Duration of Relationship and HIV Transmission

Mary Ajibola Familusi (mary@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Professor Jane Hutton
University of Warwick, UK
Co-supervised by: Dr. Wim Delva
Ghent University and SACEMA

22 May 2014

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

Abstract

HIV poses a major public health challenge all over the world and especially in South Africa. The UNAIDS estimate for 2012 shows that 240,000 people died of AIDS and 2,500,000 were made orphans due to AIDS in South Africa. Certain sexual partnering practices, such as concurrent or multiple partnerships, are recognized to increase the risk of transmitting HIV. However the effect of a long-term relationship with a partner as a means of reducing the epidemic rate of HIV has not been considered. Using data collected from three urban disadvantaged communities in the greater Cape Town area, this study examines the relationship duration of 878 persons in Cape Town within one year. Non-parametric survival analysis and lifetime regression models were used to determine whether age, age-difference, condom use, HIV status, and race were significantly associated with how long a person stays in a relationship. The result of the analysis indicated that being female, coloured, aged between 38 and 47 and having sexual intercourse (coital frequency) three times in a week are positively associated with long-term relationship.

Keywords: HIV, AIDS, UNAIDS, Survival Analysis, Relationship

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Mary Ajibola Familusi, 22 May 2014

Contents

1. Introduction

South Africa's HIV epidemic is defined by the Joint United Nations Programme on HIV/AIDS (UNAIDS) as being a "hyper-endemic epidemic", because more than 15% of the population that are within the age group 15-49 years are living with HIV (Shisana et al., 2009b). In one of the main findings of the South Africa National HIV Prevalence, an estimate of 6.4 million people were living with HIV/AIDS in 2012. In order to minimize the effect of HIV in South Africa, several studies has been done to examine factors that contribute to the spread of HIV. Some of these factors are: "poverty, inequality and social instability, high levels of sexually transmitted infections, the low status of women, sexual violence, high mobility" (AFSA). The use of ART (Antiretroviral Treatment) has contributed significantly to the ongoing reduction of HIV infections all over the world. A recent study in the rural part of South Africa established that the occurrence of HIV infection has dropped by 17% for every 10% increase in the number of people that are receiving ART (Hocking, 2013).

Research has shown high levels of understanding about how HIV can be transmitted and methods of preventing it. But, this does not result into HIV preventive practice. That is the reason why this project is set out to examine the duration of relationship and factors that may contribute to a long lasting relationship among partners. In fact, recent studies show that multiple and concurrent sex partners play an important role in promoting HIV transmission. Concurrent relationships have been defined by the working group on measuring concurrent sexual partnerships of the UNAIDS reference group on Estimates, Modelling, and Projections as "overlapping sexual partnerships in which sexual intercourse with one partner occur between two acts of intercourse with another partner" (UNAIDS). In the data collected from the survey that was conducted from June 2011 to February 2012, in three urban disadvantaged communities in the greater Cape Town area to study associations between HIV status, sexual connectedness and age-disparity, the result of the analysis showed that 46% and 41.4% of male and female participants respectively, had more than one partner. In the next section, we shall review a few yet important studies that have been done on HIV, and factors that contribute to it.

HIV and Gender

The spread of HIV in South Africa keeps on showing that the prevalence of HIV in females is more than that of males. The prevalence of HIV among males of age 15-19 year is under 1% and for females in the same age group, it is 5.6%. This indicates that females are contracting HIV by having a sexual relationship with men that are older than them. The females HIV prevalence at age 20-24 years is three times that of males of the same age range (PLHL).

At age 30-34 years, the prevalence of HIV among females rises to 36.0% while the prevalence among males in the age group of 35-39 years rises to 28.8%. The widespread of HIV in males is significantly lower at age 40-44 years than the female counterpart (PLHL).

HIV and Different Age Group

The prevalence of HIV among different age groups varied over the years. For the period 1996 to 2005, the Age-Specific HIV Prevalence data shows a rapid progression of infection across all age groups. The data reveals a surprising rate of increase from 1.5% in 1996 to 20.1% in 2005 amongst the 25-29 age groups. This age group has been known to always show the highest levels of infection over the previous five years. Also in the age group 30-34, the HIV prevalence shows an increase from 11.5% in 2003 to 18.7% in 2005, and for the year 2003 to 2005, there has been a slight reduction of 1.5% in the HIV prevalence of age group less than 20 years old (Shaikh and Smith, 2005).

HIV and Age Disparity

In Zimbabwe, young women with partners who are five or more years older, were more than seven times likely to be HIV-infected than women with same-age partners (Simon et al.). In South Africa, such age-discrepant partnering is an important factor contributing to an HIV prevalence of 24.5 percent among young adult women aged 20-24 (Audrey et al., 2005).

In South Africa, older partner exposed young women to HIV virus since the prevalence of HIV in men increases as their age increases most especially at their early thirties (Shisana et al., 2009b). However, the high rate of HIV being transmitted from older men to younger women was because older men are less likely to always use condoms unlike men that does not have age difference (Bankole et al., 2007; Luke et al., 2013).

HIV and Ethnic Group

A striking characteristic of the HIV spread in South Africa and elsewhere is differences in racial or ethnic groups. The 2005 Human Sciences Research Council representative survey of people 15 to 49 years old in South Africa found HIV prevalence among black, coloured and white populations to be 19.9%, 3.2% and 0.5%, respectively (Shisana et al., 2005). The term coloured refers to a racial category in South Africa, and consists of racially mixed descendants of Europeans, indigenous populations and slaves from South and East Asia (Delva et al., 2013). HIV prevalence among 15 to 24-year-olds in the same racial groups, were 12.3%, 1.7%, and 0.3%, respectively. With an HIV incidence rate of 4.5%, black African females aged 20-34 years had the highest incidence of HIV among the analysed population groups (HSRC).

HIV and Relationship

The effect of HIV infection on the dissolution of marriages and dis-stabilization of families cannot be overemphasized. Laura, in 2004 investigates "the influence of HIV status on the risk of separation or divorce and widowhood among women in Rakai, Uganda". Her results show that separation is high among women that are infected by HIV than those that are positive and for those that do not know their HIV status. The reason why separation is common among HIV positive people is as a result of the partners unfit to carry out the supposed roles in the sexual aspect of the relationship. It might also be as a result of stigmatization, sickness and death attached to HIV infection .

HIV and Multiple-Concurrent Sexual Partnerships

Concurrency is seen to be playing a major role in the transmission of HIV, because partner having HIV is more likely to have sexual intercourse with multiple partner during the acute phase of the infection. The acute phase of HIV infection can also be referred to as the primary phase where the amount of HIV in the blood increases rapidly within a very few days.

The rampaging of multiple and concurrent partnerships among young people is due to the lack of understanding about the risk associated with it and can also be as a consequence of low self-esteem. In 2010, Malawi Demographic and Health Survey (MDHS) reported that 0.7% of the women that are in the age group 15-24 years was reported of having more than two partners in about a year before their survey. And 6.5% of men in the same age category also have more than two partners.

1.1 Objective of the study

The purpose of this study is to explore the association between the duration of a relationship per episode in 12 months, using data collected from the three disadvantaged communities in Cape Town for the

purpose of investigating age-disparity, sexual network structure and HIV infection by considering factors, such as race, coital frequency, gender, HIV status, age difference between partners.

2. Data and Methodology

2.1 Source of Data

The data are from the South Africa Centre for Epidemiological Modelling and Analysis (SACEMA). SACEMA is a national research centre, established under the Centre of Excellence programme of the Department of Science and Technology and the National Research Foundation, South Africa, and is hosted by the University of Stellenbosch. The survey was conducted from June 2011 to February 2012 in three urban disadvantaged communities in the greater Cape Town area to study associations between HIV status, sexual connectedness and age-disparity (Delva et al., 2013). The survey took a sub-sample of participants from Zambia-South Africa TB and AIDS Reduction Study (ZAMSTAR), a community randomized trial that aimed to reduce the prevalence of TB and HIV using novel public health interventions (Beauclair et al., 2013). It was conducted using Audio Computer-Assisted Self-Interview (ASACI) technology. ASACI has a benefit of providing privacy to the participant and avoids the white-coat effect when answering questions about sensitive topics (Delva et al., 2013). Out of the target size of 1500 participants, 878 participants took the survey and reported on the total of 1128 relationships, consisting of 1567 episodes (Delva, 2012).

2.2 Limitation of the study

From the data used for this study, each group is not well represented. For example, females are more than males participant, also black are more than coloured participants. There is missing information for different variables that varies from different respondent. Moreover, the use of the R statistical software automatically excludes missing data during the analysis.

2.3 Method of Analysis: Survival Analysis

Survival analysis is the statistical method for analysing longitudinal data on the occurrence of events, such as death, injury, the onset of illness, recovery from illness or transition above or below the clinical threshold of a meaningful continuous variable; all these are referred to as the event of interest. Our event of interest in this project is discontinuation of relationship. The objective of survival analysis is to estimate time-to-event for a group of individuals, to compare time-to-event between two or more groups and to assess the relationship of co-variables to time-to-event. Survival analysis data, also referred to as lifetime data, can be analysed by survival function $S(t)$ which is defined as the probability that an individual survives for a time at least t. This function can be estimated by the empirical survivor function, given by

$$S(t) = \frac{\text{Number of individuals with survival time} \geq t}{\text{Number of individuals in the data set}}.$$

Equivalently,

$$S(t) = 1 - F(t) \tag{2.3.1}$$

where $F(t)$ is the empirical cumulative distribution function of a random variable $T > 0$ with the probability density function $f(t)$ for $t < 0$. If $f(t)$ is continuous, then

$$F(t) = \int_{-\infty}^{t} f(u)du = P(T \leq t) \tag{2.3.2}$$

this implies that the probability density function of the random variable T is the derivative of its cumulative distribution function

$$f(t) = \frac{dF(t)}{dt}. \tag{2.3.3}$$

From Equations 2.2.1 and 2.2.3, we get

$$f(t) = -\frac{dS(t)}{dt}. \tag{2.3.4}$$

Another probability function used in survival analysis is the hazard function which is also known as the instantaneous death rate or the force of mortality.

Hazard function:

$$h(t) = \lim_{\epsilon \to 0^+} Pr(T < t + \epsilon | T > t)/\epsilon$$
$$h(t, \theta) = f(t, \theta)/S(t, \theta)$$

where $Pr(T < t+\epsilon|T > t)$ is the conditional probability that an event will occur in the interval $(t, t+\epsilon)$ given that it has not occurred before and ϵ is the width of the interval.

Lifetime data may be incompletely determined for some subjects due to censoring. The following events are regarded as censoring: (a) a subject that does not experience the event of interest before the study ends, (b) a subject that is lost to follow-up during the study period, (c) a subject that withdraws from the study. All methods of analysing lifetime data must take censoring into account in order to avoid loss of information and having a biased result. In this project, we shall use parametric and non-parametric methods of survival analysis that will allow us to account for censored data and we shall discuss these methods in the next section.

2.4 Kaplan-Meier Estimate

The Kaplan-Meier estimate of the survival function is an empirical or non-parametric method of estimating survival time $S(t)$ from non-or right-censored data. It is extremely popular as it requires only very weak assumptions and yet utilises the information content of both fully-observed and right-censored data. Kaplan-Meier estimate can be calculated using a standard statistical package in R software and can also be calculated by hand using this formula.

$$S(t) = \prod_{t_i < t} \left(1 - \frac{d(t_i)}{n(t_i)}\right) \tag{2.4.1}$$

$$= \prod_{t_i < t} S(t_i), \qquad \text{for } \; i = 1, \ldots, n \tag{2.4.2}$$

where $S(0) = 1$ by definition, $d(t_i)$ is the number in the study who fail at t_i, $n(t_i)$ is the number in the study at t_i and at risk that is able to fail at t_i.

2.5 Log-Rank Test

The log-rank test is a statistical test of difference between survival times of the two groups. It tests the null hypothesis that the groups are from the same population, and it compares the observed number of events in each group with the corresponding expected numbers for each. For example, in this project we shall be testing the null hypothesis that the different HIV status group have the same proportion of continuing in a relationship, different race has the same proportion of continuing in a relationship and other independent variables that we have chosen.

The null hypothesis is $H_0 : S_1(t) = S_2(t)$, where $S_1, S_2(t)$ are the survival functions for the two groups.

The alternative hypothesis is $H_1 : S_1(t) \neq S_2(t)$ for all t.

The alternative hypothesis will accepted if the null hypothesis outcome is false; this can be determined if the p-value, that is the significant level of test, is less than the α value (which is the threshold value that we measure p-value against). In this project, we shall use 0.05 as our threshold value which means we are 95% confident of our decision.

Suppose that two different groups are being compared. The test statistics is

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2},$$

which is referred to as the χ_1^2 distribution.

Note: O_1 and O_2 are the total numbers of observed events in groups 1 and 2 respectively, and E_1 and E_2 the total numbers of expected events.

Decision rule: Do not reject H_0 if p-value $> \alpha = 0.05$.

2.6 Lifetime Regression Model

The regression models help us to find a relationship between survival time and other covariates or factors. The two most widely used models for regression analysis of lifetime data involve the proportional hazards and accelerated life models.

2.7 Proportional Hazards Model

Proportional hazards regression model is also called Cox regression. It models the incidence or hazard rate, the number of new cases of disease per population at-risk per unit time. It is called the mortality rate if the outcome is death.

The proportional hazards regression model is given by

$$h(t|X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i)$$

where $h(t)$ is the baseline hazard function, X_i are the covariates and β_i are the coefficient of the covariates.

2.8 Accelerated Life Model

Accelerated life models are essentially standard regression models applied to the log of survival time. They relate the lifetime distribution of the explanatory variable such as gender, race and age.

The accelerated life regression model is given by

$$\log T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i + \epsilon$$

where T is the dependent variable, X_i are the covariates and β_i are the coefficient of the covariates.

It is only in the case of the Weibull models that these two kinds of modelling processes coincide (that is the Proportional Hazard Model and Accelerated Life Model). If it is not appropriate to fit a Weibull model to given data, then a choice needs to be made between using accelerated lifetimes or the proportional hazards model.

In order to fit either a proportional hazards model or an accelerated live model we need to identify a distribution followed by survival times in the data set and then estimate the parameters of this distribution using the maximum likelihood estimation method. These distributions include log-logistic, log-normal, exponential and Weibull distributions.

2.8.1 Log-normal distribution. The distribution of the random variable T is log-Normal if $Y = \log(T)$ has a Normal distribution, with $E(Y) = \mu$ and $Var(Y) = \sigma^2$. Hence

$$Y = \mu + \sigma Z, \quad \text{with} \quad Z \sim N(0,1),$$

with $\mu = -\log \lambda = \gamma' \underline{x}$ and $\sigma = \frac{1}{\alpha}$.

$$Y = logT = \gamma' \underline{x} + \sigma\epsilon, \qquad \epsilon \sim N(0,1).$$
$$logT = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_i x_i + \epsilon, \qquad \epsilon \sim N(0,1)$$

where γ' is $\gamma_0,\ \gamma_1,\ \ldots,\ \gamma_i$ and $\underline{x}$ is $x_1,\ x_2,\ \ldots,\ x_i$.

The probability density function of T is

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \frac{\alpha}{t} \exp\{\frac{-(\alpha \log(\lambda t))^2}{2}\}$$

and the survival function is $S_T(t) = 1 - \Phi(\alpha \log(\lambda t))$.

2.9 Software

Data were imported into R for analysis. R is a language and environment for statistical computing and graphics. Standard R commands (survfit, survdif and survreg) were used for the Kaplan-Meier estimate, log-rank test and regression model analysis, respectively.

3. Results

In this chapter we shall present the descriptive and the advanced statistical results using the method of survival analysis.

3.1 Descriptive Statistics

Table 3.1: Summary of the variables used in the analysis from 878 participant.

Variables	Participant				Relationship			
	Male		Female		End		Ongoing	
	N	%	N	%	N	%	N	%
Age								
18-27 years	69	26.1	154	28.2	44	36.1	138	25.8
28-37 years	60	22.7	152	27.8	34	27.9	152	28.4
38-47 years	53	20.1	143	26.2	18	14.7	150	28.1
48-57 years	53	20.1	55	10.1	16	13.1	57	10.7
58-67 years	27	10.2	34	6.2	9	7.4	34	6.4
68-70 years	2	0.8	8	1.5	1	0.8	3	0.6
AGE DIFFERENCE								
Female older than male	32	16.3	19	19.0	53	43.8	175	33.2
Equal age	14	7.1	26	6.0	7	5.8	35	6.6
Male older than female	150	76.6	323	75.0	61	50.4	317	60.2
Race								
Coloured	59	22.3	141	25.0	22	16.9	116	21.8
Black	206	77.7	423	75.0	108	83.1	417	78.2
HIV status								
Positive	35	12.9	147	25.6	38	29.0	117	21.4
Negative	185	68.0	349	60.7	72	55.0	342	62.6
Not done	52	19.1	79	13.7	21	16.0	87	16.0
Coital frequency per week								
0	7	3.9	12	3.0	6	5.8	14	2.8
1	37	20.8	120	29.7	19	18.4	148	29.7
2	57	32.0	146	36.1	36	35.0	173	34.7
3	46	25.8	77	19.1	15	14.6	109	22.0
> 3	31	17.4	49	12.1	27	26.2	54	10.8
Partner number within one year								
1	115	54.0	260	58.6	11	8.4	406	74.4
2	30	14.1	92	20.7	49	37.4	74	13.6
3	35	16.4	41	9.2	38	29.0	40	7.3
> 3	33	15.5	51	11.5	33	25.2	26	4.7
Condom frequency								
Always	47	24.9	129	30.9	32	29.1	150	29.0
Irregular	57	30.2	136	32.5	45	40.9	155	30.0
Never	85	44.9	153	36.6	33	30.0	212	41.0

Table 3.1 summarises the independent variables used in the analysis. Of the full sample of women, 28.2% are within the age group 18 to 27 while the highest percentage of male participants (26.1%) fall between the age group 18 to 27. The majority of the participants are between ages 18 and 70, so the age group that is younger than 18 years and older than 70 years was left-out in this project.

27.9% of participants between age 28 and 37 was reported to have ended their relationship. Fewer participants were reported to be HIV positive and 29% of those that have ended their relationship are HIV positive while 21.4% of the ongoing relationships are HIV positive as at the time of the survey. 17.4% of men and 12.1% of women do have sexual intercourse more than three times a week. The majority of the participants are black (76.1%) whilst coloured are just 23.9%.

Also, the percentage of participants that have never used condoms is higher in males (44.9%) than females (36.6%). 40.9% of those that use condoms irregularly were reported to have ended their relationship. 68.3% of participants had one episode with their partner, 20.2% had two episodes and 11.5% had three episodes.

3.2 Non-Parametric Method of Survival Analysis

In this section, we are looking at the estimates from Kaplan-Meier and log-rank test for gender, race, HIV status, language, age, condom use and age difference.

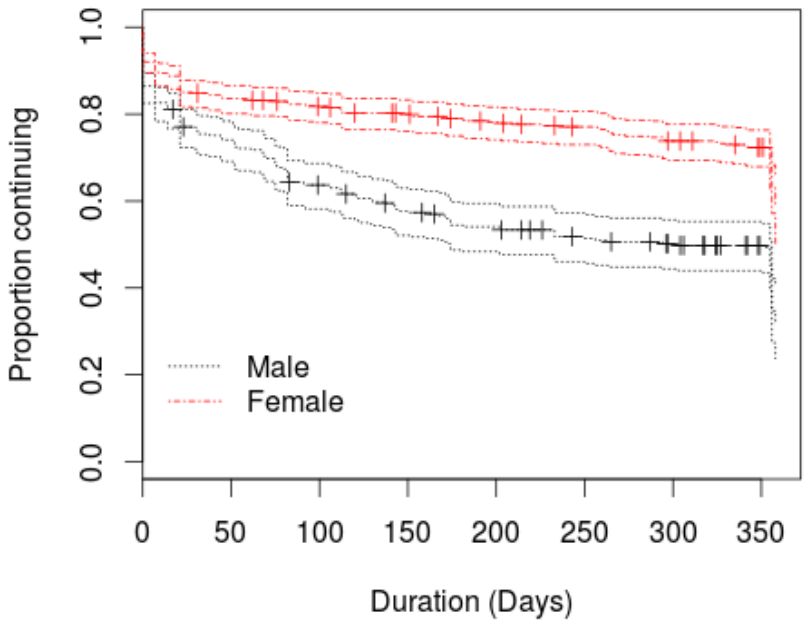

Figure 3.1: The Kaplan-Meier Curves of the first episode for Gender.

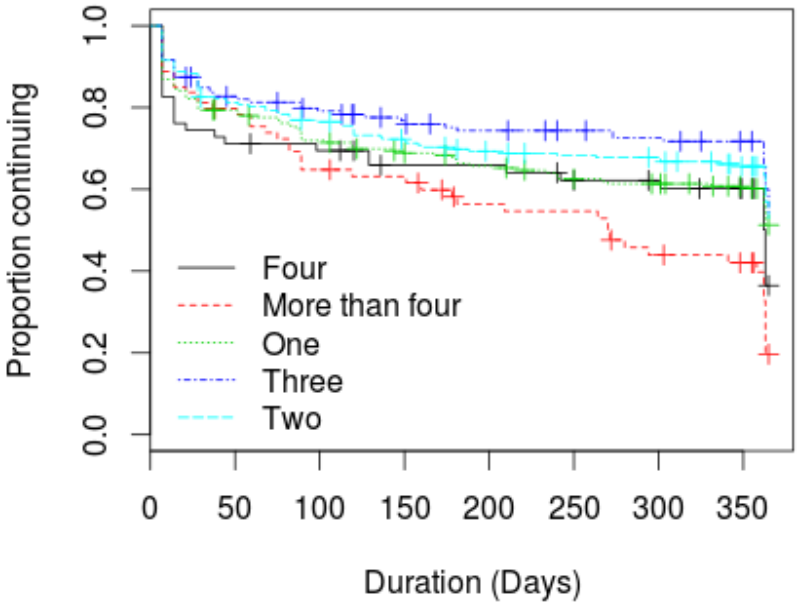

Figure 3.2: The Kaplan-Meier Curves of the first episode for Coital frequency.

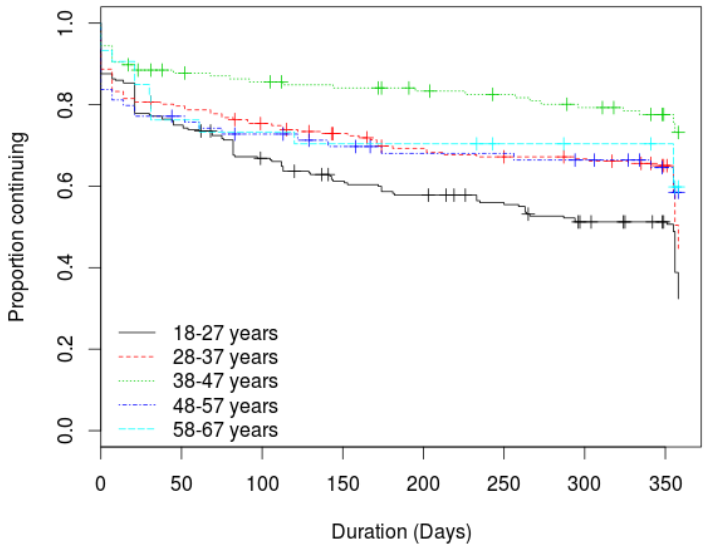

Figure 3.3: The Kaplan-Meier Curves of the first episode for Age categories.

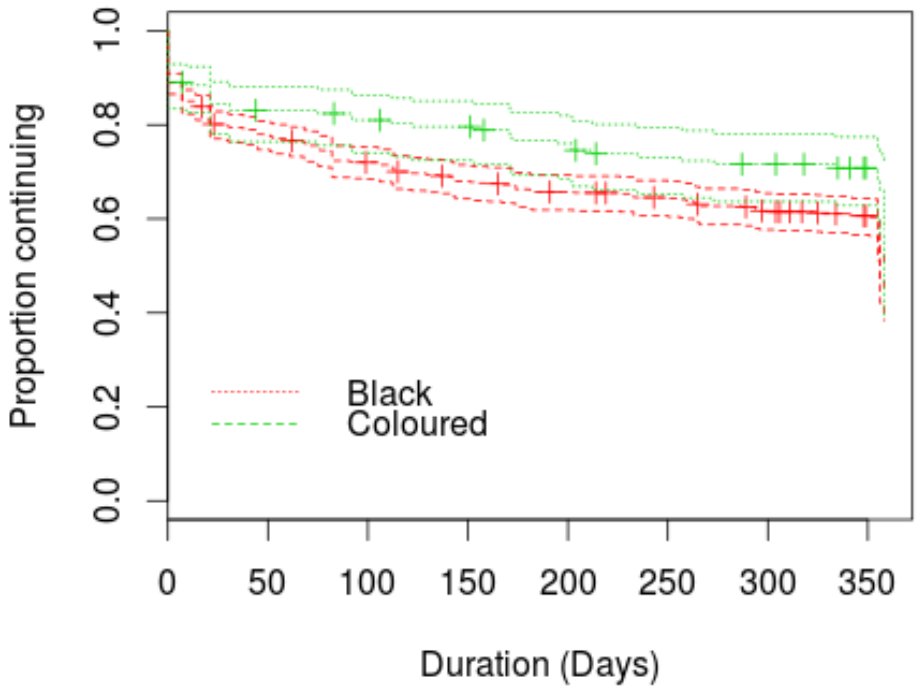

Figure 3.4: The Kaplan-Meier Curves of the first episode for race.

Kaplan-Meier Estimate

Results presented in Table 3.2 show that the participants that are coloured (67.8%), HIV negative (56.4%), never use a condom (64.2%), female (59.5%), Afrikaans (68.1%) and male older than female (53.9%) have a higher probability of continuing in a relationship up till 12 weeks. This is clearly represented in the Kaplan-Meier Curves for some of the variables above.

Moreover, looking at Table 3.1 we can observe that the percentage of coloured-race that have their relationships ongoing as at the time of the survey (21.8%) is higher than the percentage of those that have break-up relationship but for the black-race, the reverse is the case. The probability of continuing in a relationship for the participants that have never use a condom is higher than the rest of the group.

Log-Rank Test

Since $p > 0.05$ for HIV status, we cannot reject the null hypothesis of no-difference which means that HIV status cannot be used to measure if a participant will stay longer in a relationship or not. This is obvious according to the Kaplan-Meier table above since the difference between their probability of continuing in a relationship is not significantly different from each other.

The hypothesis of no-difference between the relationship duration of different races is rejected since the p-value is less than the threshold value. Also, for different age categories the log-rank test shows that there is a significant difference between the time that a participant in the age group 18-28 years, 29-39

Table 3.2: Probability first episode lasts at least 12 weeks and p-value for the log-rank test.

Variables	Estimate (%)	95% LCL	95% UCL	p-value
RACE				0.006
Coloured	67.8	60.3	74.2	
Black	52.4	48.9	55.8	
HIV STATUS				0.55
Negative	56.4	52.3	60.3	
Positive	50.9	44.0	57.4	
Not done	54.5	46.9	61.5	
CONDOM FREQUENCY				< 0.050
Always	44.7	39.2	49.9	
Irregular	55.8	50.1	61.2	
Never	64.2	58.7	69.1	
AGE(Years)				< 0.050
18-27	63.3	57.4	68.5	
28-37	73.4	67.3	78.6	
38-47	84.1	77.5	88.9	
48-57	71.3	60.2	79.8	
58-67	70.4	52.6	82.5	
AGE DIFFERENCE				0.481
Female older than male	49.3	43.1	55.1	
Equal age	45.8	25.6	64.0	
Male older than female	53.9	41.0	65.0	
GENDER				< 0.050
Male	50.3	45.1	55.2	
Female	59.5	55.4	63.5	
LANGUAGE				< 0.050
Xhosa	51.1	47.2	54.8	
English	58.2	49.8	65.7	
Afrikaans	68.1	60.3	74.7	

years, 40-50 years, 51-61 years and 62-72 years can stay in a relationship within a year, and this can be clearly seen from the Kaplan-Meier curve for age categories.

For gender, condom use and language, we shall reject the null hypothesis and conclude that there is a significant time difference between those participants that always using a condom, those that use them irregularly and those that never use them when it comes to being in a relationship.

Table 3.3: A summary table for the Kaplan-Meier Estimate of three different levels of episode in a relationship of the chosen independent variables for 3 weeks.

Variable	Episode	Group	Estimate (%)	95% LCL	95% UCL
RACE	First	Coloured	84.5	78.0	89.2
		Black	80.1	77.0	82.8
	Second	Coloured	80.7	37.7	95.4
		Black	65.1	53.3	74.6
	Third	Coloured			
		Black	63.9	44.9	77.8
CONDOM FREQUENCY	First	Always	72.7	67.2	77.4
		Irregular	83.2	78.3	87.1
		Never	86.8	82.4	90.1
	Second	Always	64.0	44.8	78.0
		Irregular	72.9	54.7	84.7
		Never	62.6	37.9	79.8
	Third	Always	70.3	43.1	86.3
		Irregular	62.0	30.0	82.7
		Never	77.6	39.0	93.4
HIV STATUS	First	Negative	82.0	78.4	85.0
		Not done	78.4	71.3	83.9
		Positive	80.0	73.6	85.0
	Second	Negative	77.9	62.1	87.7
		Not done	57.0	33.6	74.8
		Positive	60.7	39.4	76.5
	Third	Negative	77.6	50.4	91.0
		Not done	55.3	21.7	79.4
		Positive	64.3	32.4	84.1
GENDER	First	Male	76.8	72.0	80.9
		Female	85.1	81.7	87.9
	Second	Male	67.5	49.8	80.1
		Female	66.0	51.1	77.3
	Third	Male	73.5	38.4	90.5
		Female	66.3	46.2	80.3
LANGUAGE	First	Xhosa	79.5	76.1	82.6
		English	81.6	74.2	87.1
		Africans	85.6	78.9	90.3
	Second	Xhosa	63.9	51.1	74.2
		English	78.4	45.0	92.9
		Africans	74.3	24.5	93.9
	Third	Xhosa	64.4	44.1	79.0
		English			
		Africans			

Table 3.3 shows that the percentage of participants that stay with their partner in the first episode is higher than the percentage of those that stay in second and third episodes after 3 weeks.

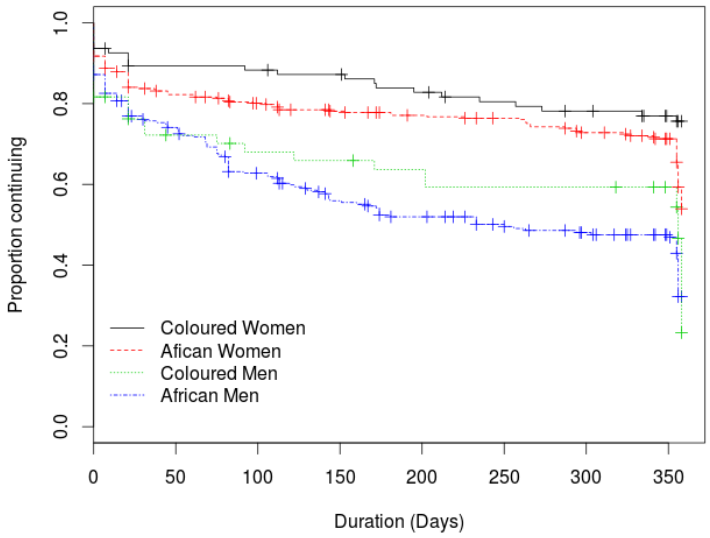

Figure 3.5: The Kaplan-Meier Curves for the Joint Effect of Race and Gender for the First Episode.

The estimate for gender and race in Table 3.2 shows that a female-coloured participant has a higher probability of continuing an episode than an African male. Also being female and coloured at the same time has a higher probability of continuing an episode as clearly shown in Figure 3.5.

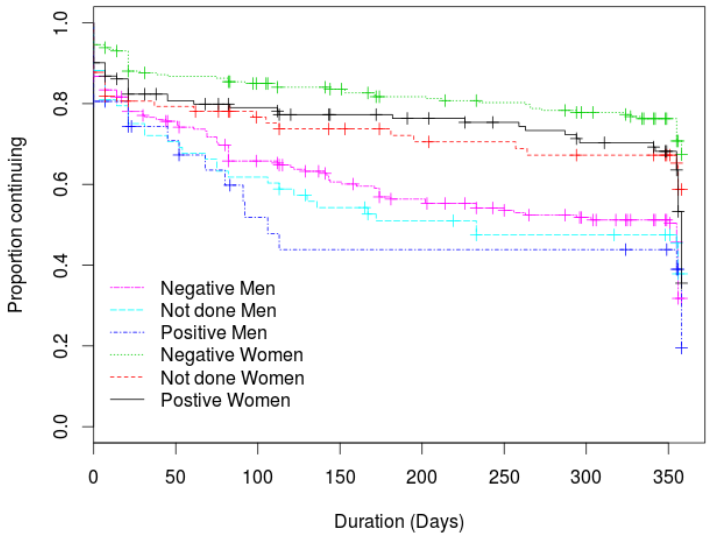

Figure 3.6: The Kaplan-Meier curves for the Joint Effect of HIV status and Gender for the First Episode.

Figure 3.6 shows that the probability of continuing an episode for women that are HIV negative is higher than the rest of the combinations, also at t less than 100 days, the probability of continuing an episode for men that are HIV positive and those that did not do the test is almost the same.

3.3 Fitting of Lifetime Regression Model

We shall use accelerated time models to determine the association between the duration of the relationship and the selected variables. The Accelerated Life Model was chosen because it has specific reference and impact on time (Lawn et al., 2011).

In order to make the choice of which distribution to use, we shall check for the distribution that has the largest likelihood and also find the Akaike's Information Criterion (AIC): $AIC = -2logL + 2p$. The AIC may be used for model selection by fitting several models and determining which has the lowest AIC.

Table 3.4: Table showing the likelihood and AIC for HIV Status for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2480.9	4961.8	4967.823
Weibull	-2401.2	4802.4	4810.406
Log-logistics	-2397.0	4794.0	4802.057
Log-normal	-2382.4	4764.8	4772.832

Table 3.5: Table showing the likelihood and AIC for Race for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2459.8	4919.6	4921.632
Weibull	-2382.2	4764.4	4767.361
Log-logistics	-2378.7	4757.4	4760.48
Log-normal	-2365.1	4730.2	4733.126

Table 3.6: Table showing the likelihood and AIC for Age for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2413.8	4827.6	4847.639
Weibull	-2339.6	4679.2	4703.296
Log-logistics	-2338.0	4676.0	4699.911
Log-normal	-2325.8	4651.6	4675.557

Table 3.7: Table showing the likelihood and AIC for Condom Use for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2442.5	4885.0	4890.932
Weibull	-2369.5	4739.0	4747.013
Log-logistics	-2364.3	4728.6	4736.513
Log-normal	-2353.0	4706	4713.906

Table 3.8: Table showing the likelihood and AIC for Language for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2473.0	4946.0	4952.087
Weibull	-2395.0	4790.0	4798.022
Log-logistics	-2392.2	4784.4	4792.394
Log-normal	-2378.6	4757.2	4765.138

Table 3.9: Table showing the likelihood and AIC for Gender for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2367.4	4734.8	4736.866
Weibull	-2302.8	4605.6	4608.665
Log-logistics	-2300.6	4601.2	4604.210
Log-normal	-2290.8	4581.6	4584.564

Table 3.10: Table showing the likelihood and AIC for Age Difference for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2379.3	4758.6	4764.539
Weibull	-2304.6	4609.2	4617.239
Log-logistics	-2301.0	4602.0	4610.045
Log-normal	-2287.2	4574.4	4582.354

Table 3.11: Table showing the likelihood and AIC for Coital Frequency for the First Episode.

Distributions	Log L	-2*Log L	AIC
Exponential	-2323.2	4646.4	4666.553
Weibull	-2252.1	4504.2	4528.297
Log-logistics	-2250.3	4500.6	4524.651
Log-normal	-2237.4	4474.8	4498.759

From Tables 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, the log-normal distribution gave the largest likelihood and at the same time lowest AIC, so this makes the log-normal distribution suitable for fitting an Accelerated Life Model to our data.

From Table 3.12, the analysis indicated that the participants that are coloured, those that use a condom irregularly and those that have never used any, women, those that have coital frequency at least three times in a week, age group 38 to 47 years, the English and Xhosa speakers were significantly related to the duration of the relationship. On the log-time table, the mean relationship duration of the coloured

Table 3.12: Univariate model of log-normal regression with chosen variables for first episode relationship.

Variables	Group name	Estimate	95% CI	p-value
HIV Status	Negative(Baseline)	1		
	Not done	0.712	(0.419, 1.212)	0.202
	Positive	0.844	(0.507, 1.405)	0.504
Race	Black(baseline)	1		
	Coloured	2.022	(1.770, 3.472)	0.009
Condom Use	Always(baseline)	1		
	Irregularly	2.008	(1.249, 3.227)	0.003
	Never	4.191	(2.567, 6.844)	<0.050
Gender	Men(baseline)	1		
	Women	3.353	(2.229, 5.045)	<0.050
Age(years)	18-27(baseline)	1		
	28-37	1.522	(0.929, 2.494)	0.090
	38-47	4.679	(2.532, 8.645)	<0.050
	48-57	1.296	(0.633, 2.651)	0.470
	58-67	2.052	(0.745, 5.658)	0.156
Age Difference	Equal age(baseline)	1		
	Male older	1.097	(0.474, 2.536)	0.825
	Female older	0.803	(0.161, 4.003)	0.635
Coital Frequency	4(baseline)	1		
	1	1.435	(0.651, 3.160)	0.361
	2	2.179	(1.106, 4.293)	0.051
	3	2.545	(1.086, 5.962)	0.028
	> 4	0.856	(0.337, 2.174)	0.739
Language	Afrikaans(baseline)	1		
	English	0.355	(0.173, 0.729)	0.004
	Xhosa	0.474	(0.266, 0.844)	0.010

race is twice that of the black. For those that have never used a condom, their mean relationship duration is four times longer than participants that always use condoms and double for those that use it irregularly. Also, the age group 38 to 47 years has their mean relationship duration almost five times higher than age group 18 to 27 years. Female participants have a mean relationship duration three times longer than that of male participants.

In order to fit the bivariate and multivariate model, we shall use a method of backward selection to select variables that should be considered. From the univariate table, age, condom use, coital frequency, race, and gender are those variables that are significantly related to the relationship duration since their p-value are less than the threshold value. Hence, these variables shall be considered in the models.

Table 3.13: Bivariate model of log-normal regression for first episode relationship.

Variables	Coefficient	Estimate	95% CI	p-value
Intercept	5.476			
Coloured	0.627	1.871	(1.084, 3.228)	0.022
Age	0.017	1.017	(1.002, 1.033)	0.027

Loglik(model)= -2362.2 Loglik(intercept only)= -2368.2

$$\log T = 5.476 + 0.627 \text{ Coloured} + 0.017 \text{ Age}$$

Variables	Coefficient	Estimate	95% CI	p-value
Intercept	4.588			
Women	1.225	3.403	(2.265, 5.113)	< 0.050
Age	0.027	1.027	(1.012, 1.043)	< 0.050

Loglik(model)= -2284 Loglik(intercept only)= -2308.5

$$\log T = 4.588 + 1.225 \text{ Women} + 0.027 \text{ Age}$$

Variables	Coefficient	Estimate	95% CI	p-value
Intercept	5.501			
Cf2	0.828	2.289	(1.028, 5.094)	0.039
Cf3	0.914	2.563	(1.093, 6.006)	0.032
Coloured	0.641	1.898	(1.898, 3.355)	0.024

Loglik(model)= -2223.9 Loglik(intercept only)= -2234.6
Cf2: coital frequency=2; Cf3: coital frequency=3

$$\log T = 5.501 + 0.828 \text{ Cf2} + 0.914 \text{ Cf3} + 0.641 \text{ Coloured}$$

Table 3.14: Multivariate model of log-normal regression for first episode relationship.

Variables	Coefficient	Estimate	95% CI	p-value
Intercept	4.051			
Cf3	0.945	2.573	(1.097, 6.037)	0.027
Age	0.026	1.026	(1.010, 1.042)	0.001
Women	1.245	2.563	(2.285, 5.282)	< 0.050
Coloured	0.564	1.758	(0.990, 3.121)	0.040

Cf3: coital frequency=3

$$\log T = 4.051 + 0.945 \text{ Cf3} + 0.026 \text{ Age} + 1.245 \text{ Women} + 0.564 \text{ Coloured}$$

3.4 Model Evaluation: Residual Plots

In addition to testing the assumption of Accelerated Life Model, the validity of both non-parametric and parametric regression models can further be assessed by checking the overall model fit using Cox-Snell residuals, identifying outliers using deviance residual and investigating the influence of individual observations on the overall model fit. But in survival analysis residual plots are difficult to understand because of censoring. So, we shall fit log-normal model to the Kaplan-Meier plots.

From the fitted log-normal model we can observe that the model fit more reasonably well after about 100 days for Figures 3.9, 3.10, 3.11, 3.12 and fit well from day one for male and black in Figure 3.7 and Figure 3.8 respectively.

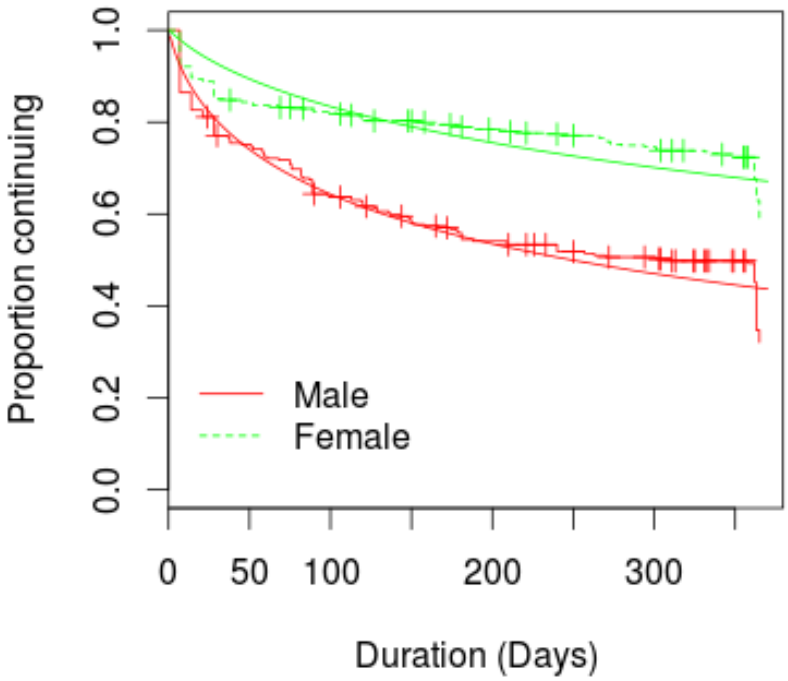

Figure 3.7: Fitting of the lognormal model estimation to the Kaplan-Meier curve for gender.

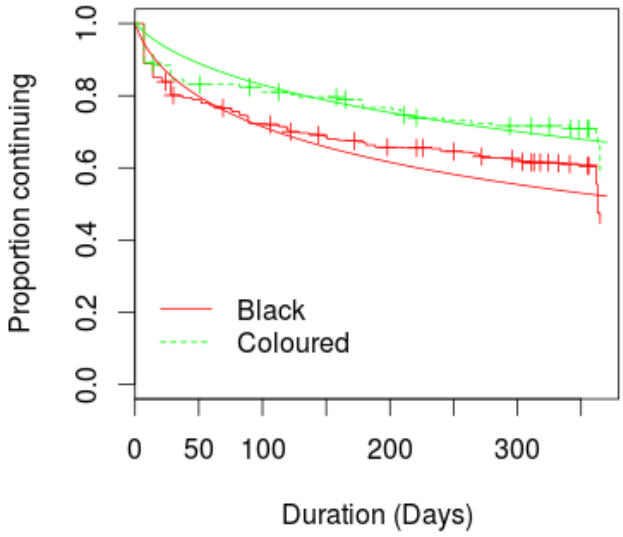

Figure 3.8: Fitting of the lognormal model estimation to the Kaplan-Meier curve for race.

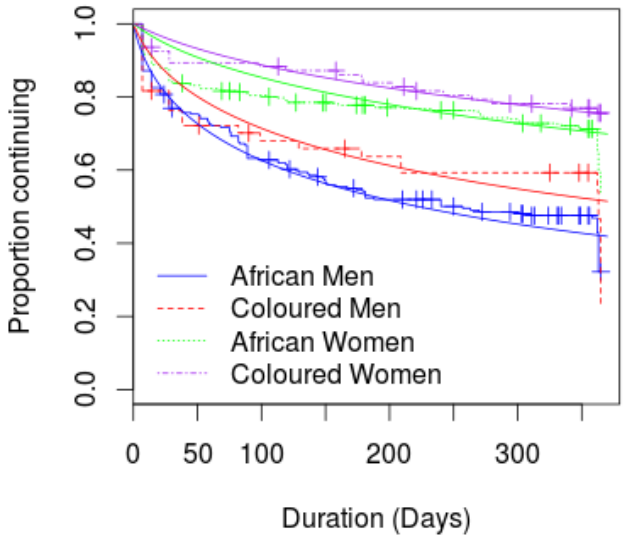

Figure 3.9: Fitting of the lognormal model estimation to the Kaplan-Meier curve for gender and race.

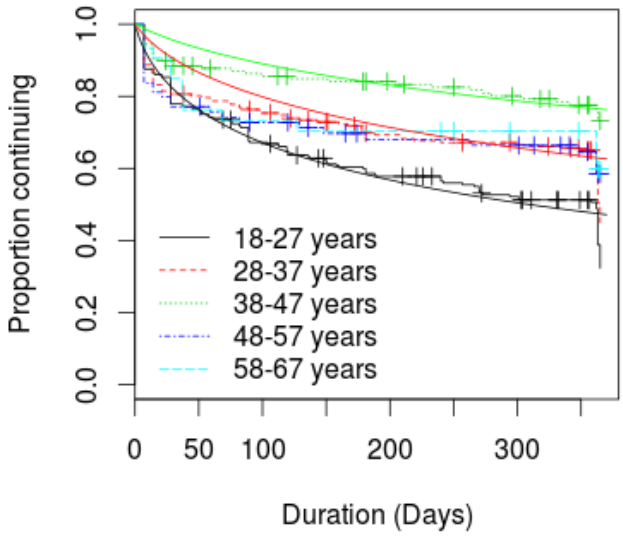

Figure 3.10: Fitting of the lognormal model estimation to the Kaplan-Meier curve for age groups.

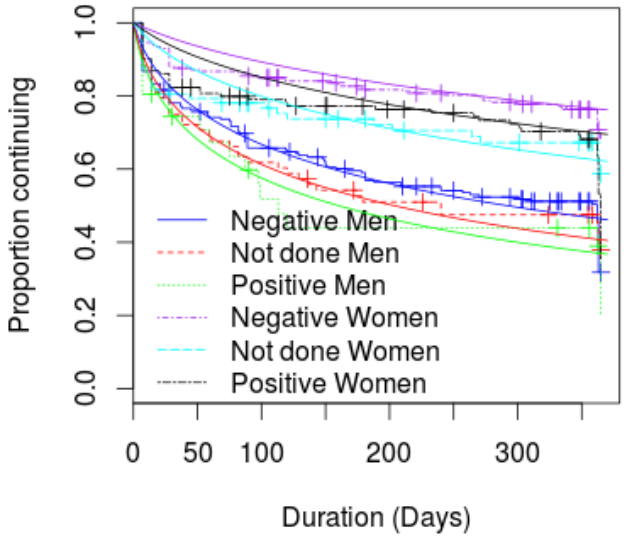

Figure 3.11: Fitting of the lognormal model estimation to the Kaplan-Meier curve for gender and HIV status.

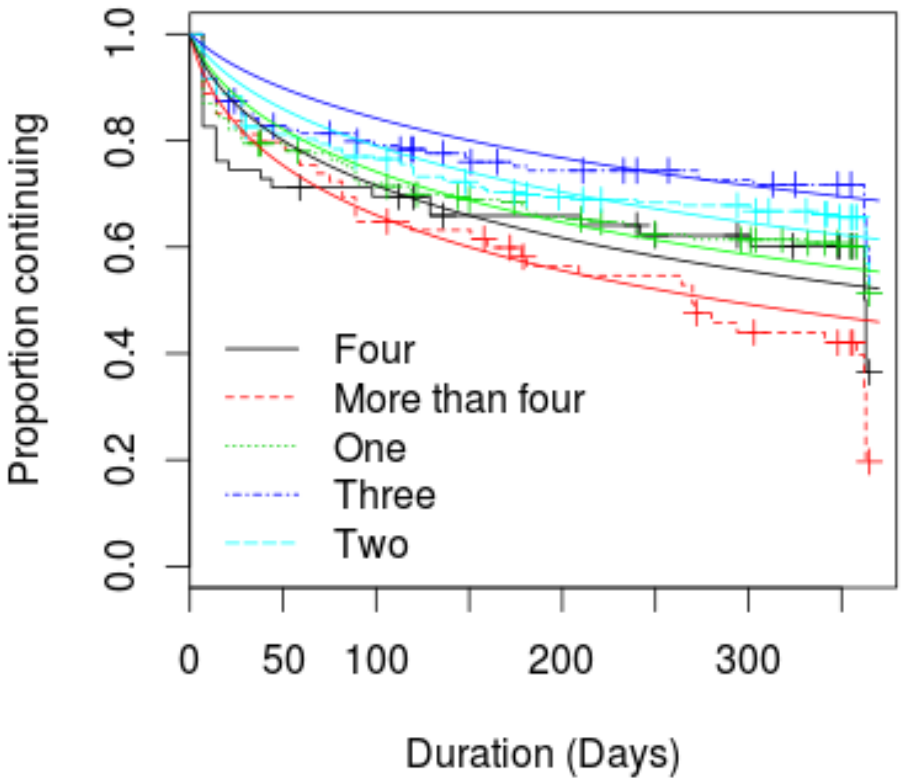

Figure 3.12: Fitting of the lognormal model estimation to the Kaplan-Meier curve for coital frequency.

4. Discussion and Conclusion

4.1 Discussion

This study established under the parametric method of survival analysis that the time spent in a relationship of an HIV positive individual is less than those that are negative and even those that did not do the test. This may be as a result of death and stigma that accompanied HIV (Serwadda et al., 1995). We also found that the coloured-race stays longer in a relationship than the black-race and some of the research findings indicated that the coloured-race tends to be less affected by HIV than other racial groups (Shisana et al., 2005). Results also indicate that participants that had sexual intercourse more than once a week are likely to persist longer in a relationship. These are in line with results from a previous study over South African, which found that " being male, coloured, having tertiary education, and having a relationship between 2 weeks and 9 months were associated with higher coital frequencies " (Delva, 2013).

For participants that are within the age group 38-47, the parametric method of survival analysis has established that they persist longer in a relationship episode than the rest of the age groups and, the log-rank test shows the significance in the test of the age groups. Also, for an Afrikaans language-speaking participant, the result has shown that they have a higher percentage of staying longer in a relationship (68.1%) than those that speaks English or Xhosa language. However, the indication of coloured-race participants and Afrikaans language-speakers having a long relationship was equally a result of the majority of coloured persons speaking Afrikaans, which make them the same people.

The result of the univariate analysis shows the significance of race, gender, condom use, age, and language to the duration of relationship. It also shows that the mean relationship duration of the coloured race is twice that of the black and for those that have never used a condom, their mean relationship duration is four times longer than a participants that always uses a condom and double for him/her that use them irregularly. In addition, the age group 38 to 47 years has their mean relationship duration almost five times higher than the age group 18 to 27 years. Moreover, female participants have a mean relationship duration three times longer than that of male participants.

The analysis in the bivariate table indicated that after adjusting for race, the effect of age is an increase of about 0.17 per decade, whereas after adjusting for gender the effect of age is about 0.27 per decade. The difference by race, which is 0.627, as about half the difference by gender, 1.225. The model with coital frequency and race describes the difference in length of first episode better than the other two models since it has the largest likelihood. Table 3.14 established that the effect of age, gender and race are alike.

4.2 Conclusion

In conclusion, we found in our study that race, gender, age and coital frequency are factors that contributes to a durable relationship. However, no evidence was found for age-difference and HIV status, the log-normal univariate analysis showed them to be insignificant to the duration of a relationship. Further more, the effect of race and language are similar, since the majority of people that belong to a particular race speaks the same language.

4.3 Future work

It would be interesting in the near future to examine factors that constitute to short term duration of other relationships episodes because the first episodes are shown to be longer than the second and third episodes in Table 3.3. In addition, one might be interested in examining other factors that contribute to the long lasting relationship of coloured-race and also estimate the rates of transitioning between relationship status options.

Acknowledgements

I give thanks to the Almighty God, the prince of peace for the peace He granted me throughout my course of studies. I also want to appreciate the founder of AIMS Prof. Neil Turok for allowing himself to be a source of blessing to others, thank you so much. I am deeply grateful to Prof Jeff Sanders AIMS Academic Director, who has been as a father to me during my stay at AIMS; thanks for your timely advice and encouragement. To all AIMS staff starting from the Director Prof. Barry Green, Jan, Igsaan, Tutors, and others that time will not permit me to mention their names, thank you for your care and support.

My appreciation also extends to my supervisor Prof. Jane Hutton not just a supervisor, but a mother. Thanks to my co-supervisor Dr. Wim Delva for his support. My sincere appreciation goes to my best friend Yusuf Owolabi thank you for your prayers, love and care. To my wonderful parents, Mr and Mrs Olaniyi Familusi, my siblings and all that have made a positive impact in my life, thanks and God bless you all.

References

AFSA. HIV/AIDS in South Africa. AIDS Foundation of South Africa, Accessed May 2014. URL http://www.aids.org.za/hivaids-in-south-africa.

H. Ahn. Log-normal regression modelling through recursive partitioning. *Computational Statistics and Data Analysis*, 21:381–398, 1996.

P. Audrey, R. Helen, K. Immo, S. Annie, M. Catherine, H.-M. Lindiwe, and V. kerri. Young people's sexual health in south africa: Hiv prvalence and sexual behaviors from a nationally representative household survey. *AIDS*, 19:1525–1534, 2005.

A. Bankole, F. H. Ahmed, S. Neema, C. Ouedraogo, and S. Konyani. Knowledge of correct condom use and consistency of use among adolescents in four countries in sub-saharan africa. *African journal of reproductive health*, 11(3):197, 2007.

R. Beauclair, F. Meng, N. Deprez, M. Temmerman, A. Welte, N. Hens, and W. Delva. Evaluating audio computer assisted self-interviews in urban south african communities: evidence for good suitability and reduced social desirability bias of a cross-sectional survey on sexual behaviour. *BMC Medical Research Methodology*, 13:2288, 2013.

C. Campbell. Male gender roles and sexuality: implications for women's aids risk and prevention. *Social Science and Medicine*, 41:197–210, 1995.

W. Delva. Description of the cape town sexual network survey data, 2012. Unpublished manuscript.

W. Delva, F. Meng, R. Beauclair, N. Deprez, M. Temmerman, A. Welte, and N. Hens. Coital frequency and condom use in monogamous and concurrent sexual relationships in cape town, south africa. *Journal of the International AIDS Society*, 16:18034, 2013.

M. K. Goel, P. Khanna, and J. Kishore. Understanding survival analysis: Kaplan-meier estimate. *International Journal of Ayurveda Research*, 14:274–278, 2010.

D. Hocking. World health organisation. *Global update on HIV treatment 2013: Results, Impact and Opportunities*, 2013.

HSRC. Launch of the 2012 south african national hiv prevalence, incidence and behaviour survey report. Human Sciences Research council, Accessed May 2014. URL http://www.hsrc.ac.za/en/media-briefs/hiv-aids-stis-and-tb/sabssm4-launch.

H. Kelly. Socioeconomic disadvantage and unsafe sexual behaviors among young women and men in south africa. *Policy Research Division Working*, page 190, 2004.

S. D. Lawn, L. Campbell, and R. Wood. Time to initiation of antiretroviral therapy among patients with hiv-associated tuberculosis in cape town, south africa. *Journal of acquired immune deficiency syndromes (1999)*, 57:136–140, 2011.

N. Luke, P. Maharaj, J. Cleland, F. Muramutsa, L. Salazar, R. DiClemente, G. Wingood, R. Crosby, K. Harrington, B. Halpern-Felsher, et al. Women hiv and aids. revised. *International Family Planning Perspectives*, 31(1):6–14, 2013.

PLHL. HIV/AIDS in South Africa, a look at the current stats. Positive Living healthier lives, Accessed May 2014. URL http://www.sapositivemagazine.co.za/hiv-numbers.

D. Serwadda, R. H. Gray, M. J. Wawer, R. Y. Stallings, N. K. Sewankambo, J. K. K.-L. B. Lainjo, and R. Kelly. The social dynamics of hiv transmission as reflected through discordant couples in rural uganda. *Aids*, 9(7):745–750, 1995.

N. Shaikh and L. Smith. The 2005 hiv antenatal provincial and area surveys western cape. 2005.

O. Shisana, T. Rehle, L. Simbayi, W. Parker, K. Zuma, A. Bhana, C. Connolly, S. Jooste, V. Pillay, et al. South african national hiv prevalence, hiv incidence, behavior and communication survey, 2005. *Cape Town, South Africa: Human Sciences Research Council Publishers*, 2005.

O. Shisana, T. Rehle, L. Simbayi, K. Zuma, and S. Jooste. South african national hiv prevalence incidence behaviour and communication survey 2008: a turning tide among teenagers. 2009a.

O. Shisana, T. Rehle, L. Simbayi, K. Zuma, and S. Jooste. South african national hiv prevalence incidence behaviour and communication survey 2008: a turning tide among teenagers? 2009b.

G. Simon, N. Constance, G. Geoff, M. PR, Z. Tom, C. Michel, C. SK, and A. R. M. Sexual mixing patterns and sex-differentials in teenage exposure to hiv infection in rural zimbabwe. lancet.2002a;359(9321):1896-903.

UNAIDS. HIV consensus indicators are needed for concurrency. UNAIDS Reference Group on Estimates Modelling and Projections *Lancet 375:621–2*, 2010.