Odds Ratio	p-value
Intercept	-5.488	0.740	0.004	<.0001
Race (ref: non-African)				
African	1.837	0.439	6.28	<.0001
Age (ref: 20 to 24)				
15 to 19	-0.999	0.193	0.37	<.0001
Education (ref: Tertiary)				
Primary	0.277	0.539	1.32	0.6084
Secondary	0.084	0.494	1.09	0.8657
Condom use at sexual debut (ref: No)				
Yes	-0.139	0.161	0.87	0.3865
Perceived Risk (ref: High)				
Low	-0.791	0.669	0.45	0.2376
Age difference (ref: ≥ 5 years)				
< 5 years	-0.669	0.160	0.51	<.0001
Education and Perceived risk				
No/primary with a low perceived HIV risk	1.372	0.757	3.94	0.0706
Secondary education with a low perceived HIV risk	0.697	0.713	2.01	0.3287

Similarly, for the perceived risk model some variables are no longer significant, that is, condom use at last sexual encounter (p-value=0.2411). These results are presented in Table 3.7.

Table 3.7: *Results from the perceived risk model with cluster size as weights*

Effect	Estimate	Standard Error	Odds ratio	p-value
Intercept	1.172	0.752	3.23	0.1196
Race (Ref: Non-African)				
African	2.451	0.450	11.60	<.0001
Age (20 to 24)				
15 to 19	-0.927	0.198	0.40	<.0001
Health (ref: Poor)				
Good	-2.375	0.303	0.09	<.0001
Condom use at sexual debut (ref: No)				
Yes	-0.172	0.147	0.84	0.2411
Duration of current relationship (ref: ≥ 1year)				
< 1 year	1.084	0.229	2.96	<.0001
Condom use at last sexual encounter (ref: No)				
Yes	-1.238	0.159	0.29	<.0001
Age at sexual debut	-0.147	0.029	0.86	<.0001
Age and Duration of a relationship				
15 to 19 year old in a < 1 year relationship	-2.845	0.414	0.06	<.0001

3.2.3 Bivariate model

HIV and the perceived risk outcomes are jointly fitted where both outcomes follow a binary distribution where the probability of being HIV positive as well as probability of being at higher perceived risk are modelled. The adaptive gaussian quadrature is used as was done for the univariate models. In this section, the main interest is to find the covariance structure which best describes the association between the perceived risk and the HIV status outcomes. This is done by using mean structures obtained from univariate models when accounting for design weights.

In addition, it is important to note that the two outcome variables have different mean structures. Furthermore, if no association is found between the two responses, that will not directly mean that there is no association within a community for each outcome but it will mean that the two outcomes are independent. Results from the analyses are presented in Table 3.8 for the random effects where random effects structures comparison are presented in Table C.1 and parameter estimates are presented in Table C.2 in Appendix C.

The significance of the random effects is done using a mixture of chi-square distribution with equal weights as shown in Table C.1 (Appendix C). Parameter estimates for the models still maintain the univariate interpretation as done before.

Table 3.8: *Random effects parameter estimates for the Bivariate models under different covariance structures*

Model	Covariance assumption	AIC	Estimate (SE)
1	Independence	1789.4	
2	Common RE	1664.1	$\hat{\sigma}^2 = 2.379$ (0.483)
3	Positive correlation	1661.8	$\hat{\sigma}^2 = 1.251$ (0.536)
			$\hat{\nu} = 1.673$ (0.448)
4	Uncorrelated Different REs	1644.7	$\hat{\sigma}^2_{11} = 1.966$ (0.640)
			$\hat{\sigma}^2_{22} = 4.249$ (1.008)
5	Correlated Different REs	1638.3	$\hat{\sigma}^2_{11} = 2.258$ (0.717)
			$\hat{\sigma}^2_{12} = 1.439$ (0.547)
			$\hat{\sigma}^2_{22} = 4.256$ (1.011)
			$\hat{\rho} = 0.464$ (0.145)

From the results it was observed that a model which assumes independence (Model 1) fits poorly compared to other models, this model has the highest AIC value. Both shared parameter models (Model 2 - 3) fit poorly when comparing them to other models with random effects. These models are followed by the model with different random effects that are uncorrelated (Model 4) with an AIC-value = 1644.7. This model assumes that the two outcomes are not associated, when this assumption is relaxed it was observed that a model which assumes correlation is better. Model

5 is the better model compared to all models in Table 3.8 according to the AIC values. Comparing model 4 and model 5 it was found that the model with different correlated random effects is better than the model with independent random effects (LRT=8.4, p-value = 0.0267) where the p-value was calculated using a mixture of χ^2 distribution with 2 and 3 degrees of freedom .

Checking whether the covariance is different from zero, the p-value gave evidence towards a non-zero covariance (p-value= 0.0088). An estimated correlation was estimated to be $\hat{\rho} = 0.4643$ with p-value = 0.0015 which implies that the two outcomes are significantly associated but with different variability. This positive correlation is also depicted on Figure 3.5, where it is seen that when the perceived risk increases, HIV infection may also increase however in different magnitudes. In addition, it can be observed that there are communities with high levels of both HIV and perceived risk and also communities with lower levels of both outcomes. It was observed that the variance for the random intercept for HIV is $\hat{\sigma}^2_{11} = 2.258$ and $\hat{\sigma}^2_{22} = 4.256$ for the perceived risk, thus, implying that there is more variability in a community for the perceived risk than there is for HIV status.

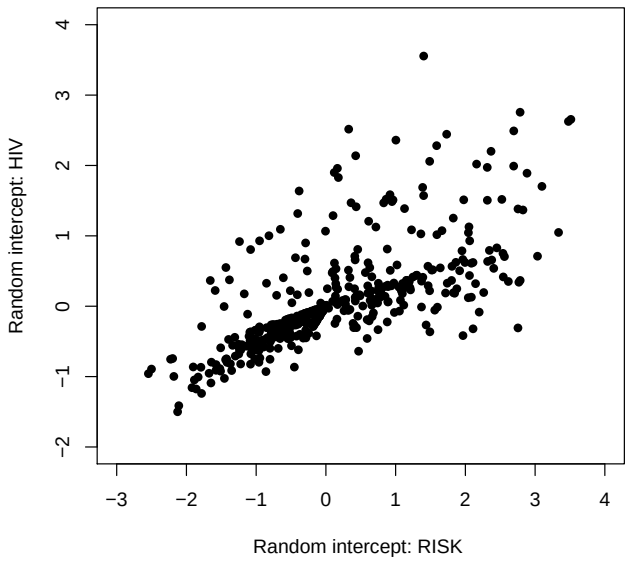

Figure 3.5: *Empirical Bayes estimates for the bivariate model with correlated random effects*

3.2.4 Complete data analysis

As was indicated that some of the observations for different variables are missing from the data. It is important to note that the GLMMs use all the available cases and thus missing information will not affect the estimates parameters. Since there was missingness in most of the covariates, MI was deemed necessary. The incom-

plete data was imputed 5 times (m = 5 imputation points) and only models where design weights are used will be fitted in this complete data. Results from these two analyses are presented in Table 3.9 for HIV and Table 3.10 for perceived risk model. Comparing these results with the available cases results, it was observed that there was not much differences between the analyses. This however gives more confidence in our results obtained from the available cases analysis.

Table 3.9: *Parameter estimates for the HIV model-Available cases and Multiple imputation*

	AVAILABLE CASES	MULTIPLE IMPUTATION
Parameter	Estimate (SE)	Estimate (SE)
Intercept	-1.839 (1.0267)	-1.969 (1.026)
Race (ref: non-African)		
African	2.265 (0.712)	2.286 (0.711)
Age (ref: 20 to 24)		
15 to 19	-1.086 (0.298)	-1.181 (0.297)
Education (ref: Tertiary)		
Primary	-0.178 (0.876)	-0.092 (0.875)
Secondary	-1.479 (0.763)	-1.339 (0.768)
Condom use at sexual debut (ref: No)		
Yes	-0.565 (0.255)	-0.480 (0.251)
Perceived Risk (ref: High)		
Low	-2.436 (1.000)	-2.415 (1.001)
Age difference (ref: ≥ 5 years)		
< 5 years	-0.851 (0.259)	-0.776 (0.255)
Education and Perceived risk		
No/primary with a low perceived HIV risk	2.769 (1.210)	2.766 (1.210)
Secondary education with a low perceived HIV risk	2.595 (1.051)	2.477 (1.062)
σ^2	1.966 (0.640)	1.963 (0.629)

Table 3.10: *Parameter estimates for the risk model-Available cases and Multiple imputation*

	AVAILABLE CASES	MULTIPLE IMPUTATION
Parameter	Estimate (SE)	Estimate (SE)
Intercept	2.046 (0.950)	1.490 (0.798)
Race (ref: Non-African)		
African	2.654 (0.543)	2.428 (0.441)
Age (ref: 20 to 24)		
15 to 19	-0.675 (0.309)	-0.567 (0.260)
Health (ref: Poor)		
Good	-1.693 (0.460)	-1.293 (0.347)
Condom use at sexual debut (ref: No)		
Yes	-0.890 (0.257)	-0.732 (0.205)
Duration of current relationship (ref: ≥ 1year)		
< 1 year	0.320 (0.415)	0.224 (0.465)
Condom use at last sexual encounter (ref: No)		
Yes	-1.003 (0.289)	-0.911 (0.225)
Age at sexual debut	-0.147 (0.042)	-0.135 (0.035)
Age and Duration of a relationship		
15 to 19 year old in a < 1 year relationship	-2.250 (0.716)	-1.756 (0.655)

Chapter 4

Discussion and Conclusion

4.1 Discussion

The aim for this study is to study factors that are associated with HIV among young females in South Africa. This was done using survey data from the third national survey which was conducted in 2008 by HSRC (Shisana et al., 2009). The analysis was based on females in age group 15 to 24 years old. Two outcome variables were studied, that is, HIV status and the perceived risk. For the second outcome, participants were asked to rate themselves as to whether they think they are at high or low risk of contacting HIV.

In the first part of the analysis the two outcomes were analysed univariately where the design of the study was taken into account by using design weights and communities (EAs) as clusters. These clusters accounts for the clustering that occurs at the community level. In addition, separate analyses where design weights were ignored and cluster size used as weights was done. From the two approaches it was observed that it is important to account for the design weights since if ignored there is a higher probability of obtaining biased results. Using cluster size in this current study as weights is not necessary since the outcome does not depend on the number of clusters. Thus, it is important that data is explored thoroughly before statistical analyses to make an informed decision as to which method to apply.

In the univariate model for HIV, risk factors that were found to be significantly associated with HIV are: race, age, condom use at sexual debut, age difference with the sexual partner, and the interaction between education and the perceived risk. Africans, 20 to 24 year old, those who did not use a condom at their sexual debut, and those having a relationship with older partners are more likely to be HIV positive. Furthermore, those having no or primary school qualification as their highest qualification and see themselves as being less likely to be infected with HIV are in fact at higher risk than those who see themselves as being at high perceived risk. Similarly those with secondary school qualification with low perceived risk are at more likely to be infected with HIV.

Several studies have indicated that Africans are significantly more likely to be infected with HIV compared to other races (Zuma et al., 2010, 2005; Mzolo, 2009; Pettifor et al., 2005; Eaton et al., 2003). This was also observed in the previous HIV surveys that has been conducted in South Africa (Shisana et al., 2005; Shisana and Simbayi, 2002). Results showed that those in age group of 20 to 24 years are at higher risk of being infected with HIV. This may be an indication that these females tend to engage in risky sexual behaviours which put them at higher risk of HIV compared to those who are in age group 15 to 19 years. Young females (15 to 19) are just entering their sexually active stage thus they may tend to be more careful than those in the older age group.

Condom use is a preventive measure that is commonly used to avoid HIV transmission. This does not guard against HIV infection only but also unwanted pregnancies. Condom use at sexual debut is an indication that one is aware of the dangers of HIV and it is most likely that the person will continue using a condom, thus placing this person at a lower risk of HIV infection and this was observed in this study. It has been reported that condom use at sexual debut is increasing among females, however this is lower among Africans then those in other race groups (Magnani et al., 2005).

HIV prevalence among 15 to 24 year old females is exacerbated by intergenerational sexual encounters where these young females get involved with older males (Leclerc-Madlala, 2008; Dunkle et al., 2004). In most cases this is done for material gain, and as a result attention to safe sex will be diverted to better life gain (Dunkle et al., 2004). Similarly in this study it was observed that those females having a sexual relationship with a partner who is at least 5 years older than them are more likely to be infected with HIV and this is in agreement with the other studies (Leclerc-Madlala, 2008).

Inter-generational sexual relationships may result in the imbalance of power in a relationship, this puts a woman at a higher risk of contacting HIV since it is difficult for her to negotiate safe sex (Leclerc-Madlala, 2008; Eaton et al., 2003; Gregson et al., 2002). Most of the times these individuals are coerced into having unprotected sexual encounters (Pettifor et al., 2005). Information about HIV is mostly conveyed via print media or televised media and through school to mention a few. The way of understanding this message will differ for people with different levels of education.

People with no or lower education are more likely to be infected with HIV than those with higher educational qualifications (Kalichman et al., 2006). People with higher education tend to be more informed about HIV than those with lower. In this study an interaction of education and perceived risk was found to be significant. Those individuals with lower education and at a lower perceived risk are more likely

to be infected with HIV than those with a high perceived risk. This is evidence of being misinformed about HIV education. This group may think they are at lower risk of being infected with HIV when in fact they are not doing enough to prevent themselves from being infected with HIV. This is especially worse for those with no education or just primary education as their highest qualification. This shows that there is more work which needs to be done about this group of people in making sure that they understand the risks of HIV. This evidence is discouraging when looking at the investments that the government has made in making sure that the youth is fully aware of HIV (Shisana et al., 2009).

Risk factors that were found to be significantly associated with the perceived risk are: race, health, condom use at sexual debut, condom use at last sexual encounter, age at sexual debut, and the interaction between age and the duration of a relationship. This response is informative in terms of the knowledge of HIV as to how well aware people are. Results indicated that Africans are more likely to say they are a higher risk than other races. This can be subjected to the fact that HIV is high among Africans than other race groups.

Those who said they are in a good health condition are less likely to say they are at high risk. Health condition is an important indicator when one is studying sexual transmitted diseases since HIV is also one of the sexually transmitted diseases which in turn implied lack of condom use (Zuma et al., 2005). Consequently having poor health status may be a direct indication of poverty which is associated with increased HIV infection (Kalichman et al., 2006). The use of a condom at sexual debut is an indication that one is aware of the dangers of HIV.

These people may tend to use condoms more consistent which is necessary for HIV prevention as lack of it may result in one being infected. In this study it was also found that people who used a condom during their sexual debut are less likely to be at higher risk. In addition, individuals who stated that they used a condom at their last sexual encounter were less likely to say they are at higher risk than those who did not. However, it is important to note that using condom at last sexual act is not enough to prevent HIV transmission but consistency of it is the most important thing.

The average age at sexual debut was found to be 18 and this is the legal age at which one can start becoming sexually active. Teenagers who are 18 years old may not be well informed about HIV and they pose a high risk of contacting HIV. In this study it was found that if the age at sexual debut is delayed the chances of being at high risk decreases. One reason for this is that when someone delays being sexually active may benefit the person, since by the time she decides to engage in sexual activity she may be more informed about HIV and more likely to make a well-

informed decision. In most cases abstinence is advised as a better option to delay age at sexual debut (Bakilana, 2005) and many national programmes are promoting delayed age at sexual debut (Zuma et al., 2010), which in turn prevents teenage pregnancy. Countries like Zimbabwe and Uganda reported an increase in age at sexual debut which coincided with the decrease in HIV prevalence (Gregson et al., 2006; Bakilana, 2005). This emphasizes the importance of delaying age at sexual debut as it proved to be beneficial. Factors that have been identified as significant determinants of age at sexual debut include age, race, geographical location, and level of education (Zuma et al., 2011).

In the results it was found that 15 to 24 year old individuals in a short term relationship are less likely to be at higher perceived risk than those in a long term relationship. This may sound strange, but it is important to note that individuals who are still getting to know each other tend to be more careful and condom use is most likely a non-negotiable option. But as time passes by they begin knowing each other more and they will start trusting each other and this 'trust' does not guarantees faithfulness. To intervene such occurrences, it will be important to emphasize the importance of safe sexual encounters as well as faithfulness among these young females having a long term relationship since trust alone will not guarantee immunity from HIV.

Bivariate model indicated that HIV and the perceived risk are positively correlated where the correlation was estimated to be 0.46. The community-specific variabilities were found to be significantly different for the two outcomes and it was observed that there is more variability for the perceived risk as there is for HIV as was shown in Table 3.8. In addition, the significance of the community-specific random intercepts implied that levels of HIV and perceived risk vary among communities.

There are those communities that are highly affected and those that are less affected. It can be stated that when there is an increase in the perceived risk there will be an increase in HIV also. Thus, as a preventive intervention it is important that individuals are aware of the risks of HIV and this should be done by focusing at community levels and more importantly paying more attention to those communities that are highly affected.

Data used is from a survey study where the design is involved a multi-stage clustering. Accounting for such a design, weights were calculated and it is crucial that the analysis conducted account for all this information. In the analysis effects of ignoring these weights as was indicated by the analyses where design weights were not used. Cluster size was used as weights and this method is used when the cluster size is informative, that is, when the response among observations in a cluster is associated with the cluster size. However, this approach is not advisable in the

current study since HIV status is not associated with the number of individuals in a community. Therefore, before one conducts any analysis it is important that the data is explored fairly well to support the methodology.

The data was analysed using the random effects model where the results are interpreted conditional on the given community and these are used since our interest was also on clustering. There are other methods that could have been used which are appropriate for such designs, like the generalized estimating equations (GEEs) which are commonly known as marginal or population-averaged models (Liang and Zeger, 1986).

Both methods account for clustering but this is done differently which makes it impossible to directly compare the results from the random effects and population-averaged models (Molenberghs and Verbeke, 2005). These models are complex but they use all the available data and are more suitable for explicative studies (Carriere and Bouyer, 2002). In addition, the random effects fall under direct likelihood methods which can be used when there exists missingness in the data (Molenberghs and Verbeke, 2005). In this study multiple imputation was used to account for missingness, however it was found that the results are not different from each other.

4.2 Conclusion

Results showed that Africans, 20 to 24 year olds, and lack of condom use at sexual debut are all associated with being likely to be HIV positive and also more likely to have a high perceived risk. Having older sexual partner and those individuals with lower educational qualification who think they are at lesser risk of HIV are in fact more likely to be infected with HIV.

Condom use at last sexual encounter as well as an increase in age at which experiences first sexual encounter are associated with being less likely to have a high perceived risk. In addition, 15 to 19 year old females who are in a short term during are less likely to be at a higher perceived risk than those in a long term relationship.

HIV infection and the perceived risk are highly associated, thus in order to be able to minimize the incidence of HIV it is important that these people who claim to be at a higher risk of HIV are educated well about HIV. The more they know about HIV and aware of its preventive measures will have an impact on the increasing HIV prevalence among 15 to 24 year old females.

Educational campaigns are always conducted nationally and it is more likely that there are communities that are missed. It will be necessary that communities that are highly affected with HIV are paid more attention in making sure that the fight

against HIV succeeds. In addition, the formula that is used to pass the information, that is, media should be revisited since it is possible that there are people in some communities that are unable to access these facilities.

4.3 Future Research work

In general joint models contain too many parameters which may result in poor estimation of the main parameters. Bayesian methods offer the advantage of borrowing information from similar studies or from experts which are then incorporated in the current analysis in the forms of prior distributions for the parameters (Wu, 2010). Such prior information helps estimating parameters that may be poorly identified by the current observed data alone, thus it would be interesting to conduct a Bayesian analysis for the bivariate model. Furthermore, an analysis which relaxes the normality assumption of the random effects is necessary since results that are obtained rely more on the assumed distribution of the random effects and if this is misspecified, results may be biased (Litiere et al., 2008).

Bibliography

Asparouhov, T. (2006). General Multi-Level Modeling with Sampling Weights. Communications in statistics. *Theory and methods*, **35(3)**:439–460.

Bakilana, A. (2005). Age at sexual debut in South Africa. *African Journal of AIDS Research*, **4(1)**:15.

Barros, A. J. D. and Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*, **3**:21.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**:9–25.

Carle, A. C. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology*, **9**:49 doi:10.1186/1471–2288–9–49.

Carriere, I. and Bouyer, J. (2002). Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons. *BMC Medical Research Methodology*, **2**:15.

Del Fava, E., Kasim, A., Usman, M., Shkedy, Z., Hens, N., Aerts, M., Bollaerts, K., Scalia Tomba, G., Vickerman, P., Sutton, A. J., Weissing, L., and Kretzschmar, M. (2011). Joint Modelling of HCV and HIV Infections among Injecting Drug Users in Italy Using Repeated Cross-Sectional Prevalence Data. *Statistical Communications in Infectious Diseases*, **3**: Issue. 1, Article 1:doi: 10.2202/1948–4690.1009.

Dunkle, K. L., Jewkes, R. K., Brown, H. C., Gray, G. E., McIntyre, J. A., and Harlow, S. D. (2004). Transactional sex among women in Soweto, South Africa: prevalence, risk factors and association with HIV infection. *Social Science and Medicine*, **59(8)**:1581 1592.

Eaton, L., Flisher, A. J., and Aaron, L. E. (2003). Unsafe sexual behaviour in South African youth. *Social Science and Medicine*, **56**:149–165.

Feuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: Pitfalls of the random-effects approach. *Statistic in Medicine*, **23**:3093–3104.

Goldstein, H. (2003). *Multi-level Statistical Models, 3rd Edition*. Arnold, London.

Gregson, S., Garnett, G. P., Nyamukapa, A. C., Hallett, T. B., Lewis, J. C., Mason, P. R., Chandiwana, S. K., and Anderson, A. M. (2006). Hiv decline associated with behavior change in eastern Zimbabwe. *Science*, **311(5761)**:664–666.

Gregson, S., Nyamukapa, A. C., Garnett, G. P., Mason, P. R., Zhuwau, T., Caral, M., Chandiwana, S. K., and Anderson, A. M. (2002). Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe. *The Lancet*, **359(9321)**:1896–1093.

Heeringa, S. G., West, B. T., and Berglund, P. A. (2010). *Applied Survey Data Analysis*. Chapman and Hall, New York.

Hodge, D. R. and Roby, J. L. (2010). Sub-Saharan African Women Living with HIV/AIDS: An Exploration of General and Spiritual Coping Strategies. *Social Work: ProQuest Social Science Journals*, **55(1)**:27–37.

Jahn, A., Floyd, S., Crampin, A., Mwaungulu, F., Mvula, H., Munthali, F., McGrath, N., Mwafilaso, J., Mwinuka, V., Mangongo, B., Fine, P., Zaba, B., and Glynn, J. (2008). Population level-impact of HIV on adult mortality and early evidence of reversal following roll-out of antiretroviral therapy in Malawi. *The Lancet*, **371(9624)**:1603–1611.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, **52**:5066–5074.

Kalichman, S. C., Simbayi, L., Kagee, A., Toefy, Y., Jooste, S., Cain, D., and Cherry, C. (2006). Association of poverty, substance use, and HIV transmission risk behaviors in three South African communities. *Social Science & Medicine,*, **62**:1641–1649.

Kreft, I. and de Leew, J. (1998). *Introducing Multi-level Modeling*. Sage, London.

Leclerc-Madlala, S. (2008). Age disparate and intergenerational sex in Southern Africa: The dynamics of hypervulnerability. *AIDS*, **22**:S17–S25.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects*. Chapman and Hall, New York.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**:13–22.

Litiere, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, **27(16)**:3125–3144.

Longford, N. T. (2005). *Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician*. Springer, New York.

Magnani, R., MacIntyre, K., Karim, A. M., Brown, L., and Hutchinson, P. (2005). The impact of life skills education on adolescent sexual risk behaviors in kwazulu-natal, south africa. *Journal of Adolescent Health*, **36**:289–304.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London.

McCullogh, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley and Sons, New York.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.

Mzolo, T. (2009). *Estimating Risk Determinants of HIV and TB in South Africa - Master thesis*. University of Kwa-Zulu Natal, Pietermaritzburg.

Peltzer, K., Matseke, G., Mzolo, T., and Majaja, M. (2009). Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. *BMC Public Health*, **9**:174 doi:10.1186/1471–2458–9–174.

Peltzer, K., Phaswana-Mafuya, N., Mzolo, T., Tabane, C., and Zuma, K. (2010). Determinants of knowledge of HIV status in South Africa: results from a population-based HIV survey. *Ethno Med*, **4(3)**:163–172.

Pettifor, A. E., Kleinschmidt, I., Levin, J., Rees, H. V., MacPhail, C., Hlongwa-Madikizela, L., Vermaak, K., Napier, G., Stevens, W., and Padian, N. S. (2005). A community-based study to examine the effect of a youth HIV prevention intervention on young people aged 15-24 in South Africa; results of the baseline survey. *Tropical Medicine and International Health*, **10**:971–980.

Pfeffermann, D., Skinner, C., Holmes, D., Goldstein, H., and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, **60**:23–40.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, New York.

Shisana, O., Rehle, T., Simbayi, L. C., Parker, W., Zuma, K., Bhana, A., Connoly, C., Jooste, S., and Pillay, V. (2005). *South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey*. HSRC Press, Cape Town.

Shisana, O., Rehle, T., Simbayi, L. C., Zuma, K., Jooste, S., van Wyk, V. P., Mbelle, N., van Zyl, J., Parker, W., Zungu, N. P., Pezi, S., and the SABSSM III Implementation Team (2009). *South African national HIV prevalence, incidence, behaviour and communication survey 2008: A turning tide among teenagers?* HSRC Press, Cape Town.

Shisana, O. and Simbayi, L. C. (2002). *Nelson Mandela/HSRC study of HIV/AIDS: South African National HIV Prevalence, Behavioural Risks and Mass Media Household Survey*. HSRC Press, Cape Town.

Snijders, T. A. B. and Bosker, R. J. (1999). *Multi-level Analysis: An Introduction to Basic and Advanced Multi-level Modelling*. Sage, London.

UNAIDS (2010). *Report on the Global AIDS Epidemic*. UNAIDS, Geneva.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

WHO (2009). *Integrating gender into HIV/AIDS programmes in the health sector : Tool to improve responsiveness to women's needs*. WHO, Geneva.

Williamson, J. M., Datta, S., and Satten, G. A. (2003). Marginal Analyses of Clustered Data When Cluster Size is Informative. *Biometrics*, **59**:36–42.

Worth, D. (1990). Sexual decision making and AIDS: why condom promotion among vulnerable women is likely to fail. *Stud Family Plann.*, **20**:297–307.

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Chapman and Hall, New York.

Zuma, K., Lurie, M. N., Williams, B. G., Mkaya-Mwamburi, D., Garnett, G. P., and Sturm, A. W. (2005). Risk factors of sexually transmitted infections among migrant and non-migrant sexual partnerships from rural South Africa. *Epidemiol. Infect*, **133**:421–428.

Zuma, K., Mzolo, T., and Makonko, E. (2011). Determinants of age at sexual debut and associated risks among South African youths. *African Journal of AIDS Research, Submitted*.

Zuma, K., Setswe, G., Ketye, T., Mzolo, T., Rehle, T., and Mbelle, N. (2010). Age at sexual debut: A determinant of multiple partnership among South African youth. *Afr. J. Reprod. Health*, **14[2]**:47–54.

Appendix A

Descriptive results

Table A.1: *Socio-Demographic factors by HIV and perceived risk*

Variable	Total	HIV+ (%)	Perceived Risk+ (%)
Overall			
Positive	234	13.88	40.26
Negative	1752	86.12	23.14
Race group			
African	1796	16.14	27.21
Non-African	1019	1.30	9.92
Age group			
15 to 19	1416	6.67	17.55
20 to 24	1399	21.14	31.30
EA geotype			
Urban formal	1628	11.34	18.30
Urban informal	384	20.99	33.14
Tribal area	650	13.98	29.64
Rural formal	153	19.22	26.03
Education			
No/primary	237	31.72	31.82
High school	2140	12.07	24.48
Tertiary	180	9.96	17.54
Employment status			
Unemployed	1895	13.45	24.84
Employed	336	17.15	22.58
Other	34	12.84	24.13
Marital status			
Single	1999	13.62	24.18
Ever Married	260	14.38	27.47
Health status			
Good	2106	12.82	23.10
Poor	151	24.46	45.46

Table A.2: *Behavioural factors by HIV and perceived risk*

Variable	Total	HIV+ (%)	Perceived Risk+ (%)
Ever had an HIV test			
Yes	1113	18.74	32.10
No	1132	7.99	11.05
Sexually active			
Yes	1368	19.33	24.94
No	845	3.06	42.07
Condom use at sexual debut			
Yes	801	14.84	32.00
No	570	25.33	32.00
Sexual activity in the past 12 months			
Yes	1104	19.10	30.83
No	269	19.44	17.79
Sexually Transmitted Diseases in the past 12 months			
Yes	94	31.19	41.12
No	1008	17.63	31.02
Multiple partnership in the past 12 months			
One partner	1029	18.96	32.55
More than 1 partners	71	21.48	21.56
Multiple partnership in the past month			
One partner	955	19.02	33.33
More than 1 partners	14	7.69	10.31
Duration of the current relationship			
Less than a year	218	10.56	23.58
Longer than a year	870	20.55	34.20
Age difference of partner (Recent partner)			
Less than 5 years	642	14.98	29.73
More than 5 years	364	26.25	37.68
Condom use at last sex			
Yes	799	18.86	27.82
No	325	17.87	40.95
Currently pregnant			
Yes	77	13.05	38.70
No	757	24.36	36.94
Pregnant in the last 12 months			
Yes	420	17.98	39.21
No	343	32.17	34.27
Ever been pregnant			
Yes	837	23.27	37.32
No	545	11.57	23.43
How often do you have sex after taking alcohol			
Always	6	0.00	80.05
Sometimes	94	11.23	20.47
Never	462	8.06	15.89
How often do you use a condom after taking alcohol			
Always	40	10.64	11.53
Sometimes	27	20.73	27.52
Never	32	2.86	32.75

Appendix B

Univariate results

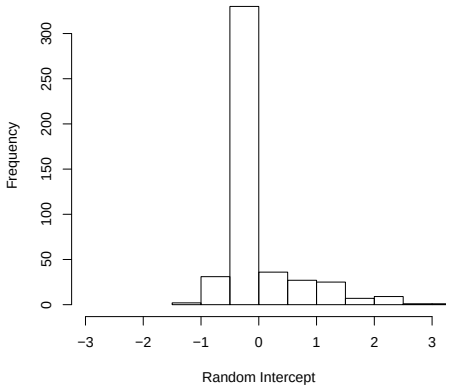

Figure B.1: *Empirical Bayes estimates for HIV model with design weights*

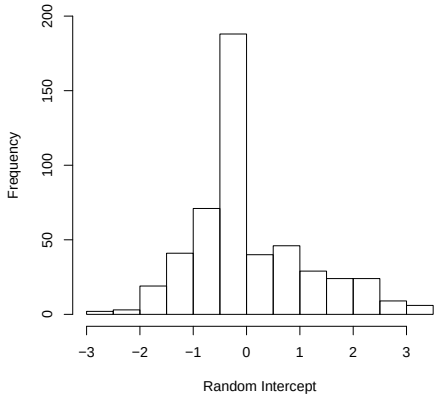

Figure B.2: *Empirical Bayes estimates for the perceived risk model with design weights*

Appendix C

Bivariate results

Table C.1: *Comparison of the random effects structures for the Bivariate models*

Model	Assumption	-2log	DF	Reference model	λ_n	Asymptotic Null Distribution	p-value
1	Independence	1751.4	19				
2	Shared	1624.1	20	1	127.3	$\chi^2_{0:1}$	< 0.0001
3	Shared + scale	1619.8	21	1	131.6	$\chi^2_{0:1}$	< 0.0001
4	Uncorrelated	1602.7	21	2	21.4	$\chi^2_{1:2}$	< 0.0001
5	Correlated	1594.3	22	4	8.4	$\chi^2_{2:3}$	0.0267

Table C.2: *Parameter estimates for the Bivariate models*

	SHARED	SHARED-SCALE	INDEPENDENT	CORRELATED
Parameter	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
HIV response				
beta10	-3.310 (0.997)	-2.827 (0.942)	-1.839 (1.027)	-2.385 (1.059)
beta11	2.822 (0.723)	2.565 (0.686)	2.265 (0.712)	2.484 (0.730)
beta12	-1.247 (0.278)	-1.106 (0.260)	-1.086 (0.298)	-1.169 (0.302)
beta13	0.180 (0.803)	0.088 (0.719)	-0.178 (0.876)	-0.100 (0.881)
beta14	-1.294 (0.690)	-1.174 (0.612)	-1.479 (0.763)	-1.518 (0.770)
beta15	-0.616 (0.246)	-0.563 (0.224)	-0.565 (0.255)	-0.585 (0.258)
beta16	-1.759 (0.968)	-1.789 (0.911)	-2.436 (1.000)	-2.162 (1.007)
beta17	-0.774 (0.245)	-0.717 (0.223)	-0.851 (0.259)	-0.842 (0.261)
beta18	3.131 (1.170)	3.023 (1.091)	2.769 (1.210)	2.872 (1.217)
beta19	3.290 (1.018)	2.971 (0.959)	2.595 (1.051)	2.914 (1.065)
perceived Risk response				
beta20	1.898 (0.817)	1.990 (0.892)	2.045 (0.949)	1.970 (0.945)
beta21	2.271 (0.456)	2.492 (0.513)	2.653 (0.543)	2.632 (0.540)
beta22	-0.521 (0.262)	-0.583 (0.290)	-0.675 (0.309)	-0.659 (0.308)
beta23	-1.342 (0.387)	-1.539 (0.436)	-1.692 (0.460)	-1.661 (0.459)
beta24	-0.841 (0.218)	-0.876 (0.241)	-0.890 (0.257)	-0.875 (0.256)
beta25	0.527 (0.349)	0.538 (0.385)	0.320 (0.415)	0.395 (0.413)
beta26	-0.991 (0.239)	-1.088 (0.268)	-1.002 (0.288)	-1.040 (0.287)
beta27	-0.141 (0.036)	-0.145 (0.039)	-0.149 (0.042)	-0.145 (0.042)
beta28	-1.951 (0.576)	-2.130 (0.645)	-2.249 (0.716)	-2.250 (0.704)

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Investigating factors associated with HIV among 15 to 24 year old females

Richting: **Master of Statistics-Biostatistics**
Jaar: **2011**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Mzolo, Thembile

Datum: **12/09/2011**