RISK FACTORS ASSOCIATED WITH AND FACTORS THAT INFLUENCE INTIMATE PARTNER VIOLENCE

A CASE STUDY OF SUB-SAHARAN REGIONS

BY

TALANI MHELEMBE (217075005)

A Thesis Submitted in Fulfilment of the Requirements of the Degree for MASTER OF SCIENCE in STATISTICS

Supervisor: PROFESSOR S. RAMROOP

and

Co-Supervisor: DR. F. HABYARIMANA

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

2021

Dedication

Glory be to the Almighty.

To my late father, Cassius Maxegana Mhelembe.

To my mother, Patricia Mantombi Ndlovu.

To all women who have experienced violence from their partner.

Declaration

I, Talani Mhelembe, declare that this thesis titled, 'Statistical models to understand risk factors associated with and factors that influence intimate partner violence - a case study of sub-Saharan regions' and the work therein is my own. I confirm that:

- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the sources are always acknowledged.
- I have acknowledged all primary sources of help.

With this, I confirm that all passages that are literally or in general matter taken out of publications or other sources, are referenced.

[redacted] 06/ 02/ 2022

Author: Talani Mhelembe Date

[redacted] 06/02/22

Supervisor: Prof. Shaun Ramroop Date

[redacted] 06/02/22

Co-Supervisor: Dr. Faustin Habyarimana Date

Submitted papers

The following papers from this thesis have been submitted for publication

1. Talani Mhelembe, Shaun Ramroop and Faustin Habyarimana. Prevalence of and risk factors associated with intimate partner violence in women of reproductive age from South Africa and Uganda. The Open Public Health Journal. Status - Published.
2. Talani Mhelembe, Shaun Ramroop and Faustin Habyarimana. Risk factors associated with and factors that influence intimate partner violence, a case study of South Africa and Uganda. Songklanakarin Journal of Science and Technology. Status – In Review

Acknowledgments

I want to express my sincerest gratitude to my supervisor, Prof Shaun Ramroop, and my co-supervisor, Dr. Faustin Habyarimana, for their guidance, encouragement, patience, advice, and time. Every consultation with each of them was an illuminating experience. They helped me understand and apply the statistical concepts correctly and efficiently. They were always available to assist me in whatever problem I experienced in this research.

I am also grateful for the facilities made available to me by the School of Mathematics, Statistics, and Computer Science, UKZN. I want to thank the School of Mathematics, Statistics, and Computer Science, Pietermaritzburg campus, for their conducive environment. I also gratefully acknowledge financial support from Absa Bank Limited.

I am also thankful to several people who directly or indirectly contributed to this work. In particular, thanks are due to Buster Mogonong, Muziwandile Mlondo, Thandokazi Jantjies, Sithandiwe Zuma, Mohanad Mohammed, and Christel Bernard.

My mother, M.P Ndlovu, my uncles, U. Mhelembe and my grandfather, B. Mhelembe, are always pillars of my strength; this is an opportunity to thank them for their sacrifices in ensuring that my studies are completed successfully.

Abstract

The reduction of intimate partner violence is critical to most societies' well-being and posterity, and for policymakers. However, in most cases, coming up with an accurate, intimate partner violence evaluation tool that focuses on vulnerable women, is a challenge for applied policy research. Intimate partner violence for women of conceptive age (15-49 years) has been measured utilizing the number of cases reported, and this approach has several underlying problems. Therefore, in this work, we came up with a rating scale from Demographic and Health Survey data as an alternate method to measure (Chapman & Gillespie, 2018) intimate partner violence, and examine different statistical methods suitable for identifying the associated factors. A generalized linear mixed model technique was utilized to elongate survey logistical regression to incorporate random effects, and account for variability amongst the primary sampling units. This was done to account for the complexity of the sampling design and the ordering of outcome variables. We have also utilized the generalized additive mixed model to ease the assumptions of normality and linearity intrinsic in linear regression models, in which categorical independent predictors were modeled by parametric model, continuous covariates, and interaction between the continuous and categorical variables by non-parametric models.

Each of these models has inherent flaws and strengths. The choice of a statistical model depends on the objectives to be achieved. The findings from this current scientific setting revealed that the following determinants are the key factors influencing intimate partner violence: age of the woman's partner, marital status, region where the woman lives, age of the woman, media exposure, size of the family, polygamy, sex of the household head, wealth index, pregnancy termination status, body mass index, marital status, cohabitation duration, partner's desire for children, partner's education level, woman's working status, and woman's earnings compared to partner's earnings.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 Introduction

Violence inflicted upon women, in particular sexual violence, and intimate partner violence, are crucial public and clinical health problems and an infringement of women's human rights (WHO, 2020). According to WHO (2020), globally, one third of women encounter sexual or physical violence in their lifetime, mainly inflicted by an intimate partner (WHO, 2002; WHO, 2020). In a report by United Nations (1993), violence inflicted upon women was described as "any form of gender-related violence that yields, or is plausible to end in: mental harm, physical harm, misery to women, and sexual harm, also considering the menace of such acts, coercion or unjustified deprivation of freedom, whether happening in general or private life" (United Nations, 1993).

Violence is a complex event; describing this event is not an exact science but a matter of judgment (WHO, 2002). Violence can be defined in many ways; it only depends on who defines it and for what purpose (WHO, 2002). WHO (1996) described violence as the premeditated usage of physical strength, threatened upon a group, that yields a significant likelihood of psychological harm, maldevelopment, injury, deprivation, or death (WHO, 1996).

Household violence is a growing public health concern in evolving countries (Koeing, et al., 2003). Scientific proof highlights the immensity of domestic violence in evolving countries and corroborates its trends globally (Koeing, et al., 2003). In sub-Saharan Africa, observed proof on the prevalence of household violence is limited to a compact number of population-based or special-population studies (Jewkes, et al., 2001; Watts, et al., 1998).

Interpersonal violence refers to violence amongst people, and is sub-divided into intimate partner violence, family violence, and group violence (Koeing, et al., 2003). The family violence category includes child mistreatment, intimate partner violence, and elderly abuse (Jewkes, et al., 2001; Koenig, et al., 2003; VPA, 2002). The last-mentioned is split into stranger and associate violence, and incorporates: teenager violence, where a teenager is assaulted by strangers; violence associated to assets crimes; and violence in places of work and other institutions (VPA, 2002). Intimate partner violence describes the behavior by the current intimate partner or ex-partner that resulted in psychological, sexual, and physical harm, incorporating physical assault, sexual coercion, psychological abuse, and governing behaviors (WHO, 2020).

In 2002, about 1.5 million people died because of deliberate acts of self-directed, interpersonal, or combined violence (Krug, et al., 2002). Many thousands more are wounded or endure other non-fatal health outcomes because of being the sufferer or witness to certain acts of violence (Krug, et al., 2002):

WHO (2010) found that about 5.8 million individuals die each year because of injuries. These results account for 10% of the world's fatalities, compared to 32% for the number of casualties that resulted from malaria, tuberculosis, and HIV/AIDS combined (WHO, 2010). Close to one-third of the 5.8 million casualties from injuries result from violence – suicide, homicide, and war; about one-third of the deaths are because of domestic violence (WHO, 2010).

About 35% of women globally have endured physical or sexual violence from a current or non-intimate partner (WHO, 2013). There are different types of violence that women may face, and they make up a great percentage of the world's women; the main type of violence is intimate partner violence (WHO, 2013). About 30% of women in relationships have endured physical or sexual violence inflicted by an intimate partner (WHO, 2013).

In certain regions, about 38% of women have encountered intimate partner violence; globally, close to 38% of murders of women are committed by intimate partners (WHO, 2013). Women who have experienced violence, either sexually or physically from their partners, report inflated rates of several basic health problems: for instance, they are about 16% more likely to have a low-birth-weight child (WHO, 2013).

From 1986-1993 (Bowman, 2003), the percentage of women who reported physical abuse by male partners was compared in some African countries and the following were the results: Tanzania 60%, Uganda 46%, Kenya 42%, and Zambia 40% (Bowman, 2003). Also, a survey conducted in Ghana in 1998 showed that one in three of the women had been beaten, slapped, or physically punished by a current intimate partner or most recent intimate partner (Bowman, 2003).

Women exposed to partner violence are twice as likely to terminate a pregnancy, and almost two times more likely to suffer from depression (WHO, 2013). In certain areas, they are about 1.5 times more likely to contract HIV compared to women who have not endured partner violence (WHO, 2013). Evidence shows that women who have endured this type of violence are about 2.3 times more likely to have the tendency to abuse alcohol, and about 2.6 times more likely to suffer from anxiety (WHO, 2013).

Intimate partner violence significantly impacts the livelihoods of people who are close to the victims or perpetrators of violence. It is evident from the figures mentioned above that this topic is of great concern to many individuals and societies. The resulting health problems to most victims suggests that communities ignore this topic or have little to no knowledge of the concept of violence. This study will address some of the issues experienced by different individuals and find the risk factors driving the increasing rate of intimate partner violence.

1.1.1 Problem Statement

Violence can be defined in different ways: it depends on which context one is dealing with at that specific moment. Many people get into relationships with the hope of having a better life and building a future with their partners. Some of the relationships end with one partner in hospital, jail, or the morgue. If a model can predict the chances of a relationship leading to violence, most people would be aware of what they are getting themselves into. If some of the risk factors leading to intimate partner violence could be eliminated, then the rate of violence could possibly be reduced.

1.1.2 Aim

The aim of this study is to investigate the risk factors that influence intimate partner violence, and determine how likely it is for some individuals to be exposed to violence in the lifespan of their relationship, in three selected sub-Saharan African countries.

1.1.3 Objectives

1. To statistically describe intimate partner violence in sub-Saharan regions.
2. To investigate some of the risk factors that influence intimate partner violence in sub-Saharan regions.
3. To ascertain the merits and demerits of statistical models relevant in modelling intimate partner violence.

1.2 Literature Review

1.2.1 Definitions

The study by Finnbogadottir, et al. uses Swahnberg & Wijma's definitions for extreme abuse, categorized as light, moderate, or critical and the type of abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where one is constantly being humiliated, degraded, and repressed is known as mild emotional abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where one is threatened or forcibly confined regarding association with others or subjected to absolute control regarding what one may and may not do, is known as moderate emotional abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where one constantly lives in fear due to persistently being threatened by a close individual, is known as critical emotional abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). Examples of light physical abuse are being hit, slapped in the face, or held in constraining restraint (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where one is kicked, punched, pushed violently, or beaten is known as moderate physical abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where one is being threatened by a weapon or is strangled is known as critical physical abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014).

Swahnberg & Wijma (2003) further define sexual abuse as mild, moderate, and severe. The type of abuse where one is poked in certain body parts, apart from the genitals, or forced to be in contact with the next person in a sexual way is known as mild sexual abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where a woman is poked in the genitals without their consent, or forced to gratify themselves in a sexual way, or touching someone's genitals is known as moderate sexual abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The type of abuse where there is forced penetration of the penis into any body part of a woman is known as critical sexual abuse (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014).

The history of violence has been defined by Finnbogadottir et al. (2014) as the lifetime encounter of emotional, physical, or sexual abuse, happening during childhood (<18 years), adulthood (≥18 years), or both, regardless of the level of abuse or the perpetrator's identity, following the operationalization of the questions in the NorAQ (Swahnberg & Wijma, 2003; Finnbogadottir, et al., 2014). The statistical method used by Finnbogadottir et al. (2014) was bivariate logistic regression. With the dependent variable being 'DV during pregnancy,' many of the variables selected were dichotomous.

1.2.2 Domestic Violence in Africa

In Africa, most women live in rural areas and follow African customary law, reinforcing the lower-ranking position of women inside the family (Bowman, 2003). Moreover, there is a greater proportion of household violence in African countries compared to some of the American countries (North America), the rate being almost double in certain regions (Bowman, 2003). For instance, the proportion of physical or sexual violence against women of conceptive age in Rwanda, showed a continual rise (2010-2016) (Habyarimana, et al., 2018). Nonetheless, in recent times a decreasing trend has been observed (Habyarimana, et al., 2018). The study included psychological kinds of abuse such as emotional violence (Habyarimana, et al., 2018). In a study by the same authors a module, based on questionnaires, was included on domestic violence for men and women aged between 15-59 and 15-49 years, respectively (Habyarimana, et al., 2018).

Gender-based violence (GBV) is usually considered the 'tip of the iceberg' or a 'silent epidemic', as victims are hesitant to speak out about their experiences of violence, due to certain barriers (Palermo, et al., 2014). The barriers leading to unreported cases of GBV are, but not limited to: fear of being humiliated, monetary barriers, lack of knowledge of available assistance, fear of retaliation, minimal law enforcement efforts, and views surrounding violence as a standard element of life (Muluneh, et al., 2020). A multi-country study by the World Health Organization (WHO) of women of conceptive age showed that the general prevalence of intimate partner violence (IPV) fluctuates between 15% in places such as Japan, to 71% in places such as Ethiopia (Pallitto, et al., 2006).

Evidence shows that the issue is most prominent in developing countries with a low socio-economic status and limited education, especially in sub-Saharan African (SSA) countries (Abrahams, et al., 2006;

Joanne, 2018). Research carried out in SSA countries concentrated on small-scale studies such as in regions and sub-regions, in particular countries that could possibly overestimate the prevalence of GBV (Pallitto, et al., 2006; Garcia-Moreno & Pallitto, 2013). Most African cultural beliefs and traditions encourage male dominance in sexual relationships, especially where marriage is concerned (Morrell, et al., 2012).

Results showed that the prevalence varies from as low as 13.9% (Schneider, et al., 2018), in a study carried out among perinatal women with signs of depression in South Africa, to as high as 97% (Ajah, et al., 2014) in a study conducted among rural women in Nigeria (Ajah, et al., 2014; Schneider, et al., 2018). Muluneh et al. (2020) found that the prevalence of GBV was greater in SSA countries than in other regions, and that emotional IPV was the most common type of violence in SSA (Muluneh, et al., 2020).

The lifetime percentage of experiencing violence from an intimate partner was 24.6% in South Africa, whereas in Nigeria, like in many developing countries where DV commonly occurs, dependable population-based data on violence against women by their partners is scarce (Obi & Ozumba, 2007). Household violence was associated with lower social class, alcohol drinking status, age difference between partners, and employment status (Obi & Ozumba, 2007). In a study by Obi & Ozumba, (2007), about 70% of the male respondents disclosed a history of household violence in their families; in 92% of cases, the female partners were the victims, while 8% of victims were male (Obi & Ozumba, 2007).

Male abuse was sky-high in the early stages of marriage, less than five years age difference and alcohol drinking status were typical risk factors (Obi & Ozumba, 2007). Frequent types of abuse were oral, physical, and forced sex (Obi & Ozumba, 2007).

1.2.3 Links to Domestic Violence (DV)

There are different links to domestic violence that are a growing concern. Several studies show the link between household violence and a range of unfavorable reproductive health outcomes; these incorporate non-use of contraception and unintended pregnancy (Gazmararian, et al., 1996; Martin, et al., 1999). Other links include poor outcomes of pregnancy and childbirth (Valdez-Santiago & Sanin-Aguirre, 1996), and gynecological morbidity (Schei & Bakketeig, 1989; Valdez-Santiago & Sanin-Aguirre, 1996). Other links are with sexually transmitted diseases (STDs) and human immunodeficiency virus (HIV) (Martin, et al., 1999; Maman, et al., 2000). A remarkable inverse relationship was observed between socio-economic status and household violence, using measures of family wealth or the education status of the male partner (Koenig, et al., 2003).

Demographic attributes are also remarkable risk factors for domestic violence, with some researchers having discovered that age disparity and the number of children is associated with a lowered risk of violence (Kim & Cho, 1992). As assessed by their educational achievements, degree of self-governance, or authority over resources, women with a well-established status are more protected from the risk of domestic violence (Koenig, et al., 2003). Some studies in developing countries have also found an association between alcohol consumption status or drugs, and the risk of violence (Rao, 1997; van der Straten, et al., 1998; Watts, et al., 1998).

Maman et al. (2000) found that HIV status is one of the key factors that influences violence amongst individuals. Research in Africa showed a high risk of violence if the man tested positive for HIV (van der Straten, et al., 1998), or if the woman believed that she was at a greater risk of getting HIV from her partner (Coker & Richter, 1998). Children who witness household violence are more likely to become victims or perpetrators of violence in adulthood (Ellsberg, et al., 1999).

Male DV was related to the financial difference favoring the woman, authoritative in-laws, literate women, and a couple within the same age category (Obi & Ozumba, 2007).

1.2.4 Health Concern

Household violence is a concerning public health issue on a worldwide scale. Its prevalence is in both the industrialised and developing countries (Habyarimana, et al., 2018). Men and women perpetuate it, but in most cases, women are more vulnerable to violence than men (Habyarimana, et al., 2018). Intimate partner violence has different impacts on society at large. From the description of intimate partner violence (VPA, 2002), intimate partner violence significantly impacts the victims' physical, psychological, and sexual well-being (VPA, 2002). An intimate partner's physical or psychological abuse is a significant public health issue affecting the medical and judicial professions (Bowman, 2003). Spousal abuse results not only in critical effects to the psychological health of the victims, but also leads to erosion of self-assurance, post-traumatic stress disorder (PTSD), fear, depression, alcohol and drug abuse, and self-harm (Bowman, 2003).

Psychological violence is more difficult to study (Bowman, 2003). In a study in Ghana, most of the women who responded reported that they had been threatened by their partners with a fist or an object (Bowman, 2003). Other responses were: being humiliated or embarrassed in front of other people, being restricted from seeing family and friends, getting barred from employment or having their earnings taken away, and having things that are important to them getting vandalized (Bowman, 2003).

In a study conducted by Brink et al. (1998) in Denmark, findings showed a significant difference in the location of injuries, the injury sustained, and the difference between the mechanisms of injury between men and women victims (Brink, et al., 1998). The mechanism of the injury was mostly blunt trauma inflicted by being punched or kicked, with broken drinking glasses causing most injuries, and knives and guns being used as weapons contributing to about 3.7% of injuries, of which 69% being craniofacial injuries (Brink, et al., 1998). Several studies have dealt with ‘minimal-scale’ injuries, the patterns of injuries, and the mechanisms of injury (Brink, et al., 1998). The sequence of injuries resulting from assault may be different between countries and within communities, due to certain cultural and social factors, and traditions (Brink, et al., 1998).

From the study by Brink, et al. (1998), most women sustained injuries to their faces. This might suggest that they either knew their abuser or they were in fights. Men sustained injuries to the head (especially their face) and hands; neck and occiput injuries were related to female victims, suggesting strangulation (Brink, et al., 1998). Male victims were mostly injured by being kicked, head-butted, or injured by drinking glasses; a broken drinking glass was the most frequent sharp object used (Brink, et al., 1998). Women are predominantly exposed to blunt violence (Brink, et al., 1998). Injuries sustained after being strangled, falling, being pushed against the wall/on the floor, were related to women (Brink, et al., 1998).

Violence against pregnant women is a severe public health issue that threatens maternal and fetal health outcomes (Finnbogadottir, et al., 2014). In Sweden, domestic violence against pregnant women differs widely, ranging from 1.2% to 66% across the different regions (Jasinki , 2004). Globally, the prevalence of intimate partner violence during pregnancy ranges between 2% and 13.5% (Devries, et al., 2010). Most violence against women happens at home, therefore, women are at higher risk of experiencing violence from their partners than from any other person (Garcia-Moreno, et al., 2006).

1.2.5 Domestic Violence-related Factors

In a study by Finnbogadottir et al. (2014), the alcohol consumption status for participants was assessed using AUDIT (Alcohol Use Disorders Identification Test). The authors used different variables in their study: age, country of origin, language, educational status, cohabiting status, employment status, and financial status. Furthermore, some maternal characteristics concerning body mass index (BMI) were amongst the variables such as smoking status, use of snuff, unintended pregnancy, and abortion/miscarriage status (Finnbogadottir, et al., 2014).

The study by Habyarimana et al. (2018) focused on whether the occurrence of violence was linked to the wife or husband, whether some community members were responsible, and whether the family members were incriminated (Habyarimana, et al., 2018). Socio-demographic attributes of women were age group, educational attainment, employment status, the number of intimate partners including husband in the previous year, asset ownership such as a house, and the woman’s income compared to her partner’s (Habyarimana, et al., 2018).

The socio-economic and demographic attributes of a woman’s partner, were also assessed, such as: their educational level, employment status, alcohol drinking status, and polygamy (Habyarimana, et al., 2018). The family or community attributes were the number of family members, wealth index, type of residence, province of residence, person who makes decisions on family visits, large household expenses, wife's healthcare, and the usage of money earned by the woman's partner (Habyarimana, et al., 2018).

Jewkes et al. (2002) found that domestic violence was significantly related to a history of violence during a woman's childhood, a woman lacking advanced education, stigma about women's roles, her alcohol drinking status, having other partners during the year, having a person a woman confides in, her partner's male child preference, the man's alcohol drinking, whether the partner is financially supporting the family, general conflicts, and the province of residence (Jewkes, et al., 2001; Jewkes, et al., 2002). The study by Habyarimana et al. (2018) showed that women with a drinking (alcohol) partner experienced more violence than women with partners who do not drink.

1.2.6 Reporting of Cases

In South Africa, about 6% of the victims reported the matter to law enforcement (Bowman, 2003). Law enforcement personnel in South Africa dragged their feet in responding to violence related calls; in some cases, the woman was counseled not to waste her time on legal proceedings, but advised to go back to her partner, and in some instances, she was even actively encouraged to drop the legal case (Bowman, 2003). The situation is even worse for women whose male partners are policemen: here, there have been a few cases in which women have been driven to commit suicide (Bowman, 2003). In one case, where a woman who was married to a policeman logged a complaint with her partner's senior (Bowman, 2003), he discarded the woman's complaint because her partner was a good policeman, and around November 1994, she committed suicide (Bowman, 2003).

In developing countries, the police are without the resources or training to deal with domestic abuse, as they are not paid enough and corruption is rife (Bowman, 2003). Thus, the victim may find that a case was not pursued after the abuser paid off the police officer involved (Bowman, 2003).

1.2.7 Limitations of other Studies

In a study by Brink et al. (1998), the injuries sustained by women or men who have a partner who drinks alcohol, are associated with domestic violence (Brink, et al., 1998; Habyarimana, et al., 2018). In several studies, the employment status of the male partner was investigated, but the question of what kind of job he was currently doing, was not examined. Some of the studies focused on the provincial level of a specific region but did not focus on the region's entire population under study.

In the relevant available studies, the count data was not modeled. The results for count data are essential to better inform the authorities of where the most significant levels of DV are to be expected. In that way, it makes it easy to deploy law enforcers who have been trained to deal with DV cases. A high visibility of law enforcement agents would scare the perpetrators from harming their partners.

1.3 Methods

1.3.1 Study Area

Three African countries have been selected for this study which are in the sub-Saharan region. The countries that have been selected are South Africa, Tanzania, and Uganda. These countries were selected based on the availability of the most recent datasets. To compare the results from these countries, datasets from the same year are required. The datasets from 2015-2016 was the most recent for the three countries at a time when this study was carried out. This study will focus on all the regions in each of the selected countries, rather than on several areas selected at random within each country.

These three countries selected fall under the category of developing countries. The prevalence of domestic violence in developing countries is higher than in developed ones (Pallitto, et al., 2006). The total population for each of these countries is about 60 million for South Africa, 60 million for Tanzania, and 46 million for Uganda (Worldometer.info, 2021).

1.3.2 Data Source

The current study will use the data from the 2016 South African Demographic and Health Survey (SADHS), the data from 2015-2016 Tanzania Demographic and Health Survey (TDHS), and the data from the 2016 Uganda Demographic and Health Survey (UDHS).

1.3.3 Model

Most of the studies utilized logistic regression models, (Adjah & Agbemafle, 2016; Audi, et al., 2008; Habyarimana, et al., 2018) amongst others, to analyze data. Logistic regression models are helpful if their assumptions are not violated (Habyarimana, et al., 2018). If the measurement from the same cluster in a complex survey are correlated, the assumptions of independence are then violated (Habyarimana, et al., 2018). In a study by Habyarimana et al. (2018), the issue was dealt with by utilizing the generalized linear mixed model (GLMM), that accounted for random effects, correlation, over-dispersion, and heterogeneity (Habyarimana, et al., 2018).

In this study, several relevant statistical models will be used, such as logistic regression, survey logistic regression, and GLMM, to account for the correlation and over-dispersion of the cases visible in some regions within a country. In addition, the generalized additive mixed model (GAMM) has been utilized to consider the non-linear relationship among the variables.

1.4 Variables

1.4.1 Dependent Variable

The dependent variable will be dichotomous, for whether an individual has experienced some form of violence, or has not experienced any.

1.4.2 Independent Variables

Several studies (Finnbogadottir, et al., 2014; Habyarimana, et al., 2018; Adjah & Agbemafle, 2016) have established some of the factors associated with DV. In Table 1.1, the independent variables are shown and described with their levels outlined.

In this study, the variables added are the use of contraceptive methods and knowledge of STIs. Also, a variable with a chi-square value of less than or equal to 0.05 will be significant for the fitted model.

1.4.3 Selected Variables

The variables that have been selected for the study are shown in Table 1.1 below. The response variable was created using other variables within the data set: It shows whether a woman has experienced some form of violence from her husband or partner. The response variable level is whether the respondent has experienced some form of violence or not. There is also a variable that has been created to check if respondents have access to media: this variable was to check how frequently a woman has watched television, listened to the radio, read a newspaper, and used the internet in the previous month.

Table 1.1: Selected variables

DESCRIPTION	**LEVELS**
Experience of any form of violence in a woman's life	Response variable (1=Yes, 0= No)
Woman's age	Continuous
Region	Different for each of the three countries
Type of place of residence	1= Rural, 2=Urban
Woman's highest educational level	0=No education, 1=Primary, 2=Secondary, 3=Higher
Number of household members	Continuous/Categorical
Sex of household head	1=Male,2=Female
Literacy	0=Cannot read, 1=Able to read
Access to the media	1=Low exposure, 2=Medium exposure, 3=High exposure
Wealth index	1=Poorest, 2=Poorer, 3=Middle, 4=Richer, 5=Richest
Ever had a terminated pregnancy	0=No, 1=Yes
Contraceptive use	0=Not using, 1=Using
Body Mass Index	1=Underweight (BMI<18.5), 2=Healthy (18.5≤BMI<25), 3=Overweight (25≤BMI<30), 4=Obese (BMI≥30)
Current marital status	0=Single, 1=Married, 2=Living with partner
Number of other wives/partners	Continuous
Cohabitation duration	1=0-4, 2=5-9
Partner's desire for children	1=Both want same, 2=Husband wants more, 3=Husband wants fewer, 8=Don't know
Partner's education level	0=No education, 1=Primary, 2=Secondary 3=Higher, 8=Don't know

Continued…

Partner's occupation	0=Unemployed, 1=Employed, 8= Don't know
Woman's occupation	0=Unemployed, 1=Employed, 8=Don't know
Partner's age	Continuous/Categorical
Woman earnings compared to partner	1=More than him, 2=Less than him, 3=About the same, 4=Husband/partner doesn't bring in money, 8=Don't know
The person who usually decides on what to do with a woman's earnings	1=Woman alone, 2=Woman and partner, 3=Woman and another person, 4=Partner alone, 5=Someone else
Knowledge of Sexually Transmitted Infections (STIs)	0=No, 1=Yes, 8=Don't know
Partner drinks alcohol	0=No, 1=Yes, 8=Don't know
Woman's father ever beaten her mother	0=No, 1=Yes, 8=Don't know
Wife-beating attitude	0=Unacceptable, 1=Acceptable, 8=Don't know

1.4.4 Statistical Software

Statistical Analysis System (SAS) and SAS Enterprise were used to model the datasets. R-studios was also used to analyze the data and plot some of the necessary and significant results using Excel, SPSS, and STATA.

1.5 Summary

In this chapter, we looked at previous studies that focused on intimate partner violence in different countries. The variable selection was carried out. The independent variables were selected based on variables used in previous studies. In Chapter 2, we will investigate how the selected independent variables are associated with the dependent variable.

Chapter 2

Data Description and Exploratory Analysis

This chapter will describe the DHS data used in this study. It also discusses the methodology used to create indicators of IPV (intimate partner violence) and ATM (access to media).

2.1 Data Source

The different datasets utilized in this study are from the Demographic and Health Survey (DHS) website (DHSProgram, 2016). Both men and women were included in the DHSs from the participating countries. This study was based on data provided by women in South Africa, Tanzania, and Uganda who were asked about their IPV experiences and contraceptive methods. Finally, the survey looked at whether the women knew anything about STIs during the 2015-16 survey. The target population included women in the age range of 15 to 49 years.

2.2 Study Design

South Africa

The current study used the data from the 2016 South African Demographic and Health survey (SADHS). The survey was conducted from the 27th of June 2016 to the 4th of November 2016. The sampling frame used for the SADHS 2016 is the Statistics South Africa Master Sample Frame (MSF), which was created using Census 2011 enumeration areas (EAs). In the MSF, EAs of manageable size were treated as primary sampling units (PSUs), whereas small neighboring EAs were pooled together to form new PSUs, and large EAs were split into conceptual PSUs. The frame contains information about the geographic type (urban, traditional, or farm) and the estimated number of residual dwelling units (DUs) in each PSU. The Sampling convention used by Stats SA (Statistics South Africa) are the DUs. One or more households may be in any given DU; recent surveys have found 1.03 households per DU on average. The SADHS 2016 followed a stratified two-stage sample design with a probability proportional to size sampling of PSUs at the first stage and systematic sampling of DUs at the second stage. The Census 2011 DU count was used as the PSU measure of size. 750 PSUs were selected from 26 sampling strata, yielding 468 selected PSUs in urban areas, 224 PSUs in traditional areas, and 58 PSUs in farm areas. The survey included a module on domestic violence. One woman of age 15 and older was selected. The survey used questionnaires to be answered by the women and men of each household. The SADHS 2016 provided women dataset among others, and we used the dataset in this study. More details on sampling techniques used in the survey and data collection can be found (SA), 2012).

Uganda

We also used the 2016 Uganda Demographic and Health Surveys (UDHS) in this study. The survey was conducted from the 20th of June 2016 to the 16th of December 2016. The sampling frame used for the 2016 UDHS is the Uganda National Population and Housing Census (NPHC) frame, conducted in 2014; the Uganda Bureau of Statistics provided the sampling frame. The census frame is a complete list of all census enumeration areas (EAs) created for the 2014 NPHC. In Uganda, an EA is a geographic area that covers an average of 130 households. The sampling frame contains information about EA location, type of residence (urban or rural), and the estimated number of residential households. At the time of NPHC, Uganda has divided administratively into 112 districts, grouped for this survey into 15 regions. The sample 2016 UDHS provides estimates of key indicators for the country, for urban and rural areas separately, and each of the 15 regions. Estimates presented for three areas: The Lake Victoria islands, the mountain districts, and greater Kampala.

The 2016 UDHS sample was stratified and selected in two stages. In the first stage, 697 EAs were selected from the 2014 Uganda NPHC: 162 EAs in urban areas and 535 in rural areas. One cluster from the Acholi sub-region was eliminated because of land disputes. Households constituted the second stage of sampling. A listing of households was compiled in each of the 696 accessible selected EAs from April to October 2016, some listings overlapping with fieldwork. Maps were drawn for each of the sampled clusters and all the listed households. The listing excluded institutional living arrangements such as army barracks, hospitals, police camps, and boarding schools. Each large EA (i.e., more than 300 households) selected for the 2016 UDHS was segmented to minimize household listing tasks. Only one segment was selected for the survey with probability proportional to segment size, and the household listing was conducted only in the selected segment. Thus, a 2016 UDHS cluster is either an EA or a segment of an EA. In total, a representative sample of 20,880 households (30 per EA or EA segment) was randomly selected for the 2016 UDHS. In addition, a subsample of one eligible woman in two-thirds of households (those households not selected for the male survey and biomarker collection) and one eligible man in one-third of households (those households selected for the male survey and biomarker collection) was randomly selected to be asked questions about domestic violence. The survey used questionnaires to be answered by the women and men of each household. The UDHS 2016 provided women dataset, among others, and was used in this study. More details on sampling techniques used in the survey and data collection can be found (Uganda Bureau of Statistics (UBOS) and ICF, January 2018).

Tanzania

The sample design for the 2015-16 TDHS-MIS was done in two stages and was intended to provide estimates for the entire country, for urban and rural areas in Tanzania Mainland, and for Zanzibar. For specific indicators such as contraceptive use, the sample design allowed the estimation of indicators for each of the 30 regions (25 regions from Tanzania Mainland and 5 regions from Zanzibar). The first stage involved selecting sample points (clusters), consisting of enumeration areas (EAs) delineated for the 2012 Tanzania Population and Housing Census. A total of 608 clusters were selected. In the second stage, a systematic selection of households was involved. A complete households listing was carried out for all 608 selected clusters prior to the fieldwork. From the list, 22 households were then systematically selected from each cluster, yielding a representative probability sample of 13,376 households for the 2015-16 TDHS-MIS. To estimate geographic differentials for certain demographic indicators, Tanzania was divided into nine geographic zones. Although these zones are not official administrative areas, this classification system is also used by the Reproductive and Child Health Section of the MoHCDGEC (MoHCDGEC, et al., 2016). Grouping the regions into zones allowed a relatively large number of people in the denominator and a reduced sampling error (NBS & OCGS, 2019).

2.3 Study Population and Sample Size

The study consisted of a population of women aged 15-49 years. The overall sample size for women in South Africa (n=8514) was selected for the domestic violence module (NDoH, et al., January 2019). From the DHS dataset, among the women who accepted to be interviewed, 85% completed the module (NDoH, et al., January 2019). In Tanzania, only 9322 women were suitable for domestic violence questions (MoHCDGEC, et al., 2016). About 2% of women eligible for the domestic violence module could not be successfully interviewed (MoHCDGEC, et al., 2016). In Uganda, a total of 9232 women aged 15-49 years responded to the domestic violence module; one percent of eligible women could not be successfully interviewed for the module because of lack of privacy or other reasons (NBS & OCGS, 2019; UBOS, 2019).

2.4 The Response Variables

The primary response variable (IPV) was the dichotomous variable representing the status of women who responded to the IPV questions. To identify the determinants of IPV, a response variable was chosen considering the following variables:

- D102 = Number of control issues
- D104 = Experienced any emotional violence
- D106 = Experienced any less severe violence
- D107 = Experienced any severe violence
- D108 = Experienced any sexual violence

The access to media (ATM) was the categorical variable measuring the level of exposure of women and was created using the following variable:

- V157 = Frequency of reading newspaper or magazine
- V158 = Frequency of listening to the radio
- V159 = Frequency of watching television
- V171B = Frequency of using internet last month

The levels for ATM were: no exposure, low exposure, medium exposure, and high exposure.

2.5 Exploratory Data Analysis (EDA)

The primary objective of EDA is to assist in understanding the data in detail before the modeling and inference stages. In this section, we present an intensive exploratory data analysis. The chi-square testing has been carried out to establish the link between the outcome variable and the independence variables. The independence variables with a p-value of less than 0.05 are not left out in the model since they are strongly associated with the outcome variable. Simple explanatory statistics such as the frequency distributions and percentages have been evaluated to explain the variables and check for missing observations.

Table 2.1: Chi-square testing for covariates associated with the response variable

	South Africa			Uganda			Tanzania		
Effect	**DF**	**ChiSq value**	**Pr > ChiSq**	**DF**	**ChiSq value**	**Pr > ChiSq**	**DF**	**ChiSq value**	**Pr > ChiSq**
Partner's alcohol drinking status	2	306.3265	<.0001	2	1080.4797	<.0001	1	908.8998	<.0001
Woman's father ever beat her mother	2	222.2671	<.0001	2	439.5897	<.0001	2	356.7751	<.0001
Access to the media	2	28.9523	<.0001	2	93.4891	<.0001	2	51.0112	<.0001
Woman's current age	1	479.5744	<.0001	34	1587.7564	<.0001	34	2027.6121	<.0001
Region	8	66.9926	<.0001	14	172.6304	<.0001	29	436.4064	<.0001
Type of place of residence	1	2.3462	0.1256	1	68.8119	<.0001	1	29.3719	<.0001
Woman's highest education level	3	3.1443	0.3699	3	211.2430	<.0001	3	526.4669	<.0001
Number of household members	1	351.2498	<.0001	22	534.3775	<.0001	30	1206.0779	<.0001
Sex of household head	1	33.7276	<.0001	1	123.4940	<.0001	1	45.4121	<.0001
Literacy	1	1.7846	0.1816	1	202.8874	<.0001	1	88.3303	<.0001
Wealth index	4	84.3448	<.0001	4	252.9049	<.0001	4	161.8325	<.0001
Ever had a terminated pregnancy	1	88.7850	<.0001	1	226.7753	<.0001	1	197.9810	<.0001
Contraceptive method used	1	78.4322	<.0001	1	176.8169	<.0001	1	474.3873	<.0001
Body Mass Index	1821	1852.0710	0.3005	1569	1350.0118	1.000	2150	2186.7632	0.2851
Current marital status	2	309.6610	<.0001	2	1851.3785	<.0001	2	2197.7765	<.0001
Number of other wives/partners	4	4.3696	0.3583	11	150.5900	<.0001	7	41.9345	<.0001
Cohabitation duration	1	13.3209	0.0003	1	113.0496	<.0001	1	95.0437	<.0001

Continued…

Partner's desire for children	3	45.6622	<.0001	3	22.5892	<.0001	3	59.2628	<.0001
Partner's education level	4	12.6660	0.0130	4	211.8207	<.0001	4	193.8716	<.0001
Partner's occupation	1	5.4330	0.0198	1	1.8016	0.1795	1	0.5954	0.4403
Woman's occupation	2	102.4633	<.0001	1	354.8114	<.0001	1	522.4364	<.0001
Partner's age	58	175.0144	<.0001	69	982.9577	<.0001	68	1217.1745	<.0001
Woman's earnings compared to partner	4	9.8774	0.0425	4	313.7857	<.0001	4	23.7439	<.0001
Knowledge of Sexually Transmitted Infections (STIs)	1	43.2351	<.0001	1	11.9220	0.0006	0	0	0
Wife-beating attitude	2	55.2004	<.0001	2	8.2269	0.0164	2	7.8961	<.0001
The person who usually decides on what to do with the woman's earnings	2	46.2568	<.0001	2	37.3685	<.0001	2	42.7895	<.0001

2.5.1 South Africa

Table 2.2 below shows the variables that have been selected. The dependence of the response variable on each variable has been investigated. The response variable dependence on some predictor variables is determined by the p-value of less than or equal to 0.05, obtained from the chi-square test.

Table 2.2 shows the prevalence of intimate partner violence (IPV) among women of reproductive age in South Africa. The current study considered 8514 women. Table 2.2 shows that the overall prevalence of IPV was 26.93%.

Table 2.2: The prevalence of intimate partner violence among women of reproductive age by category indicator variable

Indicator	**Category**	**EXPERIENCED IPV**		**P-value**
		YES - N (%)	**NO** - N (%)	
IPV		2293(26.93)	6221(73.07)	
Woman's current age	Continuous	Minimum=15		
		Mean=30.21		
		Maximum=49		
Region	Western Cape	141(1.66)	515(6.05)	<.0001
	Eastern Cape	322(3.78)	719(8.44)	
	Northern Cape	136(1.60)	582(6.84)	
	Free State	247(2.90)	607(7.13)	
	Kwazulu-Natal	314(3.69)	1024(12.29)	
	North West	270(2.90)	593(6.96)	
	Gauteng	242(2.84)	621(7.29)	
	Mpumalanga	313(3.68)	741(8.70)	
	Limpopo	308(3.62)	797(9.36)	
Type of place of residence	Rural	1263(14.83)	3542(41.60)	0.1256
	Urban	1030(12.10)	2679(31.47)	
Woman's education level	No education	58(0.68)	132(1.55)	0.3699
	Primary	245(2.88)	617(7.25)	
	Secondary	1745(20.50)	4836(56.80)	
	Higher	245(2.88)	636(7.47)	
Number of household members	Less than 5	1774(20.84)	3676(43.18)	<.0001
	More than or equal to 5	519(6.10)	2545(29.89)	

Continued…

Sex of the household head	Male	1090(12.80)	2521(29.61)	<.0001
	Female	1203(14.13)	3700(43.46)	
Literacy	Cannot read	100(1.17)	232(2.72)	0.1816
	Able to read	2193(25.76)	5989(70.34)	
Wife-beating attitude	Unacceptable	2067(24.28)	5726(67.25)	<.0001
	Acceptable	208(2.44)	332(3.90)	
	I don't know	18(0.21)	163(1.91)	
Access to the media	Low exposure	325(3.83)	981(11.52)	<.0001
	Medium exposure	1384(16.26)	3345(39.39)	
	High exposure	583(6.85)	1886(22.15)	
Wealth index	Poorest	492(5.78)	1271(14.93)	<.0001
	Poorer	573(6.73)	1292(15.18)	
	Middle	587(6.89)	1369(16.08)	
	Richer	435(5.11)	1298(15.25)	
	Richest	206(2.42)	991(11.64)	
Ever had a terminated pregnancy	No	1972(23.16)	5763(67.69)	<.0001
	Yes	321(3.77)	458(5.38)	
Contraceptive method used	No	1028(12.07)	3461(40.65)	<.0001
	Yes	1265(14.86)	2760(32.42)	
Body Mass Index	Underweight	71(0.83)	201(2.36)	0.0806
	Healthy	637(7.48)	1676(19.69)	
	Overweight	559(6.57)	1384(16.26)	
	Obese	1026(12.05)	2960(34.77)	
Current marital status	Single	1205(14.15)	4468(52.48)	<.0001
	Married	637(7.48)	1188(13.95)	
	Living with partner	451(5.30)	565(6.64)	
Number of other wives/partners	No other wives	2109(24.77)	5741(67.43)	0.8179
	One or more	65(0.76)	178(2.09)	
	I don't know	119(1.40)	302(3.550	
Cohabitation duration	0-4	2059(24.18)	5740(67.42)	0.0003
	5-9	234(2.75)	481(5.65)	
Partner's desire for children	Both want same	1077(12.65)	3434(40.33)	<.0001
	Partner wants more	453(5.32)	1041(12.23)	
	Partner wants fewer	129(1.52)	287(3.37)	
	Don't know	634(7.45)	1459(17.14)	
Partner's education level	No education	99(1.16)	279(3.28)	0.0130
	Primary	258(3.03)	588(6.91)	
	Secondary	1559(18.31)	4162(48.88)	
	Higher	364(4.28)	1147(13.47)	
	Don't know	13(0.15)	45(0.53)	
Partner's occupation status	Employed	1931(22.68)	5363(62.99)	0.0198
	Don't know	362(4.25)	858(10.08)	
Woman's occupation status	Unemployed	1261(14.81)	4148(48.72)	<.0001
	Employed	938(11.02)	1840(21.61)	
	Don't know	94(1.10)	233(2.74)	
Partner's age	Less than 25	139(1.63)	756(8.88)	<.0001
	Between 25 and 34	802(9.42)	2372(27.86)	
	35 and above	1352(15.88)	3093(36.33)	
Woman’s earnings compared to partner	More than him	442(5.19)	1151(13.52)	0.0425
	Less than him	1219(14.32)	3441(40.42)	
	About the same	369(4.33)	998(11.72)	
	Partner doesn't bring in	209(2.45)	457(5.37)	
	Don't know	54(0.63)	174(2.04)	

Continued…

Knowledge of Sexually Transmitted Infections (STIs)	No	25(0.29)	242(2.84)	<.0001
	Yes	2268(26.64)	5979(70.23)	
The person who usually decides on what to do with the woman's earnings	Woman alone	723(8.49)	1709(20.07)	0.0011
	Woman and partner	1390(16.33)	4013(47.13)	
	Partner alone	180(2.11)	499(5.86)	
Woman's father ever beat her mother	No	1699(19.96)	5432(63.80)	<.0001
	Yes	449(5.27)	551(6.47)	
	Don't know	145(1.70)	238(2.80)	

Table 2.2 shows that the province with the highest prevalence of IPV was the Eastern Cape with 3.78%, followed by Kwazulu-Natal, Mpumalanga, Limpopo, Free State, North West, Gauteng, Western Cape, and Northern Cape, with 3.69%, 3.68%, 3.62%, 2.90%, 2.90%, 2.84%, 1.66%, and 1.60% respectively. The results show that women from households where the head of the household is a woman, have a higher prevalence of IPV with 14.13%, while having a 12.80% prevalence for households with a man being the head of the household (p-value<.0001). The table above shows that women with medium, high, and low exposure to media showed 16.26%, 6.85%, 3.83% prevalence of IPV, respectively, with a p-value<.0001. This table also shows that women from the poorest, poorer, middle, richer, and richest wealth indexes have a 5.78%, 6.73%, 6.89%, 5.11%, and 2.42% prevalence of IPV, respectively, with a p-value<.0001. Women from a household with less than five family members and those with five or more family members showed a 20.84% and 6.10% prevalence, respectively, with a p-value<.0001. The table also shows that 24.28%, 2.44%, and 0.21% of women view wife-beating attitudes as unacceptable, acceptable, and unknown, respectively, with a p-value<.0001. Moreover, it reveals that the prevalence of IPV in women who terminated pregnancy is 3.77%, and 23.16% for those who have never done so, with a p-value<.0001. The results found that 14.86% of women are using contraceptives while 12.07% are not using any contraceptive method, with a p-value<.0001. The above table further shows that single, married, and those living with a partner have a 14.15%, 7.48%, and 5.30% prevalence of IPV, respectively, with a p-value<.0001. Women who stayed with a partner for 0-4 years have a 24.18% prevalence of IPV, while those who stayed with a partner for 5-9 years have a 2.75% prevalence, respectively, with a p-value=0.0003.

Table 2.2 reveals that women who have a partner who wants the same number of children, wants more, wants fewer, and those who do not know their partner's desire for children, showed a 12.65%, 5.32%, 1.52%, and 7.45% prevalence of IPV, respectively, with a p-value<.0001. A woman who had a partner with no education, primary, secondary, higher education, and those who do not know their partner's level of education, showed a 1.16%, 3.03%, 18.31%, 4.28%, and 0.15% prevalence of IPV, respectively, with a p-value=0.0130. Women who have an employed partner showed a 22.68% prevalence of IPV, and those whose partner is unemployed, 4.25%, with a p-value=0.0198. Employed women, not employed, and those who do not know, showed 14.81%, 11.02%, and 1.10% prevalence of IPV, respectively, with a p-value<.0001. We also investigated the knowledge of Sexually Transmitted Infections (STI's), and Table 2.2 shows that women who knew about STI's showed a 26.64% prevalence of IPV. Those who do not know about STIs had a prevalence of 0.29%, with a p-value<.0001. There was an IPV prevalence of 1.63%, 9.42%, and 15.88%, for women with a partner younger than 25, between 25 and 34, and 35 and above, respectively, with a p-value<.0001. The same table also revealed an 8.49%, 16.33%, and 2.11% prevalence of IPV for women who decide what to do with their earnings alone, for both the woman and her partner deciding, and for the partner alone deciding, respectively, with a p-value=0.0011. The table shows that women with a partner who drinks alcohol showed a 14.34% prevalence of IPV, those with partner who does not drink had a prevalence of 12.50%, and a prevalence of 0.09% for those who do not know if a partner drinks alcohol, with a p-value<.0001. In Table 2.2, women who have never witnessed their father beat their mother showed a 19.96% prevalence of IPV, with 5.27% for those who have witnessed their father beat their mother, and 1.70% for those who do not know if the father ever beat her mother, with a p-value<.0001.

2.5.2 Uganda

Table 2.3 shows the prevalence of IPV amongst women of reproductive age in Uganda. The study approached 18 506 women. In this table, the overall prevalence of IPV was 32.25%.

Table 2.3: The prevalence of intimate partner violence for women of reproductive age, by category of the indicator variable

Indicator	Category	EXPERIENCED IPV		P-value
		YES - N (%)	**NO** - N (%)	
IPV		5968(32.25)	12538(67.75)	
Woman's current age	Continuous	Minimum=15		
		Mean=27.94		
		Maximum=49		
Region	Kampala	290(1.57)	1010(5.46)	<.0001
	South Buganda	471(2.55)	1144(6.18)	
	North Buganda	459(2.48)	951(5.14)	
	Busoga	538(2.91)	992(5.36)	
	Bukedi	441(2.38)	764(4.13)	
	Busigu	316(1.71)	641(3.46)	
	Teso	412(2.23)	935(5.05)	
	Karamoja	275(1.49)	466(2.52)	
	Lango	402(2.17)	834(4.51)	
	Acholi	346(1.87)	764(4.13)	
	West Nile	466(2.52)	815(4.40)	
	Bunyoro	300(1.62)	913(4.93)	
	Tooro	419(2.26)	882(4.77)	
	Ankole	524(2.83)	777(4.20)	
	Kigezi	309(1.67)	650(3.51)	

Continued…

Type of place of residence	Rural	1188(6.42)	3191(17.24)	<.0001
	Urban	4780(25.83	9347(50.51)	
Woman's education level	No education	836(4.52)	1235(6.67)	<.0001
	Primary	3733(20.17	7160(38.69)	
	Secondary	1093(5.91)	3120(16.86)	
	Higher	306(1.65)	1023(5.53)	
Number of household members	Less than 5	3513(18.98)	5577(30.14)	<.0001
	More than or equal to 5	2455(13.27)	6961(37.61)	
Sex of household head	Male	4316(23.32)	8035(43.42)	<.0001
	Female	1652(8.93)	4503(24.33)	
Literacy	Cannot read	2513(13.58)	3941(21.30)	<.0001
	Able to read	3455(18.67)	8597(46.46)	
Wife-beating attitude	Unacceptable	2717(14.68)	6481(35.02)	<.0001
	Acceptable	3198(17.28)	5770(31.18)	
	Don't know	53(0.29)	287(1.55)	
Access to the media	Low exposure	2293(12.39)	4483(24.22)	<.0001
	Medium exposure	3554(19.20	7436(40.18)	
	High exposure	121(0.65)	619(3.34)	
Wealth index	Poorest	1504(8.13)	2380(12.86)	<.0001
	Poorer	1319(7.13)	2321(12.54)	
	Middle	1152(6.23)	2333(12.61)	
	Richer	1047(5.66)	2407(13.01)	
	Richest	946(5.11)	3097(16.74)	
Ever had a terminated pregnancy	No	4527(24.46)	10651(57.55)	<.0001
	Yes	1441(7.79)	1887(10.20)	
Contraceptive method used	No	3836(20.73)	9252(49.99)	<.0001
	Yes	2132(11.52)	3286(17.76)	
Body Mass Index	Underweight	536(2.90)	1085(5.86)	0.0588
	Healthy	4103(22.17)	8435(45.58)	
	Overweight	899(4.86)	2029(10.96)	
	Obese	430(2.32)	989(5.34)	
Current marital status	Single	967(5.23)	6160(33.29)	<.0001
	Married	2548(13.77)	3265(17.64)	
	Living with partner	2453(13.26)	3113(16.82)	
Number of other wives/partners	No other wives/partners	3957(21.38)	9237(49.91)	<.0001
	One other wife/partner	1722(9.31)	2883(15.58)	
	I don't know	289(1.57)	419(2.27)	
Cohabitation duration	0-4	4657(25.16)	10583(57.19)	<.0001
	5-9	1311(7.08)	1955(10.56)	
Partner's desire for children	Both want same	2092(11.30)	4672(25.25)	<.0001
	Partner wants more	2109(11.40)	3996(21.59)	
	Partner wants fewer	579(3.13)	1239(6.68)	
	Don't know	1188(6.42)	2633(14.23)	

Continued…

Partner's education level	No education	516(2.79)	748(4.04)	<.0001
	Primary	3334(18.02)	6031(32.59)	
	Secondary	1456(7.87)	3652(19.73)	
	Higher	542(2.93)	1708(9.23)	
	Don't know	120(0.65)	399(2.16)	
Partner's occupation status	Unemployed	216(1.17)	506(2.73)	0.1795
	Employed	5752(31.08)	12033(65.02)	
Woman's occupation status	Unemployed	822(4.44)	3268(17.66)	<.0001
	Employed	5146(27.81)	9270(50.09)	
Partner's age	Less than 25	594(3.21)	3410(18.43)	<.0001
	Between 25 and 34	2102(11.36)	4283(23.14)	
	35 and above	3272(17.68)	4845(26.18)	
Woman's earnings compared to partner	More than him	568(3.07)	780(4.21)	<.0001
	Less than him	4271(23.08)	7950(42.96)	
	About the same	855(4.62)	2695(14.56)	
	Partner doesn't bring in	101(0.55)	317(1.71)	
	Don't know	173(0.93)	796(4.30)	
Knowledge of Sexually Transmitted Infections (STIs)	No	6(0.03)	50(0.27)	0.0006
	Yes	5962(32.22)	12488(67.48)	
The person who usually decide on what to do with the woman's earnings	Woman alone	3439(18.58)	6163(33.30)	<.0001
	Woman and partner	2093(11.31)	5138(27.76)	
	Partner alone	436(2.36)	1237(6.69)	
Partner's drinks alcohol	No	3023(16.34)	9396(50.77)	<.0001
	Yes	2945(15.91)	3142(16.98)	
Woman's father ever beat her mother	No	3253(17.58	8799(47.55)	<.0001
	Yes	2450(13.24	3332(18.00)	
	Don't know	265(1.43)	407(2.20)	

Table 2.3 shows that the region with the highest prevalence of IPV is Busonga at 2.91%, followed by Ankole, South Buganda, West Nile, North Buganda, Bukedi, Tooro, Teso, Lango, Acholi, Busigu, Kigezi, Bunyoro, Kampala, and Karamoja regions, at 2.83%, 2.55%, 2.52%, 2.48%, 2.38%, 2.26%, 2.23%, 2.17%, 1.87%, 1.71%, 1.67%, 1.62%, 1.57%, and 1.49%, respectively. The highest IPV is prevalent in urban parts of the country, with 25.83%, compared to only 6.42% for rural areas, with a p-value<.0001. About 20.17% of the women who have experienced IPV have primary education, 5.91 % secondary education, 4.52% no education, 1.65% higher education, with a p-value<.0001. Women from a household with less than five family members and those with five or more family members, showed an 18.98% and 13.27% IPV prevalence, respectively, with a p-value<.0001. The table also shows that 14.68%, 17.28%, and 0.29% of women view wife-beating attitudes as unacceptable, acceptable, and unknown, respectively, with a p-value<.0001.

Table 2.3 also shows that a household where the head of the household is a man has a 23.32% IPV prevalence for women, but only an 8.93% for those with a woman as the head, with a p-value<.0001. We can also see that women who can read have an 18.67% prevalence of IPV, with an 13.58% one for those who cannot read, with a p-value<.0001. Women with medium exposure to the media showed a high prevalence of IPV at 19.20%, followed by those with low and high exposure at 12.39% and 0.65%, respectively, with a p-value<.0001. Moreover, the results show that women from the poorest, poorer, middle, richer, and richest wealth indexes have an 8.13%, 7.13%, 6.23%, 5.66%, and 5.11% prevalence of IPV, respectively, with a p-value<.0001. The table above reveals the prevalence of IPV in women who terminated pregnancy is 7.79%, while being 24.46% for those who have never done so, with a p-value<.0001. Table 2.3 shows that women who use contraceptive methods, have an IPV prevalence of 11.52%, while those that are not using any have a prevalence of 20.73%, with a p-value<.0001. The table mentioned above shows that single, married, and those women living with a partner have a 5.23%, 13.77%, and 13.26% prevalence of IPV, respectively, with a p-value<.0001. Women who have been staying with a partner for 0-4 years have a 25.16% prevalence of IPV, while those who stayed with a partner for 5-9 years had a 7.08% one, respectively, with a p-value<.0001. Meanwhile, there was an IPV prevalence of 21.38%, 9.31%, and 1.57%, for women whose partner has no other wives/partners, one other wife/partner, and those who do not know whether their partner has other wives/partners, respectively, with a p-value<.0001.

Moreover, Table 2.3 amongst women who have a partner who wants the same number of children, one who wants more, one who wants fewer, and those who do not know their partner's desire for children, there was an IPV prevalence of 11.30%, 11.40%, 3.13%, and 6.42%, respectively, with a p-value<.0001. A woman who had a partner with no education, primary education, secondary education, higher education, and those who do not know their partner's level of education, showed an IPV prevalence of 2.79%, 18.02%, %, 7.87%, 2.93%, and 0.65%, respectively, with a p-value<.0001. Employed and unemployed women showed 27.81% and 4.44% prevalence of IPV, respectively, with a p-value<.0001. Women who earn more than, less than, about the same as their partner, with a partner who does not bring anything, and those who do not know what their partner earns, showed a 3.07%, 23.085, 4.62%, 0.55%, and 0.93% prevalence of IPV, respectively, with a p-value<.0001. There is an IPV prevalence of 3.21%, 11.36%, and 17.68% for women with a partner aged less than 25, aged between 25 and 34, and those aged 35 and above, respectively, with a p-value<.0001. The table above also revealed 18.58%, 11.31%, and 2.36% prevalence of IPV for women who decide on what to do with their earnings alone, for cases where both the woman and her partner decide, and where the partner alone decides, respectively, with a p-value<.0001. With regards to knowledge of sexually transmitted infections (STIs), the table shows that women who knew about STIs showed a 32.22% prevalence, and those who do not know about them showed a prevalence of 0.03%, with a p-value=0.0006. Table 2.3 further shows that women with a partner who drinks alcohol had a 15.91% prevalence of IPV, while those with a partner who does not drink had a prevalence of 16.34%, with a p-value<.0001. The results also revealed that a woman who has never witnessed her father beating her mother showed a 17.58% prevalence of IPV, however, there was a 13.24% prevalence for those who have witnessed their father beat their mother, with a p-value<.0001.

2.5.3 Tanzania

Table 2.4 shows the prevalence of IPV amongst females of reproductive age in Tanzania. The current study considered 13 266 females. Table 2.4 shows that the overall prevalence of IPV was 44.70%.

Table 2.4: The prevalence of intimate partner violence amongst women of reproductive age, by category of the indicator variable

Indicator	**Category**	**EXPERIENCED IPV**		**P-value**
		YES - **N** (%)	NO - **N** (%)	
IPV		5930(44.70)	7336(55.30)	
Woman's age	Continuous	Minimum=15		
		Mean=28.69		
		Maximum=49		
Region	Dodoma	176(1.33)	167(1.26)	<.0001
	Arusha	216(1.63)	204(1.54)	
	Kilimanjaro	141(1.06)	229(1.73)	
	Tanga	157(1.18)	308(2.32)	
	Morogoro	177(1.33)	168(1.27)	
	Pwari	169(1.27)	164(1.24)	
	Dar Es Salaam	342(2.58)	455(3.43)	
	Lindi	217(1.64)	163(1.23)	
	Mtwara	202(1.52)	146(1.10)	
	Ruvuma	194(1.46)	189(1.42)	
	Iringa	158(1.19)	182(1.37)	
	Mbeya	182(1.37)	192(1.45)	
	Singida	193(1.45)	220(1.66)	
	Tabora	243(1.83)	317(2.39)	
	Rukwa	237(1.79)	188(1.42)	
	Kigoma	209(1.58)	282(2.13)	
	Shinyanga	246(1.85)	270(2.04)	
	Kagera	247(1.86)	169(1.27)	
	Mwanza	196(1.48)	300(2.26)	
	Mara	272(2.05)	259(1.95)	
	Manyanga	239(1.80)	195(1.47)	
	Njobe	175(1.32)	184(1.39)	
	Katavi	211(1.59)	255(1.92)	
	Simiyu	227(1.71)	360(2.71)	

Continued…

	Geita	235(1.77)	300(2.26)	
	Kaskazini Unguja	153(1.15)	213(1.61)	
	Kusini Unguja	153(1.15)	208(1.57)	
	Mjini Magharibi	212(1.60)	496(3.74)	
	Kaskazini Pemba	68(0.51)	270(2.04)	
	Kusini Pemba	83(0.63)	283(2.13)	
Type of place of residence	Rural	1709(12.88)	2436(18.36)	<.0001
	Urban	4221(31.82)	4900(36.94)	
Woman's highest educational level	No education	1051(7.92)	947(7.14)	<.0001
	Primary	3841(28.95)	3799(28.64)	
	Secondary	998(7.52)	2489(18.76)	
	Higher	40(0.40)	101(0.76)	
Number of household members	Less than 5	3336(25.15)	2257(17.01)	<.0001
	More than or equal	2594(19.55)	5079(38.29)	
Sex of household head	Male	4735(35.69)	5495(41.42)	<.0001
	Female	1195(9.01)	1841(13.88)	
Literacy	Cannot read	1613(12.16)	1486(11.20)	<.0001
	Able to read	4317(32.54)	5850(44.10)	
Access to the media	Low exposure	2867(21.61)	3269(24.64)	<.0001
	Medium exposure	2929(22.08)	3755(28.31)	
	High exposure	134(1.01)	312(2.35)	
Wealth index	Poorest	1071(8.07)	1073(8.09)	<.0001
	Poorer	1078(8.13)	1088(8.20)	
	Middle	1180(8.89)	1258(9.48)	
	Richer	1369(10.32)	1739(13.11)	
	Richest	1232(9.29)	2178(16.42)	
Ever had a terminated pregnancy	No	4671(35.21)	6442(48.56)	<.0001
	Yes	1259(9.49)	893(6.73)	
Contraceptive method used	Not using	3640(27.44)	5770(43.49)	<.0001
	Using	2290(17.26)	1566(11.80)	
Body Mass Index	Underweight	388(2.92)	816(6.15)	<.0001
	Healthy	3632(27.38)	4570(34.45)	
	Overweight	1237(9.32)	1236(9.32)	
	Obese	673(5.07)	714(5.38)	

Continued…

Current marital status	Single	969(7.30)	4108(30.97)	<.0001
	Married	3643(27.46)	2494(18.80)	
	Living with partner	1318(9.94)	734(5.53)	
Number of other wives/partners	No other wives/partners	4549(34.29)	5877(44.30)	<.0001
	one other wife/partner	1312(9.89)	1415(10.67)	
	Don't know	69(0.52)	44(0.33)	
Cohabitation duration	0-4	4598(34.66)	6176(46.56)	<.0001
	5-9	1332(10.04)	1160(8.74)	
Partner's desire for children	Both want same	2138(16.12)	3052(23.01)	0.0026
	Partner wants more	1577(11.89)	1961(14.78)	
	Partner wants fewer	372(2.80)	351(2.65)	
	Don't know	1843(13.89)	1972(14.87)	
Partner's education level	No education	683(5.15)	655(4.94)	<.0001
	Primary	3941(29.71)	4278(32.25)	
	Secondary	1138(8.58)	2070(15.60)	
	Higher	145(1.09)	299(2.25)	
	Don't know	23(0.17)	34(0.26)	
Partner's occupation status	Unemployed	62(0.47)	67(0.51)	0.4403
	Employed	5868(44.23)	7269(54.79)	
Woman's occupation status	Unemployed	872(6.57)	2332(17.58)	<.0001
	Employed	5058(38.13)	5004(37.72)	
Partner's age	Less than 25	412(3.11)	1908(14.38)	<.0001
	Between 25 and 34	2034(15.33)	2562(19.31)	
	35 and above	3484(26.26)	2866(21.60)	
Woman's earnings compared to partner's	More than him	418(3.15)	633(4.77)	0.0034
	Less than him	3920(29.55)	4957(37.37)	
	About the same	1309(9.87)	1436(10.82)	
	Husband/partner doesn't bring in	51(0.38)	61(0.46)	
	Don't know	232(1.75)	249(1.88)	

Continued…

The person who usually decides how to spend the woman's earnings	Woman alone	2183(16.46)	2464(18.57)	<.0001
	Both woman and partner	3193(24.07)	4047(30.51)	
	Partner alone	554(4.18)	825(6.22)	
Partner drinks alcohol	No	3854(29.05)	6388(48.15)	<.0001
	Yes	2076(15.65)	948(7.15)	
Woman's father ever beat her mother	No	3193(24.07)	5119(38.59)	<.0001
	Yes	2203(16.61)	1760(13.27)	
	Don't know	534(4.03)	457(3.44)	
Wife-beating attitude	Unacceptable	2160(16.28)	3330(25.10)	<.0001
	Acceptable	3626(27.33)	3489(26.30)	
	Don't know	144(1.09)	517(3.90)	

Table 2.4 shows that the region with the highest prevalence of IPV with 2.58% is Dar Es Salaam, followed by Mara, Kagera, Shinyanga, Tabora, Manyanga, Rukwa, Geita, Simiyu, Lindi, Arusha, Mjini Magharibi, Katavi, Kigoma, Mtwara, Mwanza, Ruvuma, Singida, Mbeya, Dodoma, Morogoro, Njobe, Pwari, Iringa, Tanga, Kaskazini Unguja, Kushini Unguja, Kilimanjaro, Kusini Pemba and Kaskazini Pemba, with levels of prevalence of 2.05%, 1.86%, 1.85%, 1.83%, 1.80%, 1.79%, 1.77%, 1.71%, 1.64%, 1.63%, 1.60%, 1.59%, 1.58%, 1.52%, 1.48%, 1.46%, 1.45%, 1.37%, 1.33%, 1.32%, 1.27%, 1.19%, 1.18%, 1.15%, 1.15%, 1.06%, 0.63%, and 0.51%, respectively, with a p-value<.0001.

The table above shows that the highest IPV levels are prevalent in urban parts of the country, with 31.82%, while being 12.88% for rural areas, with a p-value<.0001. The results also show that 28.95% of the women who have experienced IPV have primary education, followed by no education, secondary education, and higher education, with 7.92%, 7.52%, and 0.40%, respectively, for a p-value<.0001. Women from a household with less than five family members and those from one with five or more family members, showed 25.15% and 19.55% prevalence of IPV, respectively, p-value<.0001. In Table 2.4, we see that women from households where the head of the household is a man are more prone to IPV with 35.69%, while being only 9.01% for households with a woman as the head of the household, with a p-value<.0001. The results also revealed that women who cannot read have a 12.16% prevalence of IPV, with an almost three times higher level of 32.54% for those who can read, with a p-value<.0001. The table further shows that women with medium exposure to media showed a high prevalence of IPV of 22.08%, followed by low exposure, and high exposure, with 21.61% and 1.01%, respectively, with a p-value<.0001. The table above also shows that women from a richer class of the wealth index showed a high IPV prevalence of 10.32%, followed by the richest, middle, poorer, and poorest, with 9.29%, 8.89%, 8.13%, and 8.07%, respectively, with a p-value<.0001. The prevalence of IPV in women who have terminated pregnancy is 9.49%, while for those who have never done so, it is 35.21%, with a p-value<.0001.

Women's use of contraceptives was also investigated. It was found that there is a 17.26% prevalence of IPV for women who are using contraceptives and a far higher 27.44% for women who are not using them, with a p-value<.0001. The table shows an IPV prevalence of 2.92%, 27.38%, 9.32%, and 5.07%, for women who are underweight, healthy, overweight, and obese, respectively, with a p-value<.0001. The results also show that single, married, and women living with a partner, have a 7.30%, 27.46%, and 9.94% prevalence of IPV, respectively, with a p-value<.0001. In the table above, an IPV prevalence of 34.29%, 9.89%, and 0.52% was found, for women whose partner has no other wives/partners, those who have one other wife/partner, and those who do not know whether their partner has other wives/partners, respectively, p-value<.0001. Women who stayed with a partner for 0-4 years have an IPV prevalence of 34.66%, and those who stayed with a partner for 5-9 years had a 10.04% prevalence of IPV, respectively, with a p-value<.0001. In the table, we see that women who have a partner who wants the same number of children, those who have one who wants more, those who have one who wants fewer, and those who do not know their partner's desire for children, showed a 16.12%, 11.89%, 2.80%, and 13.89% prevalence of IPV, respectively, with a p-value=0.0026. Women who had a partner with no education, a primary one, a secondary one, higher education, and those who do not know their partner's level of education, showed an IPV prevalence of 5.15%, 29.71%, 8.58%, 2.80%, and 13.89%, respectively, with a p-value<.0001. Meanwhile, there was an IPV prevalence of 3.11%, 15.33%, and 26.26%, for women with a partner less than 25 years, one between 25 and 34 years, and those with a partner aged 35 years and above, respectively, p-value<.0001. Employed women and those not employed showed an IPV prevalence of 38.13% and 6.57%, respectively, with a p-value<.0001. Women who earn more than their partner, less than them, about the same as them, whose partner does not bring earnings, and those who do not know what their partner earns, showed a 3.15%, 29.55%, 9.87%, 0.38%, and 1.75% prevalence of IPV, respectively, with a p-value=0.0034.

Table 2.4 shows that women with a partner who drinks alcohol had a 15.65% prevalence of IPV, while those with a partner who does not drink had a prevalence of 29.05%, with a p-value<.0001. The results have also revealed that a woman who has never witnessed her father beating her mother showed a 24.07% prevalence of IPV, with a 16.61% one for women who have witnessed their father beating their mother, and 4.03% for women who do not know if their father beats their mother, with a p-value<.0001. The table also shows levels of IPV prevalence of 16.28%, 27.33%, and 1.09% for women who view wife-beating attitudes as unacceptable, acceptable, and unknown, respectively, with a p-value<.0001.

2.6 Prevalence of IPV

The comparison of IPV for the different countries showed that Tanzania had 44.70% prevalence, followed by Uganda with a prevalence of 32.25%, and finally South Africa with the lowest prevalence of 26.93%.

2.7 Missing Values

The issue of non-response to one or more questions in any survey may be problematic when the data is used in regression analysis (Haitovsky, 1968). Practical and statistically advanced ways have been formulated to manage the missing data problems, depending on their nature and magnitude (Haitovsky, 1968). Three general cases are recognizable:

1. Randomly missing values.
2. Missing classes, meaning there is no response available because the question refers to some non-existing class in the responding unit.
3. Non-randomly missing values: this is a case in which the researcher has reason to believe that neither one of the above is correct, and that some other reason exists for non-response to a specific question (Haitovsky, 1968).

Missingness Mechanisms

There are three different classes of missing data (Rubin, 1976; Ibrahim, et al., 2011). These are: missing completely at random (MCAR), not missing at random (NMAR), and missing at random (MAR) (Rubin, 1976; Ibrahim, et al., 2011). In the following, let us suppose there is subject t, with y_t as the univariate result of interest, and let $\boldsymbol{x}_t$ be the $p \times 1$ vector of predictors comparable to y_t (Rubin, 1976; Ibrahim, et al., 2011).

MCAR

'Missing completely at random' refers to data where the missingness mechanism is not dependent on the variable of interest, or any variable present in the data (Scheffer, 2002). For instance, in logistic regression, say that y_t has complete observations, where some elements of $\boldsymbol{x}_t$ are missing for subject t (Rubin, 1976; Ibrahim, et al., 2011). Therefore, missing values for $\boldsymbol{x}_t$ are said to be MCAR if the probability of observed $\boldsymbol{x}_t$ does not depend on y_t, and is independent of observations of $\boldsymbol{x}_t$, that are observed or would have been observed (Ibrahim, et al., 2011). In other words, the observed data is just a random sample of all the data points (Ibrahim, et al., 2011).

MAR

'Missing at random' refers to data where the missing values are missing at random (Scheffer, 2002). Data is claimed to be MAR only if it is conditional on the recorded data, and failure to observe a value is not dependent on the unobserved data (Ibrahim, et al., 2011). The conditional probability of missingness may solely depend on observed data (Ibrahim, et al., 2011), while the unconditional probability of missing observations may depend on unobserved data (Ibrahim, et al., 2011). Taking for instance, as before, y_t having complete observations, while some elements of $\boldsymbol{x}_t$ are missing (Ibrahim, et al., 2011); observations of $\boldsymbol{x}_t$ are then said to be MAR only if, conditional on the observations, the probability of $\boldsymbol{x}_t$ being observed is independent of the values of $\boldsymbol{x}_t$ that would have been recorded, but this probability is not entirely independent of y_t and the recorded values of $\boldsymbol{x}_t$ (Rubin, 1976; Ibrahim, et al., 2011). Therefore, the unconditional probability of $\boldsymbol{x}_t$ being observed is dependent on $\boldsymbol{x}_t$ (Rubin, 1976; Scheffer, 2002; Ibrahim, et al., 2011). MAR is the most realistic assumption compared to MCAR, but adjustments must be made since the observed predictors are no longer a random sample (Rubin, 1976; Ibrahim, et al., 2011).

NMAR

'Not missing at random', also known as informatively missing, occurs when the missingness mechanism depends on the exact value of the missing data (Schafer & Graham, 2002). This is known as a complex condition to model for (Schafer & Graham, 2002).

Nonignorable Missing Data

'Nonignorable missing data' occurs if the failure to record an entry depends entirely on the observation that would have been recorded (Rubin, 1976; Ibrahim, et al., 2011). For instance, suppose that some elements of $\boldsymbol{x}_t$ are missing, and y_t is completely observed (Rubin, 1976; Ibrahim, et al., 2011); missing observations of $\boldsymbol{x}_t$ are then said to be nonignorable, only if, depending on the recorded data, the probability that $\boldsymbol{x}_t$ is missing, depends on the missing observations of $\boldsymbol{x}_t$ (Ibrahim, et al., 2011).

Overview of Approaches to Imputation Methods

Multivariate Normal Imputation (MVNI)

'Multivariate normal imputation' presumes that the variables in the imputation model jointly follows a multivariate normal distribution (Lee & Carlin, 2010; Rubin, 1987). The implementation utilized a Bayesian approach to get the imputed observations from the approximate multivariate normal distribution, qualifying for uncertainty in the predicted model parameters, for 'proper' imputation (Rubin, 1987; Lee & Carlin, 2010). The presumption of multivariate normality is frequently not plausible, more specifically in the presence of categorical or binary variables (Lee & Carlin, 2010). Schafer (1997) proposed that as MVNI may at times be rational, even if multivariate normality is not valid, nonetheless, it has been put in contexts where the data does not qualify to be multivariate normal (Choi, et al., 2008; Seitzman, et al., 2008; Lee & Carlin, 2010; Schafer, 1997).

Fully Conditional Specification (FCS)

The fully conditional specification method is the most adjustable method that does not depend on the presumption of multivariate normality (van Buuren, et al., 1999; Raghunathan, et al., 2001; Lee & Carlin, 2010). From there, conditional distributions (regression models) are defined for each predictor with missing observations, conditional on other predictors in the imputation model (Lee & Carlin, 2010). Imputations are created by predicting each conditional distribution, utilizing recorded values for the predictor in consideration, and imputed observations for all other predictors, at that particular iteration and imputing missing observations (Lee & Carlin, 2010). This approach is very useful, since it does not limit the conditional distributions to being normal, so that univariate regression models can be appropriately tailored, for instance, in the use of logistic regression for binary predictors and ordered logistic regression for ordinal predictors (Lee & Carlin, 2010).

Nonnormal Continuous Variables and Prediction Matching

In the approaches above, continous predictors with clearly skewed distributions are less likely to be adequately managed without any exceptional treatment. This is because multivariate normality suggests a normal marginal distribution for each predictor, and standard FCS draws imputed observations for a continous predictor, utilizing a normal linear regression on other predictors specified (Lee & Carlin, 2010). The substitute for FCS is to impute observations utilizing prediction matching (Royston, 2004; Little & Rubin, 2002; Lee & Carlin, 2010), where the missing observation is substituted by a nonmissing observation for the case whose predictive mean is closest to that of the case with the missing observation (Lee & Carlin, 2010). Within multivariate normal imputation, a fundamental approach to skewness is to use normalizing transformations; these can be applicable when utilizing FCS (Lee & Carlin, 2010).

Methods

Multivariate imputation by chained equations, also known as 'sequential regression multiple imputations', has appeared in literature as a proper method of dealing with missing observations (Azur, et al., 2011). Generating multiple imputations, compared to generating single imputations, takes into account the statistical uncertainty in the imputations (Azur, et al., 2011). The chained equations method is adjustable and can deal with predictors of different types and complexities, such as bounds or survey patterns (Azur, et al., 2011).

When there is less than 5% missingness in a specific circumstance and the observations are missing at random, and do not rely on observed or unobserved data points, complete case may be an acceptable approach for dealing with missing observations (Azur, et al., 2011; Graham, 2009). Practically, these circumstances hardly occur (Graham, 2009; Azur, et al., 2011). Complete case may be effortless to carry out; it depends on a stronger presumption of missing observations presumption compared to multiple imputations, resulting in biased estimates and reduction of power (Graham, 2009; Azur, et al., 2011). The single imputation method, such as mean imputation, is an advancement but does not take into account uncertainty in the imputations (Azur, et al., 2011). Sometimes, the maximum likelihood approach is viable in dealing with missing observations (Azur, et al., 2011; Graham, 2009); nevertheless, this approach is only available for certain types of models, such as longitudinal or structural equations models (Azur, et al., 2011).

Multiple imputations have a few advantages compared to the other missing data approaches (Azur, et al., 2011). They involve filling in the missing observations multiple times, generating multiple 'complete' datasets. As explained by Schafer and Graham (2002), missing observations are imputed based on the recorded values for different individuals, and the relations recorded in the data for other participants, presuming the observed predictors are incorporated in the imputation model (Schafer & Graham, 2002; Azur, et al., 2011). Multiple imputation approaches, particularly MICE, are adjustable and can be utilized in various settings (Azur, et al., 2011; Graham, 2009). They involve generating multiple predictions for each missing observation, the analysis of multiple imputed data, and consider the uncertainty in the imputations and results in accurate standard errors (Azur, et al., 2011).

The MICE Approach

The MICE method is well presented in a study by van Buuren et al., 1999. MICE works under the assumption that the predictors utilized in the imputation approach, the missing observations, are missing at random (MAR), which implies that the probability that an observation is missing, relies only on observed and not on unobserved data points (Azur, et al., 2011; Schafer & Graham, 2002). After accounting for all the variables in the available data, any missingness is entirely at random (Graham, 2009; Azur, et al., 2011). If MICE is implemented when data is not MAR and could yield biased estimates (Azur, et al., 2011; Schafer & Graham, 2002).

These multiple imputed data sets are analyzed using standard approaches for complete data, and by combining the results from these analyses (Yang, 2011). No matter which complete data analysis is utilized, the process of integrating the results of parameter estimates, and their related standard errors from different imputed data sets, is essentially the same (Azur, et al., 2011). This process results in valid statistical inferences that reflect the uncertainty due to missing values (Yang, 2011). *Proc mi* in SAS Enterprise was utilized for missing values in this study. In this study we assumed that the data was missing completely at random.

MICE Steps

MICE is a specific imputation method (Raghunathan, et al., 2001; Van Buuren, 2007; Azur, et al., 2011). The chained equation process follows the four following steps:

1. A basic imputation, such as imputing the average/mean, is carried out for all missing observations in the data set. These average/mean imputations can be conceptualized as 'place holders' (Azur, et al., 2011).
2. The 'place holder' (average/mean) imputations for one predictor 'var', are set back to missing (Azur, et al., 2011).
3. The values observed from the predictor 'var' in step 2, are regressed on the other variables in the imputation model, which may or may not constitute all the variables in the dataset. In simple terms, 'var' is the response variable in a regression model, and all the others are independent variables in the model (Azur, et al., 2011).
4. Missing values for 'var' are substituted with predictions (imputations) from the regression model. 'Var' is later utilized as an independent variable in the regression models for other variables, so that the observed and imputed values can be utilized (Azur, et al., 2011).

Missing data challenges

In this study, the challenge that has been encountered with regards to missing values is that some participants did not respond to the variables that were vital in creating the response variable. In the current study, the implication of the abovementioned was that the response variable had a few responses from participants and therefore made it difficult to model the data and get reliable inference. The missing values in each country was about 5-10%.

Technique used

Imputation techniques have been discussed in this section, therefore in this study the multiple imputation by chained equations (MICE) has been used in addressing the issue of missing values. The values were assumed to be missing at random (MAR), and therefore the steps for MICE were carried out in this study.

2.8 Summary

This section found that the IPV prevalence for South Africa, Uganda, and Tanzania is 26.93%, 32.25%, and 44.70%, respectively. The results also revealed that women who have a higher educational level showed a lower prevalence of IPV. Meanwhile, women who have higher exposure to the media have a lower prevalence of IPV. Women from all the different wealth index classes showed a balanced prevalence of intimate partner violence. Imputation was also carried out, and the main reason was to end up with complete data that will be used in the chapters that follow. The results are well addressed in Chapter 3, Chapter 4, and Chapter 5, using generalized linear models, generalized linear mixed models, and generalized additive mixed models, respectively.

Chapter 3

Generalized Linear Models

After developing the response variable IPV in Chapter 2, we apply logistic and survey logistic regression to assess the association between the response and the covariates. Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables (Hosmer Jr et al., 2013). It is often the case that the outcome variable is discrete, taking on two or more possible values. The logistic regression has become, in many fields, the standard method of analysis in this situation (Hosmer Jr et al., 2013). In this study, we want to know what the relationship is of one or more exposure variables to IPV.

3.1 Introduction

Regression techniques have become an integral element of any data analysis that describes the association between dependent variables and different predictors (Cucchiara, 2012). Logistic regression is a particular case of the Generalized Linear Model. More specifically, the outcome variable in logistic regressions is binary or dichotomous (Cucchiara, 2012). In this study, we assess the extent to which IPV is associated with knowledge of sexually transmitted infections and contraceptive methods used by women, accounting for additional variables.

3.2 Statistical Modeling

3.2.1 The Exponential Family of Distributions

Several distributions belong to the exponential family. These include discrete distributions like the Poisson or Bernoulli distribution, and Gamma or Gaussian (normal) distribution. Suppose $Z_a(a = 1, \dots, n)$ is a set of random response predictors. Z_a is part of the exponential family provided that its probability density function (pdf) can be written as:

$$f(z_a, \xi_a) = r(z_a)s(\xi_a)\exp[t(z_a)u(\xi_a)] = \exp\,[t(z_a)u(\xi_a) + v(z_a) + w(\xi_a)],$$

where ξ_a is the location parameter (Lindsey, 1997).

Furthermore, let $y = t(z)$ and $\theta = \mu(\xi)$, then the canonical form is obtained, and the model becomes:

$$f(y_a, \theta_a) = \exp[y_a\theta_a - b(\theta_a) + c(y_a)],$$

where $b(\theta_a)$ is known as the normalizing constant and $Y_a(a = 1, \dots, n)$ is a set of random response predictors with mean (μ_a). Therefore, $y_a = \mu_a + \epsilon_a$. The generalization of the exponential family can be obtained by letting ϕ be a constant scale parameter such that:

$$f(y_a, \theta_a, \phi) = \exp\left[\frac{y_a\theta_a - b(\theta_a)}{a_a(\phi)} + c(y_a, \phi)\right],$$

where θ_a is the natural parameter or canonical form of the location parameter, and some function of the mean μ_a (Lindsey, 1997; McCullagh & Nelder, 1989). $a_a(\phi)$ has the form $a_a(\phi) = \frac{\phi}{w_a}$ for known weight w_a, where ϕ is the dispersion parameter (Lindsey, 1997). When y_a is a mean of n_a independent readings, then $w_a = n_a$. The dispersion parameter ϕ is also known as the nuisance parameter, which can be used in exponential family distributions such as the normal or gamma, but is not required for one-parameter families, such as the binomial and Poisson ones (Lindsey, 1997). If an outcome Y is in the exponential family of distribution, then there is a special association between the mean and variance (Lindsey, 1997). The association between the mean and variance is given for any likelihood function $L(\theta_a, \phi_a; y_a) = f(y_a, \theta_a, \phi)$. The first derivative of its logarithm for one observation is given by:

$$U_a = \frac{\partial \log\,[L(\theta_a, \phi; y_a)]}{\partial \theta}$$

This is known as the score function. Equating the score function to zero, the resulting score equations yield the maximum likelihood estimates (Lindsey, 1997). Standard inference theory shows that $E[U_a] = 0$ and $var[U_a] = E[U_a^2] = E[-\frac{\partial U_a}{\partial \theta_a}]$. In the exponential dispersion family,

$$\log[L(\theta_a, \phi, y_a)] = \exp\,[\frac{y_a\theta_a - b(\theta_a)}{a_a(\phi)} + c(y_a, \phi)]$$

Then for θ_a, $U_a = \frac{y_a - \frac{\partial b(\theta_a)}{\partial \theta_a}}{a_a(\phi)}$ so that

$$E[Y_a] = \frac{\partial b(\theta_a)}{\partial \theta_a}$$

$$= \mu_a$$

Let $U = \frac{\partial l}{\partial \theta}$ then

$$U_a' = -\frac{\frac{\partial^2 b(\theta_a)}{\partial \theta_a^2}}{a_a(\phi)}$$

Thus, we have variance:

$var[U_a] = \frac{var[Y_a]}{a_a^2(\phi)} = \frac{\frac{\partial^2 b(\theta_a)}{\partial \theta_a^2}}{a_a(\phi)}$ (Lindsey, 1997).

There are three components of the Generalized Linear Models as outlined by McCullagh & Nelder (1989):

- The random component explains the conditional distribution of the outcome *Y* with independent variables, for an element of the exponential family such as Normal, Poisson, gamma, and binomial.
- The systematic component involving the explanatory variables $x_1, x_2, \dots, x_p$ is used as a linear predictor.
- The third component is the link function g that links the covariates to the natural mean of the outcome variable *Y* (McCullagh & Nelder, 1989).

Linear Predictor

A set of *p+1* unknown parameters, $\beta_a (a = 0,1,2, \dots, p)$, and the design matrix of known independent variables $\mathbf{X}_{n\times(p+1)}$, describe a linear predictor η given by:

$$\eta = \mathbf{X}\beta,$$

where $\mathbf{X}\beta$ is the linear structure. The a^{th} row of **X** is given by $x_a = (1, x_{a1}, \dots, x_{ap})'$ with $x_{aj}, a = 1, \dots, n$; equal to the observation of the j^{th} predictor or independent variables $x_j, j = 1, \dots, p$, and $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ is a vector of regression coefficients comprising the constant β_0 corresponding to $\mathrm{X}_0 = 1$ (Lindsey, 1997).

Link Function

The link function that is $g_a(\cdot)$, provides the association between the mean of the a^{th} observation and its linear predictor so that:

$$\eta_a = g_a(\mu_a) = \mathbf{X}'_a\beta,$$

where the function needs to be monotonic and differentiable (Lindsey, 1997). The canonical link function is the function that makes η_a, like the canonical parameter θ_a, which is in the exponential family (Lindsey, 1997). With the canonical link function, the unknown parameters in β have enough statistics, if the outcome variable distribution is in the exponential family with known scale parameters (Lindsey, 1997).

3.2.2 Model Selection

Stepwise, forward, and backward selection procedures were utilized to select important variables related to the outcome variable (Dlamini, 2016). The procedures gave similar variables/factors that were identified to be important in the model.

Akaike's Information Criteria (AIC)

One technique to evaluate a model is to utilize the Information Criteria (IC). The criterion tries to quantify how the model best fits the estimated data (Dlamini, 2016). The Akaike's Information Criterion is a statistic that helps in comparing the corresponding different models in fitting the data (Dlamini, 2016). The statistic has been proposed by Akaike, (1974) and is given by:

$$AIC = -2logLikelihood + 2k$$

where k represents the number of parameters in a model (Habyarimana, 2016; Dlamini, 2016). This approach penalizes the log-likelihood for the number of variables estimated (Akaike, 1974; Dlamini, 2016; Habyarimana, 2016). A model that minimizes the AIC is preferred (Dlamini, 2016), therefore this method is useful when comparing non-nested models (Habyarimana, 2016; Dlamini, 2016).

3.3 The Logistic Regression Model

3.3.1 Binary Data and Responses

The response variable Y can take two possible outcomes: either a 'success' or a 'failure', denoted by 1 or 0, respectively. Let π_a and $1 - \pi_a$ be the probabilities of success and failure respectively, then on the $a^{th}(a = 1, \dots, N)$, observational units $\Pr(Y_a = 1) = \pi_a$, and $\Pr(Y_a = 0) = 1 - \pi_a$. These are the probabilities of 'success' and 'failure,' respectively. In statistics, the objective is to look into the association between the outcome probability $\pi = \pi(x)$, and the independent variables $x = (x_1, \dots, x_p)$. Binary data is ungrouped data that lists individual experimental units (McCullagh & Nelder, 1989).

3.4 Stages in Building the Logistic Regression Model

There are four stages in building a logistic regression model of survey data. These are:

1) Specifying the model.
2) Estimating the parameters and standard errors.
3) Evaluating and diagnosing the model.
4) Interpreting the results and inferring based on the selected model.

3.4.1 Model Specification

The best logistic regression model for the survey data is formed by identifying the predictors and evaluating them individually and in the multivariate context, with other relevant independent variables (Lemeshow & Hosmer, 2000). To specify the initial model, the Lemeshow & Hosmer, (2000) incremental process is recommended, then the final logistic model is predicted by using the following method:

1) Initially, perform a bivariate analysis of y (outcome) relationship to individual predictor variable candidates.
2) Using the significance $p < 0.05$ as candidates for the main effects and selected predictors to have a bivariate association.
3) Using the Wald test to assess the contribution of each predictor to the multivariate model.
4) Check the assumption as to whether there is a linear relationship between the continuous variables.
5) The interactions among the predictors should be justified scientifically.

A final step recommended by Lemeshow & Hosmer (2000) is to apply the polynomial functions and smoothing splines, to test whether the logistic model is linear in the logit (Lemeshow & Hosmer, 2000).

3.4.2 The Likelihood Function

The likelihood function L is the function of the unspecified parameters denoted as $L(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ denotes the vector of unspecified parameters predicted in the model (Heeringa, et al., 2010; Moeti, 2010). The joint probability or likelihood of observing the data that has been collected, is given as $L = L(\boldsymbol{\beta})$ and $logit(\pi(x)) = \mathbf{X}_i'\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)$ (Heeringa, et al., 2010; Moeti, 2010).

3.4.3 Estimation of Model Parameters

After specifying our model, the following step is to predict the model's parameters and standard errors. Maximum Likelihood (ML) estimation is one of the methods used to predict the parameters in a mathematical model (Heeringa, et al., 2010; Moeti, 2010). Another method is the least-squares (LS) estimation, mainly useful in predicting the parameters in a classified linear or multiple linear regression model (Moeti, 2010; Dlamini, 2016). The predictors can be nominal, ordinal, and lastly interval when using ML estimation (Heeringa, et al., 2010).

When choosing the ML method, the number of parameters relative to the total number of subjects in one's model plays an important role (Heeringa, et al., 2010). Unconditional ML estimation is favored if the number of parameters in the model is low in relation to the number of subjects (Dlamini, 2016; Heeringa, et al., 2010). Conditional ML prefers large numbers of parameters that are more than clusters or groups, like in the survey of a country relative to the subjects (Heeringa, et al., 2010).

For example, if we have a logistic regression that takes the form $logit\big(\pi(x)\big) = \beta_0 + \mathrm{X}_{i1}\beta_1 + \cdots + \mathrm{X}_{ok}\beta_k$; given that the sample data follows a simple random sample (SRS), the parameters $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are estimated using the Maximum Likelihood (ML) approach (Moeti, 2010). Likelihood functions for an SRS with *n* values on a binary predictor *y*, with likely values 0 or 1, is established on the binomial distribution given by:

$$\boldsymbol{L}(\boldsymbol{\beta}|\boldsymbol{x}) = \prod_{\boldsymbol{a}=\boldsymbol{1}}^{\boldsymbol{n}} \boldsymbol{\pi}(\boldsymbol{x_a})^{y_a}[\boldsymbol{1} - \boldsymbol{\pi}(\boldsymbol{x_a})]^{\boldsymbol{1}-y_a}, \tag{3.4.1}$$

where:

$$\pi(x_a) = \frac{e^{(x_a\boldsymbol{\beta})}}{1 + e^{(x_a\boldsymbol{\beta})}} \tag{3.4.2}$$

Maximum Likelihood Estimation is not possible when survey data has been gathered under a complex sample design (Dlamini, 2016). In this case, we require sampling weights to work out the logistic regression parameters (Moeti, 2010; Dlamini, 2016). The weighted least squares (WLS) estimation and the pseudo-maximum likelihood estimation (PMLE), are the two methods used to estimate model parameters for logistic models of complex survey data (Heeringa, et al., 2010; Moeti, 2010). If the population is finite, then the likelihood equations are maximized by the regression parameters for $a = 1,\cdots,N$ (where N is the population size) components in the population under survey (Heeringa, et al., 2010; Moeti, 2010).

Letting $\boldsymbol{\beta}$ be the finite population model parameters, the population likelihood for a binary dependent variable y is given by:

$$L(\boldsymbol{\beta}|x) = \prod_{a=1}^{N} \pi(x_a)^{y_a}[1 - \pi(x_a)]^{1-y_a}, \tag{3.4.3}$$

where:

$$\pi(x_a) = \frac{e^{(x_a\boldsymbol{\beta})}}{1 + e^{(x_a\boldsymbol{\beta})}}$$

Maximizing the estimates for the population likelihood, which is the weighted function of the sample data and $\pi(x_a)$ observations, we get the predictors of the finite population regression parameters as shown below (Moeti, 2010):

$$PL(\boldsymbol{\beta}|x) = \prod_{a=1}^{n} \{\pi(x_a)^{y_a} \cdot [1 - \pi(x_a)]^{1-y_a}\}^{w_a} \tag{3.4.4}$$

where:

$\pi(x_a) = \frac{e^{(x_a\boldsymbol{\beta})}}{1+e^{(x_a\boldsymbol{\beta})}}$ and $w_a = \frac{n_a}{N}$.

When using logistic regression models to analyze complex survey data, it is essential to estimate the sampling variances and covariates of the parameter estimates (Heeringa, et al., 2010; Moeti, 2010;

Dlamini, 2016). The sampling variance is estimated because in the above-mentioned model, the random effects are not taken into account and hence the model might not be a good fit for the data. And there is a greater variability within clusters.

The Newton-Raphson Method

The method is useful for the approximation of continuous variables, these variables are easily approximated to find the linear relationship. One of the methods for solving nonlinear equations is the Newton-Raphson method. The examples are likelihood equations that show where a function is maximized. The Newton-Raphson method determines the value $\hat{\beta}$ of β that maximizes a function $\tau(\beta)$. Let $g' = \left(\frac{\partial_\tau}{\partial\beta_1}, \frac{\partial_\tau}{\partial\beta_2}, \cdots\right)$, and let the Hessian $\boldsymbol{H} = h_{xy}$ denote the matrix of the second derivative that is $\boldsymbol{H} = \frac{\partial_\tau^2}{\partial\beta_x\partial\beta_y}$.

Let $g^{(m)}$ and $\boldsymbol{H}^{(\boldsymbol{m})}$ be the values evaluated at $\beta^{(m)}$, the m^{th} guess for $\hat{\beta}$. The m^{th} iteration process $(m = 0,1,2,\cdots)$, $\tau(\beta)$ is approximated near $\beta^{(m)}$ by the terms up to second order in its Taylor series expansion (Heeringa, et al., 2010; Moeti, 2010)

$$Q^{(m)}(\beta) = \tau\left(\beta^{(m)}\right) + g^{(m)'}\left(\beta - \beta^{(m)}\right) + (\frac{1}{2})(\beta - \beta^{(m)})'\boldsymbol{H}^{(\boldsymbol{m})}\left(\beta - \beta^{(m)}\right), \qquad (3.4.5)$$

then:

$$\frac{\partial Q^{(m)}}{\partial\beta} = g^{(m)} + \boldsymbol{H}^{(\boldsymbol{m})}\left(\beta - \beta^{(m)}\right) = 0 \qquad (3.4.6)$$

after solving for equation (3.4.6, we then get:

$$\beta^{(m+1)} = \beta^{(m)} - (\boldsymbol{H}^{(m)})^{(-1)}g^{(m)} \qquad (3.4.7)$$

assuming $\boldsymbol{H}^{(\boldsymbol{m})}$ is non-singular. $\beta^{(0)}$ denotes the first estimate of $\hat{\beta}$ (Heeringa, et al., 2010; Dlamini, 2016). Then each iteration $\beta^{(m)}$ is utilized to obtain $H^{(m)}$and $g^{(m)}$, which are then utilized to estimate $\beta^{(m+1)}$, which in turn is used to obtain $\beta^{(m+2)}$, and the process carries on until convergence (Heeringa, et al., 2010; Moeti, 2010).

3.4.4 Goodness of Fit and Diagnostics

The measure of discrepancy called deviance was introduced by Nelder and Wedderburn (1972). Let μ denote the mean value parameter, θ the canonical parameter, and ϕ some dispersion parameter. Let $l(\hat{\mu}, \phi, y)$ be the log-likelihood maximized over the vector of parameters β for a fixed value of ϕ, and $l(\hat{\mu}, \phi, y)$ be the maximum log-likelihood achievable in the saturated model (Dlamini, 2016). The scaled deviance is as follows:

$$D^{*} = -2\frac{[log(\hat{\mu}, \phi, y) - log(y, \phi, y)]}{\phi}. \quad (3.4.8)$$

The scaled deviance D^{*} is the deviance expressed as a multiple of the dispersion parameter ϕ (Dlamini, 2016). If $\phi = 1$ then the deviance is described as:

$$D = -2[log(\hat{\mu}, \phi, y) - log(y, \phi, y)] \quad (3.4.9)$$

If $\phi = 1$ or $\phi \neq 1$ (but the value of ϕ is known), we can measure the closeness of the fit of a model to the data (Heeringa, et al., 2010; Dlamini, 2016). If there is small deviance with the log-likelihood close to $\log(y, \phi, y)$, then the model describes the data well. Large deviance shows that the data is not well fitted by the model. D has an approximate χ^2 distribution with $n - p - 1$ degrees of freedom, where p is the number of independent variables in the linear predictor (Heeringa, et al., 2010; Moeti, 2010; Dlamini, 2016). The fitted model is adequate if $D \leq \chi^2_{\alpha, n-p-1}$. If the deviance ratio to its degrees of freedom $n - p - 1$ is close to 1, we can conclude that the fitted model is sufficient (Moeti, 2010; Heeringa, et al., 2010; Dlamini, 2016). A substantial value means that the model is incorrectly specified. Another critical measure of inconsistency is the generalized Pearson χ^2 statistic (Heeringa, et al., 2010). It is given by:

$$\chi^2 = \sum_{a=1}^{n} \frac{(y_a - \hat{\mu}_a)^2}{V(\hat{\mu}_a)} \quad (3.4.10)$$

where $(\hat{\mu}_a)$ is the predicted variance function for the distribution concerned (Dlamini, 2016). For normal-theory linear models, the deviance, and the generalized Pearson χ^2 statistic, have the exact χ^2 distribution with $n - p - 1$ degrees of freedom (Heeringa, et al., 2010; Dlamini, 2016). The deviance as a measure of inconsistency has the advantage that it is an additive for nested sets of models, if maximum likelihood estimates are utilized (McCullagh & Nelder, 1989; Heeringa, et al., 2010; Moeti, 2010; Dlamini, 2016).

Many approaches can be used to test whether the logistic regression model fits the data. Several approaches use the idea of comparing the recorded number of individuals with the expected one, if the fitted model is the valid one (Dlamini, 2016). The observed (O) and the expected (E) numbers are amalgamated to form a χ^2 goodness-of-fit statistic (Dlamini, 2016). Substantial values of the test statistic show that the model is a poor fit as with small p-values. The goodness-of-fit test statistic of the classic approach is given by:

$$\chi^2 = \sum 2O(ln\frac{O}{E}) \quad (3.4.11)$$

The goodness-of-fit test statistic (Hosmer Jr, et al., 2013; Dlamini, 2016) is given by:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \tag{3.4.12}$$

Predictive Accuracy

It is essential to inspect the degree to which the estimated probability resembles the outcomes (Moeti, 2010; Dlamini, 2016). The goal is to come up with a model that maximizes the probability and sensitivity of recognizing individuals whose intervention is justified (Moeti, 2010; Dlamini, 2016). This means that we are interested in reducing the proportion of individuals classified incorrectly as having an outcome or failure. We can validate the model by checking the prediction accuracy, which could be done by checking how accurately the model predicts the outcome.

To inspect for the estimative accuracy of the SAS procedure, PROC LOGISTIC yields other model statistics such as Somer's D, Gamma, C, and Tau-a (Dlamini, 2016). All these statistics span between 0 and 1 (Moeti, 2010; Dlamini, 2016). Larger values correspond to a strong relationship between estimated and observed values (Moeti, 2010; Dlamini, 2016). These measures of relationship are given by:

$$Tau - a = \frac{C - D}{N},$$

$$Gamma = \frac{C - D}{C + D},$$

$$Somer's\ D = \frac{C - D}{C + D + T}$$

where *C* is the percentage of observation set with distinct observed results (Dlamini, 2016; Moeti, 2010). The model properly predicts high probabilities for observations with the event result, compared to probabilities for non-event observations (Dlamini, 2016; Moeti, 2010). An outcome of 1 implies the model allocates high probabilities to observations with the event observed, compared to those with a non-event observed (Moeti, 2010; Dlamini, 2016). Concordant and discordant sets are used to define the association between sets of observed values (Dlamini, 2016). The sets are known to be concordant if the subject is ranked higher for the independent variable X, and is also ranked higher for the outcome predictor Y (Moeti, 2010; Dlamini, 2016). The sets are discordant if the subject ranked higher for independent variable X, and ranks lower for the outcome predictor Y (Moeti, 2010; Dlamini, 2016). The set is tied (T) if subjects have the same categorization for independent and outcome variables (Dlamini, 2016). C represents the receiver operating characteristics (ROC) curve in the binary response case, which is defined below (Simundi´c, 2008; Dlamini, 2016).

Area Under the Receiver Operating Characteristics (AUROC)

Specificity and sensitivity depend on the cut-off point to categorize the outcome as positive (Lemeshow & Hosmer, 2000; Dlamini, 2016). In plotting the ROC, one is required to plot sensitivity versus 1-specificity (Lemeshow & Hosmer, 2000; Dlamini, 2016). Sensitivity is used to measure the percentage of correctly categorized positive results or events the study focuses on (Dlamini, 2016; Moeti, 2010). Specificity is used to measure the percentage of correctly categorized non-event results (Moeti, 2010; Dlamini, 2016). The ROC is useful in providing a complete description of classification accuracy and is most useful as a graphical display of the model prediction accuracy (Vittinghoff, et al., 2011; Simundi´c, 2008; Dlamini, 2016). It also defines the model's ability to categorize between event and non-event subjects (Dlamini, 2016) The area under the curve (AUC) is between 0 and 1, as shown in Figure 3.1 (Simundi´c, 2008; Dlamini, 2016)..

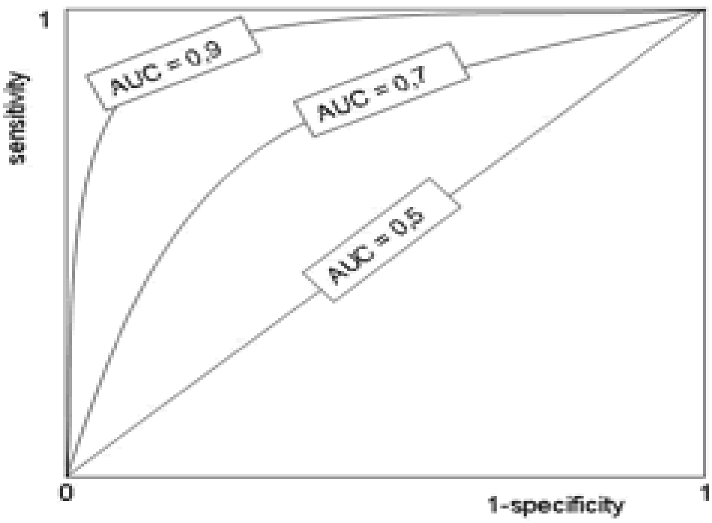

Figure 3.1: Receiver Operating Characteristic (ROC) curve

Source: (Simundi´c, 2008; Dlamini, 2016).

The AUC is known as the global measure of diagnostic accuracy (Dlamini, 2016). The area is used to measure the model prediction accuracy (Dlamini, 2016). However, AUC does not tell us anything about individual parameters (Simundi´c, 2008). Suppose the AUC is substantial, then the diagnostic accuracy of the test improves (Simundi´c, 2008; Dlamini, 2016). Let us assume that three different models were fitted, with model 1 yielding an AUC of 0.55, model 2 yielding an AUC of 0.73, and model 3 yielding an AUC of 0.94. Model 3 can be categorized as the best model since it has great diagnostic accuracy, therefore, it has the best accuracy (Simundi´c, 2008; Dlamini, 2016). An AUC of 0.5 is bad since the test cannot correctly classify positive outcomes or those falsely classified as positive (Simundi´c, 2008; Dlamini, 2016). Table 3.1 below shows the relationship between the AUC and diagnostic accuracy (Dlamini, 2016).

Table 3.1: Relationship between area under the curve and diagnostic accuracy

Area	Diagnostic Accuracy
0.9≤AUC< 1.0	Excellent
0.8≤AUC< 0.9	Very good
0.7≤AUC< 0.8	Good
0.6≤AUC<0.7	Sufficient
0.5≤AUC < 0.6	Bad
<0.5	Test not useful

Source: (Simundi´c, 2008; Dlamini, 2016)

3.4.5 Interpretation and Inference

Logistic regression modeling mostly uses Wald χ^2 test and confidence intervals to formulate inferences about the significance of independent covariates (McCullagh & Nelder, 1989; Dlamini, 2016). They also give details on the strength and uncertainty related to the predicted effects of individual independent variables (Dlamini, 2016). The Confidence Interval (CI) for the logistic regression model is computed as follows:

$$CI_{1-\alpha}(B_j) = \hat{B}_j \pm Z_{1-\frac{\alpha}{2}} \times se(\hat{B}_j) \tag{3.4.13}$$

where $\alpha = 5\%$ and the df is based on the design, this being a 95% confidence interval for the parameter, which includes the actual population value of $\hat{B}_j$ (McCullagh & Nelder, 1989; Dlamini, 2016). A logistic regression model with one predictor, x_1, can give an estimate of the unadjusted odds ratio. It is given by:

$$\hat{\psi} = exp\ (\hat{B}_1) \tag{3.4.14}$$

If a logistic regression has multiple predictors, then the result is an adjusted odds ratio:

$$\hat{\psi}_j | \hat{B}_{k\pm j} = exp\ (\hat{B}_j) \tag{3.4.15}$$

CI limits for the adjusted odds ratio are given by:

$$CI_{\psi_j} = exp\ (\hat{B}_j \pm Z_{1-\frac{\alpha}{2}} \times se(\hat{B}_j)) \tag{3.4.16}$$

Categorical variables, ordinal, and continuous variables can be estimated using adjusted odd ratios and CI (McCullagh & Nelder, 1989).

3.5 Limitations of Logistic Regression

When using logistic regression, there are no assumptions made with respect to the distribution of the covariates, although covariates are not supposed to be highly correlated to each other (Dlamini, 2016). The correlation may lead to estimation problems (Dlamini, 2016). Also, the greater the number of independent variables there are, the greater the need for a larger sample size (Dlamini, 2016). The smaller the sample size, the less powerful the Hosmer-Lemeshow test will be (Hosmer Jr, et al., 2013; Dlamini, 2016). The other constraint is the existence of a non-linear relationship between the log odds and covariates, leading to the possibility of the results obtained being invalid (Dlamini, 2016); furthermore, ordinary logistic regression does not consider the complex nature of the survey design, leading to invalid inference (Dlamini, 2016).

3.6 Fitting the Logistic Regression Model

In fitting the model, the logistic procedure in SAS was used; univariate models were fitted to identify potential candidate variables associated with the outcome, without considering the combined effects of covariates on the dependent variable. The multiple logistic models were then fitted with all the predictors that were significant in the univariate analysis. The goodness-of-fit was tested utilizing the Hosmer-Lemeshow test (Dlamini, 2016), and the model's predictive accuracy was assessed through the ROC. Interaction effects to determine the association between age of the woman and contraceptive use, and wealth index against contraceptive use have been included in the model.

South Africa

Table 3.6 displays results from the univariate models fitted. The results shown in this table confirm some of the bivariate results in Section 2.5.1 South Africa. The significant covariates had p-values that were less than 0.05. In this section, only covariates significant (p-value<0.05) to IPV have been interpreted under each category. The significant covariates are relevant to a specific country, and it makes sense to interpret what is relevant to a specific country.

Model Checking

Multicollinearity was checked for the variables in the model, and it was not detected. There are no variables that contain the same information in the dataset used for this study. To check how good the model is, we can utilize the Hosmer-Lemeshow test (Dlamini, 2016). The Chi-square statistic for this test with a corresponding p-value of 0.3393, is shown in Table 3.3. From this table, there is insufficient proof to assess if the model is the best fit; thus, we can infer that the model fits the data adequately. Predicted probabilities are approximately the same as the observed values.

Model Selection

Table 3.2 shows the model fit statistics that are used to compare the two models. From the results, the full model has a smaller AIC than the model with only the intercept. Hence, the model containing the intercept and covariates performs the best.

Table 3.2: Model fit statistics

Criterion	Intercept Only	Intercept and Covariates (Full Model)
AIC	9922.214	8497.566
SC	9929.263	8821.842
-2 Log L	9920.214	8405.566

Table 3.3: Hosmer and Lemeshow Goodness-of-fit test

Chi-Square	DF	P-value
9.0357	8	0.3393

Prediction Accuracy of the Model

Table 3.***4*** shows the relationship between predicted probabilities and observed outcomes, with an AUC of $\mathrm{C} = 0.761$ and a concordant rate of 76.1, showing us how good the model is for separating the 0's and 1's from the selected model. Figure 3.2 shows the ROC curve of the fitted model and the area under the curve $\mathrm{c} = 0.761$, which implies that 76.1% of the probabilities are predicted correctly, which is a good predictive accuracy. The value for Gamma is 0.523, which suggests that there is no perfect association. It is interpreted as giving 51.6% fewer errors in prediction, utilizing the estimated probabilities, compared to a chance alone. One weakness of this statistic is the tendency to overstate the strength of association between probabilities and outcomes. The value for Somer's D is 0.523, this shows that not all pairs are concordant, and we may utilize it to compare models.

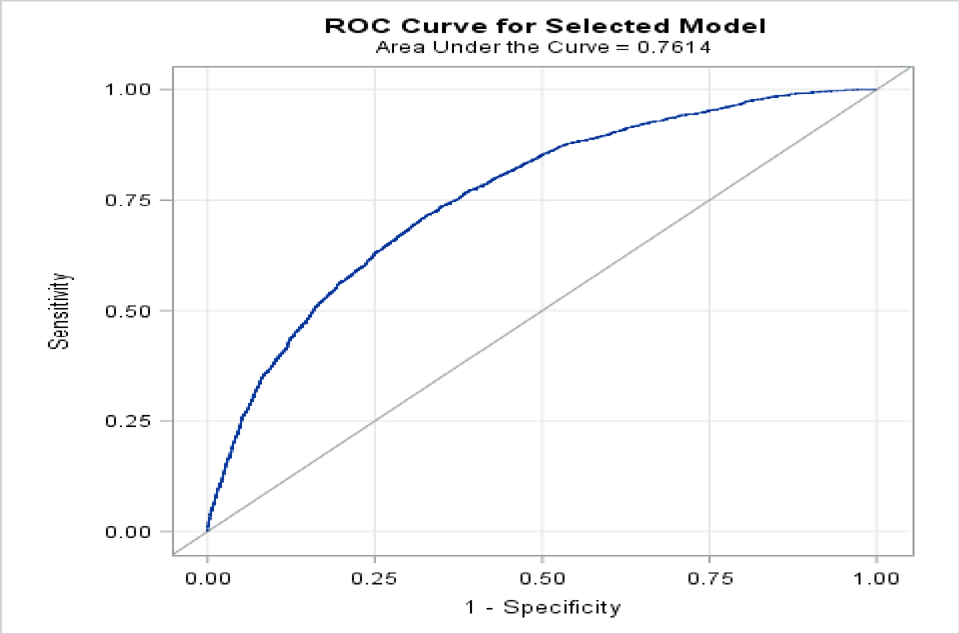

Figure 3.2: ROC curve for selected model

Table 3.4: Association of predicted probabilities and observed responses

Concordant	76.1%	Somers' D	0.523
Discordant	23.9%	Gamma	0.523
Tied	0.0%	Tau-a	0.206
Pairs	14264753	C	0.761

Interpretation of the Results

The covariates used in the model are shown in Table 3.5 with their p-values. A covariate with a p-value of less than 0.05 has a significant association with the response variable.

Table 3.5: Type 3 analysis of effects

Effect	**DF**	**Wald Chi-square**	**Pr > ChiSq**
Partner drinks alcohol	2	259.4493	<.0001
Woman's father ever beat her mother	2	113.9519	<.0001
Access to the media	2	12.6659	0.0018
Wife-beating attitude	2	48.8873	<.0001
Woman's current age	1	32.9220	<.0001
Region	8	56.5075	<.0001
Number of household members	1	182.9187	<.0001
Sex of the household head	1	10.7935	0.0010
Wealth index	4	77.2688	<.0001

Ever had a terminated pregnancy	1	21.0563	<.0001
Contraceptive method used	1	49.9150	<.0001
Body Mass Index	1	10.1313	0.0015
Current marital status	2	64.1680	<.0001
Cohabitation duration	1	5.7664	0.0163
Partner's desire for children	3	49.2085	<.0001
Woman's occupation	2	17.1661	0.0002
Partner's age	1	10.2855	0.0013
Woman's earnings compared to partner	4	13.2058	0.0103
Knowledge of Sexually Transmitted Infections (STIs)	1	12.5045	0.0004
Interaction effects			
Woman's age by contraceptive use	1	27.2953	<.0001
Wealth index by contraceptive use	4	12.6044	0.0134

***DF=Degrees of Freedom**

Table 3.6 shows that a woman whose partner does not drink alcohol is 0.44 (OR=0.440, p-value<.0001) times less likely to experience IPV, compared to one whose partner does drink it. A woman who has never witnessed her father beating her mother is 0.44 (OR=0.439, p-value<.0001) times less likely to experience IPV, compared to one who has witnessed her father beating her mother. A woman who does not know if her father beating her mother is 0.73 (OR=0.725, p-value=0.0182) times less likely to experience IPV, compared to one who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.77 (OR=1.7727, p-value<.0001) times more likely to experience IPV, compared to one who views wife-beating as unacceptable. A woman who does not know whether wife-beating is acceptable or unacceptable is 0.32 (OR=0.3249, p-value<.0001) times less likely to experience IPV, compared to one who views wife-beating as unacceptable.

A woman with medium exposure to the media is 1.34 (OR=1.336, p-value=0.0007) times more likely to experience IPV, compared to one with low exposure to the media. A woman with high exposure to the media is 1.25 (OR=1.254, p-value=0.0279) times more likely to experience IPV, compared to a woman with low exposure to the media. A unit increase in the woman's age increases the chances of her experiencing IPV by 0.0122 units.

A woman from the Western Cape province is 0.65 (OR=0.645, p-value=0.0008) times less likely to experience IPV, compared to a woman from the Eastern Cape province. A woman from the Northern Cape region is 0.53 (OR=0.530, p-value<.0001) times less likely to experience IPV, compared to one from the Eastern Cape region. A one-member increase in the number of members in a household decreases a woman's chances of experiencing IPV by 0.1607 units. A woman from a household where the head of the house is male has 1.24 (OR=1.240, p-value=0.0009) times more likely to experience IPV, compared to one from a household where the head of the house is female.

A woman from the poorest wealth index class is 1.85 (OR=1.853, p-value<.0001) times more likely to experience IPV, compared to a woman from the class with the richest wealth index. A woman from a poorer wealth index class is 2.13 (OR=2.130, p-value<.0001) times more likely to experience IPV, compared to a woman from the class with the richest wealth index. A woman from a middle wealth index class is 1.69 (OR=1.689, p-value=0.0002) times more likely to experience IPV, compared to a woman from the class with the richest wealth index. A woman from a richer wealth index class is 1.58 (OR=1.579, p-value=0.0011) times more likely to experience IPV, compared to a woman from the class with the richest wealth index.

A woman who has never terminated a pregnancy is 0.66 (OR=0.660, p-value<.0001) times less likely to experience IPV, compared to a woman who has done so. A unit increase in a woman's body mass index decreases her chances of experiencing IPV by 0.0022 units. A married woman is 1.70 (OR=1.704, p-value<.0001) times more likely to experience IPV, compared to a single woman. A woman living with her partner is 1.81 (OR=1.805, p-value<.0001) times more likely to experience IPV, compared to a single woman. A woman who stayed with her partner for 5-9 years is 0.80 (OR=0.802, p-value=0.0322) times less likely to experience IPV compared, to a woman who stayed with her partner for 0-4 years.

A woman with a partner who wants more children compared to her is 1.36 (OR=1.356, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does. A woman with a partner who wants fewer children compared to her is 1.29 (OR=1.289, p-value=0.0412) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does. A woman who does not know the number of children her partner wants is 1.36 (OR=1.356, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does.

A woman who has an employed partner is 0.84 (OR=0.842, p-value=0.0241) times less likely to experience IPV, compared to a woman who does not know if her partner is employed or not. An employed woman is 1.30 (OR=1.295, p-value<.0001) times more likely to experience IPV, compared to an unemployed woman. A one-year increase in the age of a woman's partner's decreases the chances of experiencing IPV by 0.0123 units. A woman who earns about the same as her partner is 1.19 (OR=1.186, p-value=0.0262) times more likely to experience IPV, compared to a woman who earns less compared to her partner. A woman whose partner does not bring in earnings is 1.41 (OR=1.411, p-value=0.0007) times more likely to experience IPV, compared to a woman who earns less compared to her partner.

Table 3.6: Logistic regression model coefficients, standard errors, and odds ratios (South Africa)

Parameter	Estimate	Standard error	Chi-Square value	P-value	Odds ratio
Intercept	0.5656	0.2815	4.0357	0.0445	
Partner drinks alcohol (ref=Yes)					
No	-0.8209	0.0553	220.1755	<.0001	0.440
Don't know	0.6884	0.6263	1.2083	0.2717	1.991
Woman's father ever beat her mother (ref=Yes)					
No	-0.8233	0.0773	113.4144	<.0001	0.439
Don't know	-0.3209	0.1359	5.5775	0.0182	0.725
Access to the media (ref=Low exposure)					
Medium exposure	0.2897	0.0859	11.3842	0.0007	1.336
High exposure	0.2262	0.1029	4.8313	0.0279	1.254
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,5725	0,105	29,7377	<.0001	1,7727
I don't know	-1,1243	0,269	17,4678	<.0001	0,3249
Woman's current age	0.0122	0.00620	3.8707	0.0491	1.012
Region (ref=Eastern Cape)					
Western Cape	-0.4393	0.1315	11.1606	0.0008	0.645
Northern Cape	-0.6347	0.1294	24.0690	<.0001	0.530
Free State	0.0550	0.1144	0.2311	0.6307	1.057
Kwazulu-Natal	0.0336	0.1035	0.1052	0.7457	1.034
North West	-0.0256	0.1117	0.0527	0.8184	0.975
Gauteng	0.00839	0.1153	0.0053	0.9420	1.008
Mpumalanga	0.0102	0.1072	0.0090	0.9244	1.010
Limpopo	0.1858	0.1069	3.0221	0.0821	1.204
Number of household members	-0.1607	0.0115	196.2019	<.0001	0.852
Sex of household head (ref=Female)					
Male	0.2149	0.0647	11.0441	0.0009	1.240
Wealth index (ref=Richest)					
Poorest	0.6168	0.1509	16.7171	<.0001	1.853
Poorer	0.7562	0.1431	27.9193	<.0001	2.130
Middle	0.5239	0.1392	14.1607	0.0002	1.689
Richer	0.4569	0.1405	10.5751	0.0011	1.579
Ever had a terminated pregnancy (ref=Yes)					
No	-0.4158	0.0861	23.3347	<.0001	0.660
Contraceptive method used (ref=Yes)					
No	-1.6123	0.2594	38.6411	<.0001	0.199
Body Mass Index	-0.00220	0.000985	4.9805	0.0256	0.998
Current marital status (ref=Single)					
Married	0.5331	0.0784	46.2055	<.0001	1.704
Living with partner	0.5908	0.0875	45.5616	<.0001	1.805
Cohabitation duration (ref=0 to 4 years)					
5 to 9 years	-0.2212	0.1033	4.5872	0.0322	0.802

Continued…

Partner's desire for children (ref=Both want same)					
Partner wants more	0.3046	0.0736	17.1111	<.0001	1.356
Partner wants fewer	0.2540	0.1244	4.1680	0.0412	1.289
Don't know	0.3093	0.0667	21.5236	<.0001	1.362
Woman's occupation (ref=Unemployed)					
Employed	0.2582	0.0612	17.7918	<.0001	1.295
Partner's age	-0.0123	0.00456	7.2262	0.0072	0.988
Woman's earnings compared to partner (ref=Less than him)					
More than him	0.0220	0.0723	0.0928	0.7607	1.022
About the same	0.1709	0.0768	4.9443	0.0262	1.186
Partner does not bring in	0.3442	0.1013	11.5487	0.0007	1.411
Don't know	-0.0847	0.1755	0.2328	0.6295	0.919
Knowledge of Sexually Transmitted Infections (STIs) (ref=Yes)					
No	-0.8482	0.2279	13.8527	0.0002	0.428
Interaction effects					
Woman's age by contraceptive use(ref=Using)	0.0285	0.00579	24.2090	<.0001	1.029
Wealth index by contraceptive use					
Poorest by not using contraceptives	0.4954	0.2035	5.9286	0.0149	1.641
Poorer by not using contraceptives	0.2314	0.2001	1.3370	0.2476	1.260
Middle by not using contraceptives	0.5496	0.1977	7.7299	0.0054	1.733
Richer by not using contraceptives	0.2709	0.2046	1.7533	0.1855	1.311

Table 3.6 shows that a woman who has never heard of STIs is 0.43 (OR=0.428, p-value=0.0002) times less likely to experience IPV than a woman who has heard of STIs.

Interaction Effects

Figure 3.3 shows that IPV increases with increasing age, whether a woman is using contraceptives or not. We observe from the same figure that IPV is higher among women using contraceptives than women not using them.

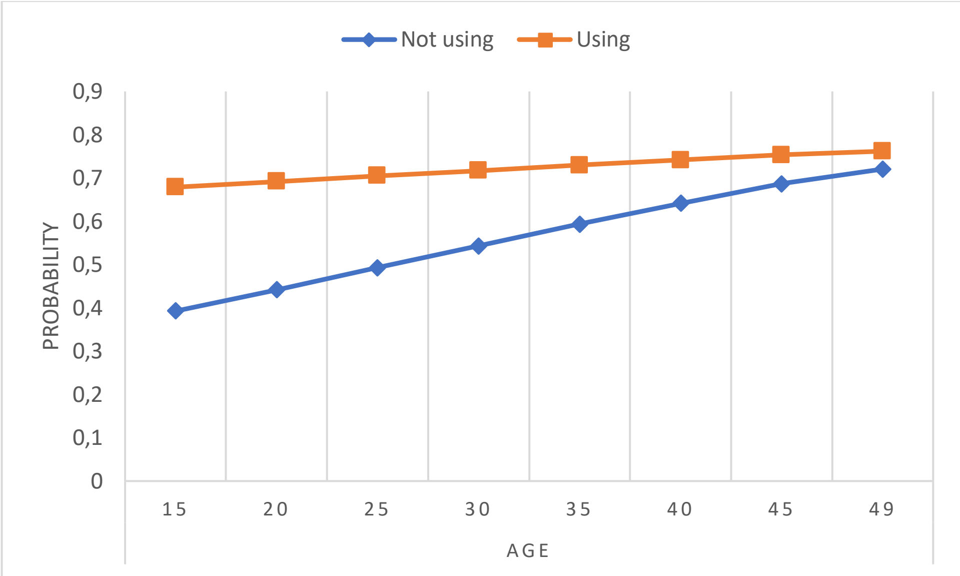

Figure 3.3: Predicted probability of experiencing IPV by woman's age and contraceptive use

Figure 3.4 shows that, for a woman not using contraceptives, IPV decreases for women from the poorest to a poorer wealth index class; it then increases from a poorer to a middle wealth index class and decreases from a middle to a richer wealth index class. For a woman using contraceptives, we observe from the same figure that IPV increases for a woman from the poorest to a poorer wealth index class, and then decreases for a woman from a poorer, middle, and richer wealth index class.

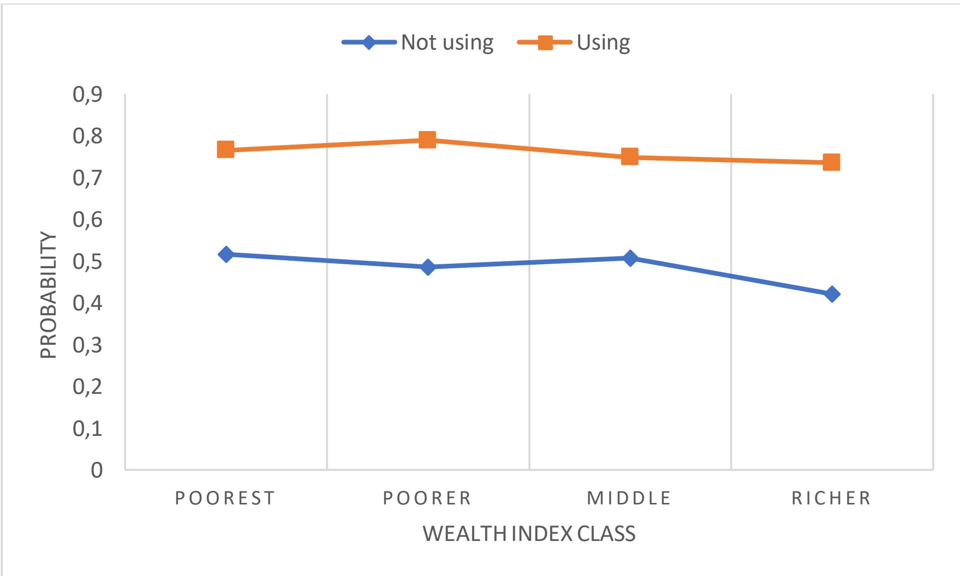

Figure 3.4: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Tanzania

Model Selection

Model Checking

Multicollinearity was checked for the variables in the model, and there were no variables found to have similar information. This means that no variable can be dropped from the model. The Chi-square statistic for the Hosmer-Lemeshow test with the corresponding p-value of <.0001, is shown in Table 3.8. This indicates that there is enough proof to claim that the model is not a good fit for the data; in other words, predicted probabilities are not the same as the observed values.

Table 3.7 below shows the model fit statistics that are used to compare the two models. The best model is the one with the intercept and covariates, since it has the smallest AIC compared to the AIC model with the intercept only.

Table 3.7: Model fit statistics

Criterion	**Intercept only**	**Intercept and covariates (full model)**
AIC	18243.286	13270.439
SC	18250.778	13697.537
2 Log L	18241.286	13156.438

Table 3.8: Hosmer and Lemeshow goodness-of-fit test

Chi-Square	DF	P-value
83.2413	8	<.0001

Prediction Accuracy of the Model

Table 3.9 shows the relationship between predicted probabilities and observed outcomes, with the area under the curve being $c = 0.831$ and a concordant rate of 83.1, which tells us how good the model is for separating the 0's and 1's, using a selected model. Figure 3.5 shows the ROC curve of the fitted model and the area under the curve $c = 0.831$, which implies that 83.1% of the probabilities are predicted correctly, which is a very good predictive accuracy. The value for Gamma is 0.663, which suggests that there is no perfect relationship. It is interpreted as meaning that 66.3% fewer errors are made in prediction by using the estimated probabilities, than by using a chance alone. One weakness of this statistic is the tendency to overstate the strength of relationship between probabilities and outcomes. The value for Somer's D is 0.663. This shows that not all pairs are concordant, and we may use it to compare models.

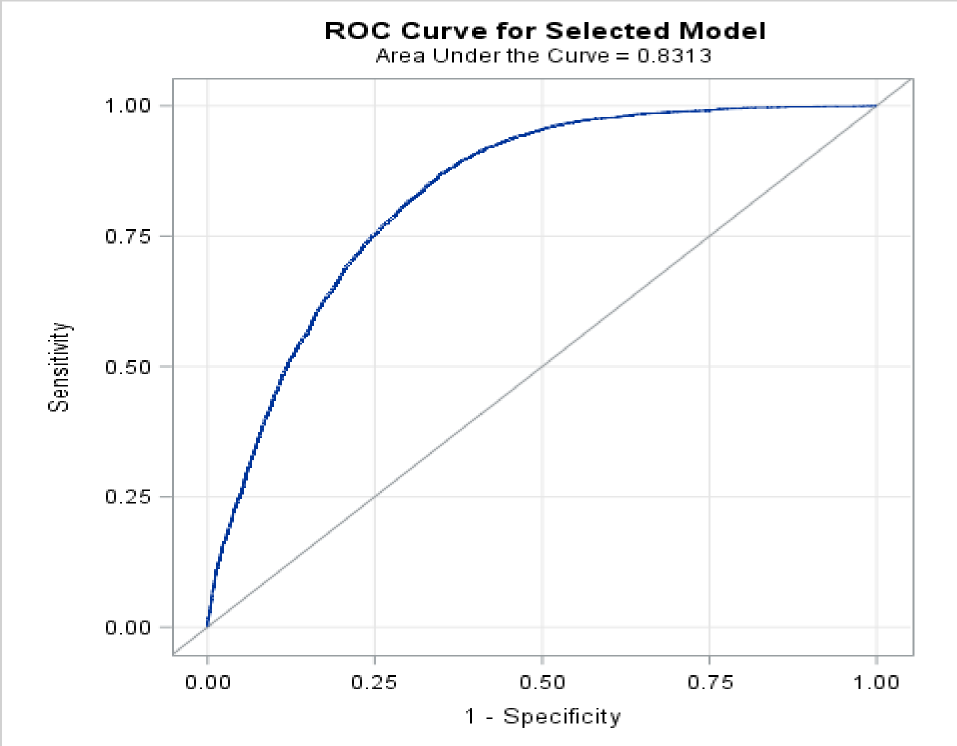

Figure 3.5: ROC curve for selected model

Table 3.9: Association of predicted probabilities and observed responses

Concordant	83.1%	Somers' D	0.663
Discordant	16.9%	Gamma	0.663
Tied	0.0%	Tau-a	0.328
Pairs	43502480	C	0.831

Interpretation of the Results

The covariates used in the model are shown in Table 3.10 with their p-values. A covariate with a p-value of less than 0.05 has a significant association with the response variable.

Table 3.10: Type 3 analysis of effects

Effect	**DF**	**Wald Chi-square**	**P-value**
Partner drinks alcohol	1	278.0779	<.0001
Woman's father ever beat her mother	2	128.8678	<.0001
Wife-beating attitude	2	103.3104	<.0001
Woman's current age	1	51.2735	<.0001
Region	29	168.6027	<.0001
Woman's level of education	3	46.2212	<.0001

Continued…

Number of household members	1	666.2697	<.0001
Sex of the household head	1	6.7188	0.0095
Wealth index	4	18.8412	0.0008
Ever had a terminated pregnancy	1	7.9283	0.0049
Contraceptive method used	1	60.7510	<.0001
Current marital status	2	894.0617	<.0001
Number of other partners	1	16.6220	<.0001
Cohabitation duration	1	82.0519	<.0001
Woman's occupation	1	30.1374	<.0001
Partner's age	1	9.1468	0.0025
The person who usually decides how to spend a woman's earnings	3	36.1463	<.0001
Interaction effects			
Woman's age by contraceptive use	1	30.5095	<.0001

***DF=Degrees of Freedom**

Table 3.11 shows that a woman whose partner does not drink alcohol is 0.38 (OR=0.379, p-value<.0001) times less likely to experience IPV, compared to a woman whose partner does drinks it. A woman who has never witnessed her father beating her mother is 0.60 (OR=0.596, p-value<.0001) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.50 (OR=1.4962, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A woman who does not know whether wife-beating is acceptable or unacceptable is 0.66 (OR=0.6602, p-value=0.0007) times less likely to experience IPV, compared to a woman who views wife-beating as unacceptable.

A woman from Kilimanjaro province is 0.49 (OR=0.485, p-value<.0001) times less likely to experience IPV, compared to a woman from Arusha province. A woman from Tanga province is 0.60 (OR=0.598, p-value<.0001) times less likely to experience IPV, compared to a woman from Arusha province. A woman from Mara province is 1.42 (OR=1.424, p-value=0.0279) times more likely to experience IPV, compared to a woman from Arusha province. A woman from Njobe province is 0.52 (OR=0.523, p-value=0.0002) times less likely to experience IPV, compared to a woman from Arusha province. A woman from Kaskazini Pemba province is 0.41 (OR=0.410, p-value<.0001) less likely to experience IPV, compared to a woman from Arusha province. A woman from Kusini Pemba province is 0.61 (OR=0.612, p-value=0.0087) times less likely to experience IPV, compared to a woman from Arusha province.

A woman with secondary education is 0.714 (OR=0.714, p-value<.0001) times less likely to experience IPV, compared to a woman who has no education. A woman who has a higher education is 0.371

(OR=0.371, p-value<.0001) times less likely to experience IPV, compared to a woman who has no education. A one-member increase in the number of household members decreases the chances of a woman experiencing IPV by 0.1936 units. A woman from a household where the head of the house is male is 0.83 (OR=0.833, p-value=0.0029) times less likely to experience IPV, compared to a woman from a household where the head of the house is female.

A woman from the poorest wealth index class is 1.40 (OR=1.397, p-value<.0001) times more likely to experience IPV, compared to a woman from the richest wealth index class. A woman from a poorer wealth index class is 1.28 (OR=1.284, p-value=0.0019) times more likely to experience IPV, compared to a woman from the richest wealth index class. A woman from a middle wealth index class is 1.33 (OR=1.325, p-value=0.0002) times more likely to experience IPV, compared to a woman from the richest wealth index class. A woman from a richer wealth index class is 1.24 (OR=1.239, p-value=0.0019) times more likely to experience IPV, compared to a woman from the richest wealth index class.

A woman who has never terminated a pregnancy is 0.86 (OR=0.858, p-value=0.0070) times less likely to experience IPV, compared to a woman who has done so. A married woman is 6.24 (OR=6.240, p-value<.0001) times more likely to experience IPV, compared to a single woman. A woman living with her partner is 5.81 (OR=5.811, p-value<.0001) times more likely to experience IPV, compared to a single woman. A woman who stayed with her partner for 5-9 years is 0.511 (OR=0.511, p-value<.0001) times less likely to experience IPV, compared to a woman who stayed with her partner for 0-4 years.

A woman with a partner who wants more children than she does is 1.18 (OR=1.177, p-value=0.0035) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does. A woman with a partner who wants fewer children compared to her is 1.32 (OR=1.320, p-value=0.0045) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does. A woman who does not know the number of children her partner wants is 1.29 (OR=1.293, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as she does.

Table 3.11: Logistic regression model coefficients, standard errors, and odds ratios (Tanzania)

Parameter	**Estimate**	**Standard error**	**Wald Chi-square**	**P-value**	**Odds ratio**
Intercept	0.9472	0.2430	15.1995	<.0001	2.578
Partner drinks alcohol (ref=Yes)					
No	-0.9695	0.0547	313.7345	<.0001	0.379
Woman's father ever beat her mother (ref=Yes)					
No	-0.5177	0.0492	110.6903	<.0001	0.596
Don't know	-0.1222	0.0862	2.0108	0.1562	0.885

Continued…

Wife-beating attitude (ref=Unacceptable)					
Acceptable	0,4029	0,0471	73,1671	<.0001	1,4962
I don't know	-0,4152	0,1225	11,4834	0,0007	0,6602
Woman's current age	0.0200	0.00620	10.4573	0.0012	1.020
Region (ref=Arusha)					
Dodoma	-0.2886	0.1747	2.7290	0.0985	0.749
Kilimanjaro	-0.7236	0.1784	16.4619	<.0001	0.485
Tanga	-0.5134	0.1645	9.7417	0.0018	0.598
Morogoro	0.00459	0.1751	0.0007	0.9791	1.005
Pwari	0.0400	0.1748	0.0523	0.8191	1.041
Dar Es Salaam	0.0524	0.1495	0.1229	0.7259	1.054
Lindi	0.1387	0.1717	0.6518	0.4195	1.149
Mtwara	0.2283	0.1756	1.6896	0.1937	1.256
Ruvuma	-0.2257	0.1714	1.7336	0.1880	0.798
Iringa	-0.3506	0.1801	3.7912	0.0515	0.704
Mbeya	-0.1714	0.1757	0.9518	0.3293	0.842
Singida	-0.2440	0.1688	2.0895	0.1483	0.784
Tabora	0.1075	0.1595	0.4545	0.5002	1.113
Rukwa	-0.1110	0.1691	0.4310	0.5115	0.895
Kigoma	-0.2369	0.1621	2.1352	0.1439	0.789
Shinyanga	0.1914	0.1609	1.4141	0.2344	1.211
Kagera	0.1055	0.1730	0.3716	0.5421	1.111
Mwanza	0.1650	0.1665	0.9821	0.3217	1.179
Mara	0.3532	0.1607	4.8330	0.0279	1.424
Manyanga	0.2431	0.1686	2.0785	0.1494	1.275
Njobe	-0.6485	0.1750	13.7402	0.0002	0.523
Katavi	-0.2281	0.1659	1.8897	0.1692	0.796
Simiyu	-0.00987	0.1593	0.0038	0.9506	0.990
Geita	0.0510	0.1629	0.0982	0.7540	1.052
Kaskazini Unguja	0.3444	0.1772	3.7785	0.0519	1.411
Kusini Unguja	0.3441	0.1800	3.6533	0.0560	1.411
Mjini Magharibi	0.2741	0.1603	2.9242	0.0873	1.315
Kaskazini Pemba	-0.8904	0.1925	21.4061	<.0001	0.410
Kusini Pemba	-0.4907	0.1869	6.8930	0.0087	0.612
Woman's highest education level (ref=No education)					
Primary education	0.0522	0.0625	0.6968	0.4038	1.054
Secondary education	-0.3373	0.0823	16.8118	<.0001	0.714
Higher education	-0.9905	0.2273	18.9793	<.0001	0.371
Number of household members	-0.1936	0.00749	668.6563	<.0001	0.824

Continued…

Sex of household head (ref=Female)					
Male	-0.1825	0.0612	8.8921	0.0029	0.833
Wealth index (ref=Richest)					
Poorest	0.3343	0.0846	15.6269	<.0001	1.397
Poorer	0.2498	0.0806	9.5988	0.0019	1.284
Middle	0.2817	0.0768	13.4633	0.0002	1.325
Richer	0.2145	0.0689	9.6874	0.0019	1.239
Ever had a terminated pregnancy (ref=Yes)					
No	-0.1531	0.0568	7.2776	0.0070	0.858
The contraceptive method used (ref=Yes)					
No	-1.3795	0.1713	64.8680	<.0001	0.252
Current marital status (ref=Single)					
Married	1.8310	0.0624	862.1317	<.0001	6.240
Living with partner	1.7597	0.0747	554.8580	<.0001	5.811
Number of other partners	0.1368	0.0243	31.6214	<.0001	1.147
Cohabitation duration (ref=0 to 4 years)					
5 to 9 years	-0.6720	0.0745	81.4740	<.0001	0.511
Partner's desire for children (ref=Both want same)					
Partner wants more	0.1631	0.0558	8.5468	0.0035	1.177
Partner wants fewer	0.2779	0.0979	8.0607	0.0045	1.320
Don't know	0.2568	0.0541	22.5016	<.0001	1.293
Woman's occupation (ref=Unemployed)					
Employed	0.3486	0.0599	33.8266	<.0001	1.417
Partner's age	-0.0114	0.00343	11.1476	0.0008	0.989
The person who usually decides how to spend a woman's earnings (ref=Woman alone)					
Woman and partner	-0,2524	0,0501	25,4007	<.0001	0,7769
Partner alone	-0,4221	0,0825	26,1781	<.0001	0,6557
Someone else	-0,8591	1,3558	0,4015	0,5263	0,4235
Interaction effects					
Woman's age by contraceptive use	0.0302	0.00526	32.9608	<.0001	1.031

An employed woman is 1.42 (OR=1.417, p-value<.0001) times more likely to experience IPV, compared to an unemployed woman. A one-year increase in the age of a woman's partner decreases the chances of experiencing IPV by 0.0114 units. A woman who decides with her partner on how to spend her earnings is 0.78 (OR=0.7769, p-value<.0001) times less likely to experience IPV, compared to a woman who decides alone how to spend them. A woman whose partner alone decides how to spend her earnings is 0.66 (OR=0.6557, p-value<.0001) times less likely to experience IPV, compared to a woman who decides alone how to spend them.

Interaction Effects

Figure 3.6 shows that women between the ages of 15 to 45 years who are using contraceptives, are at a higher risk of experiencing IPV, compared to women not using any. From the same figure, there is a slight decrease in IPV for women between the ages of 46 to 50 years. Also, for the same age group, women who are not using contraceptives are at a higher risk of experiencing IPV.

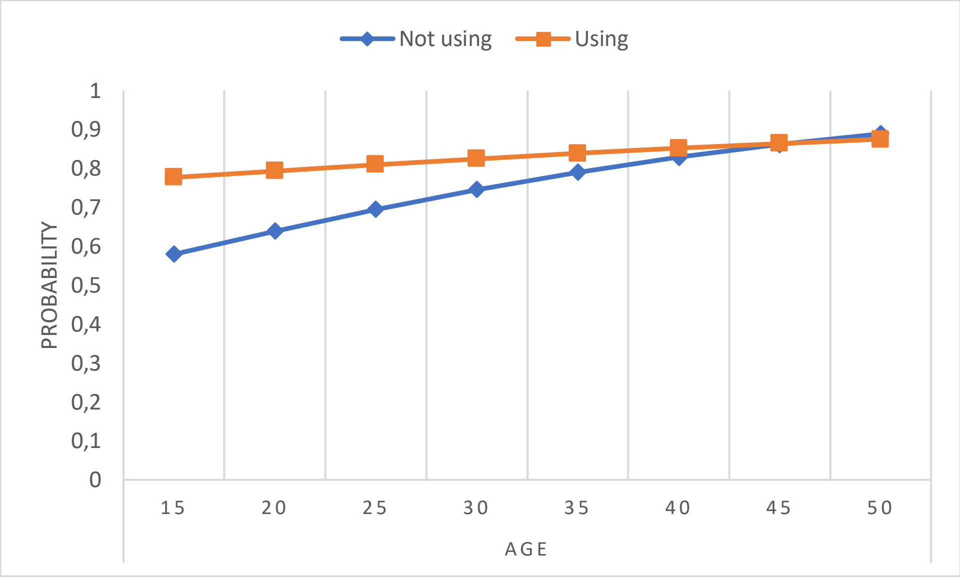

Figure 3.6: Predicted probability of experiencing IPV by woman's age and contraceptive use

Uganda

Model Checking

Multicollinearity was checked for the variables in the model, and the outcome showed no collinearity in the variables fitted in the model. The Chi-square statistic for the Hosmer-Lemeshow test is 140.5064, with 8 degrees of freedom, and a corresponding p-value of <.0001, as shown in Table 3.13. This indicates that there is sufficient evidence to claim that the model does not fit the data adequately. Thus, we can conclude that the model is not a good fit for the data; in other words, predicted probabilities are not the same as the observed values.

Model Selection

Table 3.12 below shows the model fit statistics that are used to compare the two models. The AIC for the full model is small compared to the AIC for the one that only has the intercept, showing that the model is the best fit.

Table 3.12: Model fit statistics

Criterion	Intercept only	Intercept and covariates (full model)
AIC	23272.633	19081.671
SC	23280.459	19480.789
-2 Log L	23270.633	18979.671

Table 3.13: Hosmer and Lemeshow goodness-of-fit test

Chi-square	DF	P-value
140.5064	8	<.0001

Prediction Accuracy of the Model

Table 3.14 shows the relationship between predicted probabilities and observed outcomes, with the area under the curve being $c = 0.779$ and a concordant rate of 77.9, which tells us how good the model is for separating the 0's and 1's, using a chosen model. The value for Gamma is 0.557, which suggests that there is no perfect relationship. It is interpreted as meaning that 55.7% fewer errors are made in prediction by using the estimated probabilities, compared to using a chance alone. One weakness of this statistic is the tendency to overstate the strength of relationship between probabilities and outcomes. The value for Somer's D is 0.557. This shows that not all pairs are concordant, and we may use it to compare models. Figure 3.7 shows the ROC curve of the fitted model and the area under the curve $c = 0.779$, implying that 77.9% of the probabilities are predicted correctly, which is a very good predictive accuracy.

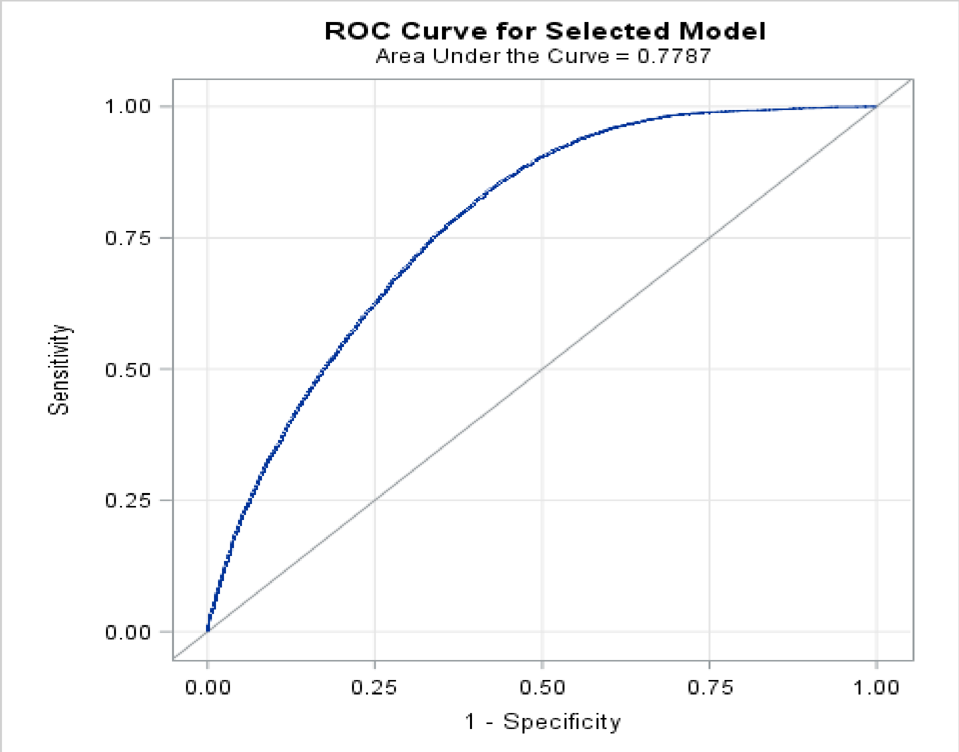

Figure 3.7: ROC curve for selected model

Table 3.14: Association of predicted probabilities and observed responses

Percent concordant	77.9	Somers' D	0.557
Percent discordant	22.1	Gamma	0.557
Percent tied	0.0	Tau-a	0.244
Pairs	74826784	C	0.779

Interpretation of the Results

The covariates used in the model are shown in Table 3.15 with their p-values. A covariate with a p-value of less than 0.05 has a significant association with the response variable.

Table 3.15: Type 3 analysis of effects

Effect	**DF**	**Wald Chi-square**	**P-value**
Partner drinks alcohol	1	520.1294	<.0001
Woman's father ever beat her mother	2	256.9705	<.0001
Wife-beating attitude	2	36.3511	<.0001
Media exposure	2	9.1905	0.0101
Woman's current age	1	69.9349	<.0001
Region	14	109.3158	<.0001

Continued…

Number of household members	1	361.5399	<.0001
Sex of the household head	1	16.5597	<.0001
Wealth index	4	39.2864	<.0001
Ever had a terminated pregnancy	1	10.3764	0.0013
Contraceptive method used	1	52.0373	<.0001
Current marital status	2	654.3645	<.0001
Number of other partners	1	28.5442	<.0001
Cohabitation duration	1	41.8784	<.0001
Partner's desire for children	3	16.8301	0.0008
Partner's education level	4	35.3007	<.0001
Woman's occupation	4	20.1706	0.0005
The person who usually decides how to spend a woman's earnings	3	47.6569	<.0001
Woman's earnings compared to her partner	4	61.4821	<.0001
Interaction effects			
Woman's age by contraceptive use	1	40.7170	<.0001
Wealth index by contraceptive use	4	14.2668	0.0065

***DF=Degrees of Freedom**

Table 3.16 shows that a woman with a partner who does not drink alcohol is 0.43 (OR=0.426, p-value<.0001) times less likely to experience IPV, compared to a woman with a partner who does drinks it. A woman who has never witnessed her father beating her mother is 0.54 (OR=0.535, p-value<.0001) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.22 (OR=1.2169, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A woman who does not know whether wife-beating is acceptable or unacceptable is 0.65 (OR=0.6532, p-value=0.013) times less likely to experience IPV, compared to a woman who views wife-beating as unacceptable.

A woman with high exposure to the media is 0.70 (OR=0.702, p-value=0.0041) times less likely to experience IPV, compared to a woman with low exposure to the media. A one-year increase in a woman's age increases the chances of her experiencing IPV by 0.0102 units. A woman from the Kampala province is 1.30 (OR=1.301, p-value=0.0260) times more likely to experience IPV, compared to one from Kigezi province. A woman from South Buganda province is 1.26 (OR=1.261, p-value=0.0242) times more likely to experience IPV, compared to one from Kigezi province. A woman from the Busoga province is 1.64 (OR=1.636, p-value<.0001) times more likely to experience IPV, compared to one from Kigezi province. A woman from Bukedi province is 1.49 (OR=1.488, p-value=0.0001) times more likely to experience IPV, compared to one from Kigezi province. A woman from West Nile province is 1.36 (OR=1.356, p-value=0.0033) times more likely to experience IPV, compared to one from Kigezi province. A woman from Ankole province is 1.52 (OR=1.523, p-value<.0001) times more likely to experience IPV, compared to one from Kigezi province.

A one-member increase in the number of household members reduces a woman's chances of experiencing IPV by 0.1371 units. A woman from a household where the head is male is 1.130 (OR=1.130, p-value=0.0072) times more likely to experience IPV, compared to a woman from a household where the head is female. A woman who has never had a terminated pregnancy is 0.87 (OR=0.873, p-value=0.0024) times less likely to experience IPV, compared to a woman who has had one. A married woman is 3.56 (OR=3.555, p-value<.0001) times more likely to experience IPV, compared to a single woman. A woman living with her partner is 3.61 (OR=3.606, p-value<.0001) times more likely to experience IPV, compared to a single woman. A unit increase in the number of other wives increases a woman's chances of experiencing IPV by 0.0556 units. A woman who has stayed with her partner for 5-9 years is 0.67 (OR=0.666, p-value<.0001) times less likely to experience IPV, compared to a woman who has stayed with her partner for 0-4 years.

A woman who does not know the number of children her partner wants is 0.85 (OR=0.854, p-value=0.0017) times less likely to experience IPV, compared to a woman who wants the same number of children as her partner. A woman who has a partner with a secondary level of education is 0.76 (OR=0.762, p-value=0.0006) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who has a partner with a higher level of education is 0.60 (OR=0.599, p-value<.0001) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who does not know her partner's level of education is 0.49 (OR=0.491, p-value<.0001) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who has an employed partner is 0.78 (OR=0.775, p-value=0.0089) times less likely to experience IPV compared to a woman who has an unemployed partner.

Table 3.16: Logistic regression model coefficients, standard errors, and odds ratios (Uganda)

Parameter	Estimate	Standard error	Wald Chi-square	P-value	Odds ratio
Intercept	-0,04	0.1953	0.0419	0.8379	0,9608
Partner drinks alcohol (ref=Yes)					
No	-0.8544	0.0380	506.4011	<.0001	0.426
Woman's father ever beat her mother (ref=Yes)					
No	-0.6257	0.0384	264.9138	<.0001	0.535
Don't know	-0.0583	0.0937	0.3874	0.5337	0.943
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,1963	0,0379	26,7825	<.0001	1,2169
I don't know	-0,4259	0,1714	6,1756	0,013	0,6532

Continued…

Access to the media (ref=Low exposure)					
Medium exposure	-0.0384	0.0392	0.9563	0.3281	0.962
High exposure	-0.3541	0.1232	8.2557	0.0041	0.702
Woman's current age	0.0102	0.00460	4.9579	0.0260	1.010
Region (ref=South Buganda)					
Kampala	0.2631	0.1145	5.2850	0.0215	1.301
North Buganda	0.1532	0.1020	2.2528	0.1334	1.166
Busoga	0.2320	0.1029	5.0827	0.0242	1.261
Bukedi	0.4920	0.1002	24.1012	<.0001	1.636
Bugisu	0.3977	0.1044	14.5162	0.0001	1.488
Teso	0.0890	0.1105	0.6483	0.4207	1.093
Karamoja	0.1592	0.1046	2.3168	0.1280	1.173
Lango	-0.1621	0.1232	1.7330	0.1880	0.850
Acholi	0.00939	0.1039	0.0082	0.9280	1.009
West Nile	-0.0297	0.1080	0.0757	0.7833	0.971
Bunyoro	0.3047	0.1036	8.6451	0.0033	1.356
Tooro	-0.1567	0.1089	2.0708	0.1501	0.855
Ankole	0.1545	0.1028	2.2586	0.1329	1.167
Kigezi	0.4206	0.1007	17.4406	<.0001	1.523
Number of household members	-0.1371	0.00737	345.8402	<.0001	0.872
Sex of household head (ref=Female)					
Male	0.1225	0.0456	7.2246	0.0072	1.130
Ever had a terminated pregnancy (ref=Yes)					
No	-0.1353	0.0445	9.2362	0.0024	0.873
The contraceptive method used (ref=Yes)					
No	-1.1260	0.1422	62.6822	<.0001	0.324
Current marital status (ref=Single)					
Married	1.2682	0.0549	532.9146	<.0001	3.555
Living with partner	1.2826	0.0532	582.1179	<.0001	3.606
Number of other partners	0.0556	0.0105	27.8114	<.0001	1.057
Cohabitation duration (ref=0 to 4 years)					
5 to 9 years	-0.4071	0.0631	41.6643	<.0001	0.666
Partner's desire for children (ref=Both want same)					
Partner wants more	0.0243	0.0432	0.3164	0.5738	1.025
Partner wants fewer	-0.0854	0.0645	1.7529	0.1855	0.918
Don't know	-0.1582	0.0505	9.7999	0.0017	0.854
Partner's education level					
Primary education	-0.1251	0.0737	2.8814	0.0896	0.882
Secondary education	-0.2713	0.0788	11.8373	0.0006	0.762
Higher education	-0.5130	0.0927	30.6050	<.0001	0.599
Don't know	-0.7123	0.1366	27.1687	<.0001	0.491
Partner's occupation (ref=Unemployed)					
Employed	-0.2553	0.0976	6.8397	0.0089	0.775

Continued…

The person who usually decides what to do with woman's earnings(ref=Woman)					
Woman and partner	-0,2161	0,0402	28,8991	<.0001	0,8057
Partner alone	-0,3903	0,0695	31,501	<.0001	0,6769
Someone else	-0,1801	0,4899	0,1351	0,7132	0,8352
Woman's earnings compared to her partner (ref=Less compared to him)					
More compared to him	0.2129	0.0656	10.5237	0.0012	1.237
About the same	-0.4263	0.0492	74.9519	<.0001	0.653
Partner does not bring in anything	-0.2840	0.1321	4.6194	0.0316	0.753
Don't know	-0.5357	0.0972	30.3507	<.0001	0.585
Interaction effects					
Woman's age by contraceptive use	0.0328	0.00451	52.9710	<.0001	1.033
Wealth index by contraceptive use					
Poorest by not using contraceptives	-0.2659	0.1237	4.6178	0.0316	0.767
Poorer by not using contraceptives	0.0742	0.1184	0.3933	0.5306	1.077
Middle by not using contraceptives	0.1583	0.1190	1.7714	0.1832	1.172
Richer by not using contraceptives	0.1163	0.1185	0.9630	0.3264	1.123

A woman who earns more compared to her partner is 1.24 (OR=1.237, p-value=0.0012) times more likely to experience IPV, compared to a woman who earns less than him. A woman who earns about the same as her partner is 0.65 (OR=0.653, p-value<.0001) times less likely to experience IPV, compared to a woman who earns less than him. A woman who has a partner who does not bring in earnings is 0.32 (OR=0.316, p-value=0.0316) times less likely to experience IPV, compared to a woman who earns less than him. A woman who decides with her partner how to spend her earnings is 0.81 (OR=0.8057, p-value<.0001) times less likely to experience IPV, compared to a woman who decides alone how to spend them. A woman whose partner alone decides how to spend her earnings is 0.68 (OR=0.68, p-value<.0001) times less likely to experience IPV, compared to a woman who decides alone how to spend them.

Interaction Effects

Figure 3. 8 shows that IPV increases with increasing age, whether a woman is using contraceptives or not. We observe from the same figure that IPV is higher among women using contraceptives aged between 15-35 years, compared to women of the same age group not using contraceptives. Women who are 35 years old have an equal chance of experiencing IPV, whether using contraceptives or not. From about 37 years of age, the IPV for women not using contraceptives is higher compared to those using contraceptives.

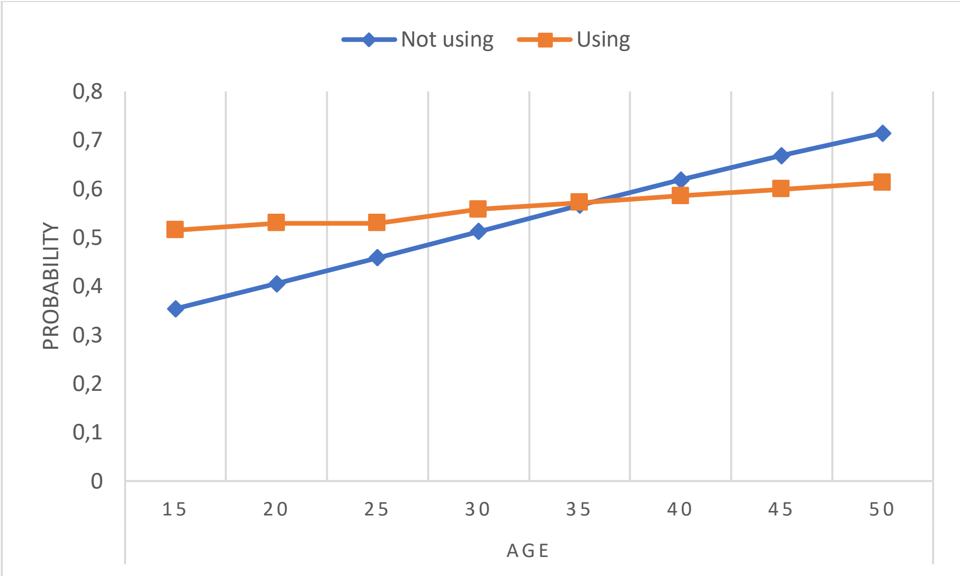

Figure 3. 8: Predicted probability of experiencing IPV by woman's age and contraceptive use

3.7 Summary

In the section above, based on the 2016 South Africa Demographic and Health Survey, 2015-2016 Tanzania Demographic and Health Survey, and the 2016 Uganda Demographic and Health Survey data, we used logistic regression to identify the critical determinants of intimate partner violence amongst women of reproductive age in these three countries. The findings of this study show that in all three different countries, some determinants are common: partner's alcohol drinking status, whether the woman's father ever beat her mother, wife-beating attitudes, the current age of the woman, the region where the woman resides, the number of household members, the sex of the household head, the wealth index, the pregnancy termination status, the contraceptive method used, current marital status, cohabitation duration, and a woman's occupation.

The following countries have certain determinants in common. South Africa and Tanzania: partner's age. South Africa and Uganda: media exposure, partner's desire for children, and a woman's earnings compared to her partner. Tanzania and Uganda: number of other partners and the person who usually decides how to spend a woman's earnings. In Uganda, the partner's education level had a significant influence on intimate partner violence. In South Africa, body mass index and the knowledge of STIs were significant to intimate partner violence. In Tanzania, a woman's education level was significant in determining intimate partner violence. In the next section, the logistic regression is extended into survey logistic regression so as to investigate the variability within clusters or strata.

3.8 Survey Logistic Regression Model

Several statistical analyses assume that most of the data being analyzed is from a finite population by simple random sampling, in which each unit in the population has the same chance of being selected

(Moeti, 2010; Dlamini, 2016). In real life, survey data is collected from a finite population, and the stratification of the population by some variable of interest is performed. The latter ensures a balance in the number of respondents for each variable (An, 2002). Survey logistic regression considers the complexity of the survey design (Moeti, 2010; Dlamini, 2016); a valid statistical inference can be made utilizing survey logistic regression, which considers the different components of the survey design (An, 2002; Moeti, 2010; Dlamini, 2016).

If survey complexity is ignored when modeling, there may be an overestimation or underestimation of the standard errors, leading to narrow or wider confidence intervals (Lemeshow & Hosmer, 2000). The main advantage of stratification is that the survey is easier to carry out, and the parameters can be estimated for each stratum in which they can be significant (Lemeshow & Hosmer, 2000). The population division into strata can possibly reduce the estimator's variance of the population total (An, 2002; Lemeshow & Hosmer, 2000).

3.8.1 Parameter Estimation

If the design of a survey is complex, then assumptions for independence are not applicable (Dlamini, 2016); in a cluster, the correlation between observations may be evident. The standard errors related to the model coefficients need to be estimated; to do so, the complex part of the design needs to be taken into account (Dlamini, 2016). The data considered the primary sampling units that were sampled in the first stage, per stratum. In stage two, households were sampled (Moeti, 2010; Dlamini, 2016). The response variable is specified as $y_{lmvz}(l = 1,2, \dots, H_{zvm}; o = 1,2, \dots, m_z; z = 1,2, \dots, K)$, which is one of the events that occurred in l^{th} individual within the m^{th} element, in the v^{th} primary sample units within the z^{th} stratum, and 0 otherwise (Dlamini, 2016). The total observations are given by $n = \sum_{z=1}^{K} \sum_{v=1}^{m_z} n_{zv}$ and the sampling design weight for the $zvml^{th}$ is given in the dataset, which is denoted by w_{zvml} (Dlamini, 2016).

Let the probability that an event was observed in the l^{th} individual, within m household in the v^{th} primary sample units, nested within z^{th} stratum, be $\pi_{zvml} = P(y_{lmvz} = 1)$, and the probability that the event did not occur in the l^{th} individual, within the m^{th} household within the v^{th} primary sample units, nested within z^{th} stratum, be $1 - \pi_{zvml} = P(y_{lmjz} = 0)$ (Moeti, 2010; Dlamini, 2016). The pseudo-maximum likelihood is constructed as the product of individual contributions to the likelihood estimators (Lemeshow & Hosmer, 2000; Dlamini, 2016). The contribution of a single value utilizing pseudo-maximum likelihood, is given by $\pi_{zvml}^{w_{zvml}y_{zvml}}(1 - \pi_{zvml})^{(1-w_{zvml}y_{zvml})}$ (Dlamini, 2016).

Thus, the pseudo-likelihood function is given by:

$$L(\beta; \boldsymbol{Y}) = \prod_{z=1}^{K} \prod_{v=1}^{m_z} \prod_{m=1}^{n_{zv}} \prod_{l=1}^{H_{zvm}} \pi_{zvml}^{w_{zvml}y_{zvml}}(1 - \pi_{zvml})^{(1-w_{zvml}y_{zvml})} \tag{3.8.1}$$

The pseudo-log-likelihood function is given by:

$$l(\beta; \boldsymbol{Y}) = \sum_{z=1}^{K}\sum_{v=1}^{m_z}\sum_{m=1}^{n_{zv}}\sum_{l=1}^{H_{zvm}}\left\{w_{zvml}y_{zvml}log\left(\frac{\pi_{zvml}}{1-\pi_{zvml}}\right) - log\left(\frac{1}{1-\pi_{zvml}}\right)\right\} \tag{3.8.2}$$

Differentiating the log-likelihood concerning unknown regression coefficients, we get the vector of $p + 1$ score equations, which are compactly written as:

$$\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{\pi}) = 0 \tag{3.8.3}$$

where $\boldsymbol{X}$ is the $n \times (p + 1)$ matrix of covariate values, $\boldsymbol{W}$ is an $n \times n$ diagonal matrix containing weights, $\boldsymbol{y}$ is the $n \times 1$ vector of observed outcome values, and $\boldsymbol{\pi} = [\pi_{1111}, \dots, \pi_{zm_zn_{zv}H_{zvm}}]'$ is the $n \times 1$ vector of logistic probabilities (Dlamini, 2016). The survey logistic regression model is given by:

$$logit(\pi_{zvml}) = log\left\{\frac{\pi_{zvml}}{1-\pi_{zvml}}\right\} = \boldsymbol{X}'_{zvml}\boldsymbol{\beta} \tag{3.8.4}$$

where, $\boldsymbol{X}_{zvml}$ is the vector that corresponds to the characteristics of the l^{th} individual within the m^{th} household within the v^{th} primary sample units, nested within the z^{th} stratum and $\boldsymbol{\beta}$ is the vector of unknown model coefficients (Dlamini, 2016).

3.9 Survey Logistic Model Selection and Checking

3.9.1 Model Selection

In SAS (the survey logistic regression procedure), there are no variable selection procedures implemented yet (Dlamini, 2016); but it is possible to add such conditions manually, or take out one variable at a time in the model, by utilizing the type 3 analysis of effects, and observing the effect of the remaining variables (Dlamini, 2016). Type 3 analysis of effects is often used when the effect of one explanatory variable influences the effect of another explanatory variable.

Testing Hypothesis about β

Computing standard errors of the parameter estimates, that are utilized in constructing the confidence intervals and performing statistical tests, is a complex algorithm, if the data is from a complex survey data (Moeti, 2010; Dlamini, 2016). The estimation of the covariance matrix of the estimator of coefficients is given by:

$$\widehat{Var(\hat{\beta})} = (\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X})^{-1}\boldsymbol{S}(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X})^{-1} \tag{3.9.1}$$

where $\boldsymbol{D} = \boldsymbol{W}\boldsymbol{X}$ is the $n \times n$ diagonal matrix with general components:

$$w_{zvml}\pi_{lvml}(1 - \pi_{zvml}).$$

Matrix $\boldsymbol{S}$ is a pooled, within-stratum estimator of the covariance matrix on the left side of the equation (3.8.3) (Dlamini, 2016). Let the general component of the vector of the score equation be denoted as $Z'_{zvml} = w_{zvml}\pi_{zvml}(1-\pi_{zvml})$ (Dlamini, 2016). Thus:

$$z_{zv} = \sum_{m=1}^{n_{zv}} z_{zvml}. \qquad (3.9.2)$$

The stratum-specific average is given by:

$$\bar{z_z} = \frac{1}{m_z}\sum_{v=1}^{m_z} z_{zv}.$$

The within-stratum estimator for the z^{th} stratum variance is given by:

$$S_z = \frac{m_z}{M_z}\sum_{v=1}^{m_z}(z_{zv}-\bar{z}_z)(z_{zv}-\bar{z}_z)'. \qquad (3.9.3)$$

$S = \sum_{z=1}^{K}(1-f_z)S_z.(1-f_z)$, is the pooled estimator of the finite population correction factor, and $f_z = \frac{m_z}{M_z}$ is the ratio of the number of sampling units to the total number of primary sampling units in the stratum z (Dlamini, 2016). In general, if M_z is unknown, we can presume that M_z is sufficiently large so that f_z approaches zero, thus making the correction factor be one (Lemeshow & Hosmer, 2000; Dlamini, 2016). The Wald statistic for testing if all coefficients in the fitted model are equal to zero, is given by:

$$Wald = \hat{\beta}'[\widehat{var(\hat{\boldsymbol{\beta}})}_{p\times p}]^{-1}\hat{\boldsymbol{\beta}} \qquad (3.9.4)$$

where $\hat{\boldsymbol{\beta}}$ is the vector of p slope coefficients, and $\widehat{var(\hat{\boldsymbol{\beta}})}_{p\times p}$ is the sub-matrix from a $(p+1)(p+1)$ matrix of $\widehat{var(\hat{\boldsymbol{\beta}})}$, and the p-value can be computed using χ^2 distribution, with p degrees of freedom, implying:

$p - value = P(\chi^2(p) \geq wald)$ (Dlamini, 2016).

The SAS procedure PROC SURVEYLOGISTIC produces the covariance of parameters, through the Taylor expansion approximation procedure (Vittinghoff, et al., 2011; Heeringa, et al., 2010; Moeti, 2010). This procedure estimates the variance between clusters, and calculates the overall variance through pooling the stratum variance together (Dlamini, 2016; Heeringa, et al., 2010; Moeti, 2010). The Wald statistic is used for testing and constructing the confidence intervals given by:

$$\hat{\beta}_v \pm Z_{1-\frac{\alpha}{2}}\sqrt{V_v} \qquad (3.9.5)$$

where α is the significance level, $Z_{1-\frac{\alpha}{2}}$ is the $100(1-\frac{\alpha}{2})$ percentile of the standard normal distribution, and V_v is the variance resulting from the diagonal of the variance-covariance matrix (Dlamini, 2016). The PROC SURVEYLOGISTIC uses both Taylor expansion (linearization method) and maximum likelihood (Vittinghoff, et al., 2011; Dlamini, 2016). The procedure has been used in order to construct a logistic regression model that accounts for the complex nature of the survey (Dlamini, 2016; Moeti, 2010). The odds ratio is still obtained, as described earlier in this chapter.

3.9.2 Model Checking

Model Fit Test

The Hosmer-Lemeshow statistic is not produced in PROC SURVEYLOGISTIC. Since this statistic is not yet available, however, the AIC can be used to compare the goodness-of-fit of two nested models (Moeti, 2010). The goodness-of-fit test for logistic regression that is applied to complex survey data, is obtained as follows: once we fit the logistic regression model, the residuals $\hat{r}_{vm} = y_{vm} - \hat{\pi}(x_{vm})$ can be obtained (Dlamini, 2016). The goodness of fit test is based on the residuals because of the significant difference between the observed and estimated values, indicating a lack of fit (Hosmer & Lemeshow, 2004; Moeti, 2010; Dlamini, 2016). If a grouping strategy is used, the observed values are sorted into deciles based on their weight and estimated residuals (Dlamini, 2016). The survey estimates of sum of the residual by decile of risk $\widehat{\boldsymbol{T}} = (\hat{T}_1, \hat{T}_2, \dots, \hat{T}_{10})$, are obtained such that $\hat{T}_g = \sum_v \sum_m \overline{w}\, \hat{r}_{vm} (g = 1, \dots, 10)$ (Dlamini, 2016). The related estimated variance-covariance matrix $\widehat{\boldsymbol{V}}(\widehat{\boldsymbol{T}})$ is obtained using linearization (Dlamini, 2016).

The linearization method can be used to approximate the functional form of the estimated population characteristics (Dlamini, 2016). In step one, the functional form of the estimated population characteristics is approximated by a 1^{st} order Taylor series (Dlamini, 2016). The result is an approximation that is linear in the sample observation (Lemeshow & Hosmer, 2000). The design-based methods are used to estimate its variance (Lemeshow & Hosmer, 2000; Dlamini, 2016). Using the method outlined, the F-adjusted can be estimated as

$$F_{adjusted} = \frac{f - g + 2}{f_g} \widehat{\boldsymbol{T}} \boldsymbol{V}(\widehat{\boldsymbol{T}})^{-1} \widehat{\boldsymbol{T}} \qquad (3.9.\ 6)$$

where f is the sampled clusters minus the strata, and g is the groups (Dlamini, 2016). The assumption is that the covariance is zero (Dlamini, 2016).

Predictive Accuracy/Ability of the Model

To assess for the predictive accuracy, SAS procedure PROC SURVEYLOGISTIC produces other statistics, namely, Somer's D, Gamma, c, and Tau-a (Dlamini, 2016). The statistics above range between 0 and 1. Larger values corresponds to a stronger relationship between predicted and observed values (Dlamini, 2016).

3.10 Design Effects (DEFF)

3.10.1 Background

The sample size and design determine how precise the estimated parameters are (Dlamini, 2016). Due to the impracticality of the simple random sampling (SRS) technique, it is not recommended in a large-scale survey (Shackman, 2001; Dlamini, 2016). In a national survey, a complex design is adopted. The problem with complex design is that sampling errors for the estimates cannot be computed easily (Shackman, 2001; Dlamini, 2016). The loss of effectiveness in using complex design rather than using SRS is known as design effects (Dlamini, 2016). The design effect is calculated by using the actual variance, found in the sampling method usually used, divided by the variance found under the assumption of SRS (Shackman, 2001; Dlamini, 2016). The technique widely used in survey sampling for planning a sample design in estimation and analysis, is called the design effect (Park & Lee, 2001; Dlamini, 2016).

In PROC SURVEYLOGISTIC, we can use DEFF in the model statement (Dlamini, 2016). The design effect is calculated for the regression coefficients (Dlamini, 2016). The design effect is given by:

$$DEFF = \frac{variance\ under\ complex\ design}{variance\ under\ the\ simple\ random\ sampling}. \qquad (3.10.1)$$

The denominator of equation (3.10.*1* is found under the assumption that the design follows an SRS, where we do not account for clustering, stratification, or weighting (Dlamini, 2016). If the sampling weights and population total are considered for analysis, the sampling rates (or population total) under the assumption of SRS are given by (Dlamini, 2016):

$$f_{srs} = \frac{n}{w}$$

where n is the sample size, and w estimates the population size (Dlamini, 2016). If the estimated population size is smaller than the sample size, then f_{srs} is set to zero (Dlamini, 2016).

3.10.2 Design Effect Interpretation

Design effects (DEFF) may be used in comparing variance under the assumption that data was collected using SRS in a complex survey (Dlamini, 2016). We can also use DEFT, which is the square root of DEFF (Dlamini, 2016). DEFT is used to minimize variability, since it is less variable compared to DEFF (Dlamini, 2016). Meanwhile, DEFF can be used to estimate the confidence interval directly (Shackman, 2001; Dlamini, 2016; Moeti, 2010). The DEFT reveals how the standard errors and confidence intervals increase (Dlamini, 2016). If DEFT is equal to *k*, then we say that the confidence interval must be *k* times as large as it would be for SRS (Park & Lee, 2001; Dlamini, 2016).

3.11 Limitations of Survey Logistic Regression

The survey logistic may present some limitations due to the unavailability of the Hosmer-Lemeshow test (Lemeshow & Hosmer, 2000; Dlamini, 2016). It makes it difficult to test if the model is a good fit or not. The variable selection is made manually, and it is time-consuming when there are many variables to work with, which may result in multiple errors. The model must be chosen based on the AIC or SC; then a penalty must be introduced to the -2log-likelihood (-2logL) of having many parameters (Lemeshow & Hosmer, 2000). Since these parameters have -2logL terms in their formulation, they are only used in the case of ungrouped data (Lemeshow & Hosmer, 2000; Dlamini, 2016).

3.12 Comparison of Logistic and Survey Logistic Regression

The selected individuals were not from SRS, and the parameter estimates for both models are not similar (Dlamini, 2016). One assumption for the logistic regression is that observations are independent, but for complex survey design, the assumption is violated (Dlamini, 2016); hence the better model may be the one fitted using PROC SURVEYLOGISTIC. The PROC SURVEYLOGISTIC considers the complexity of the design.

3.13 Fitting the Survey Logistic Regression Model

Multiple logistic regression was fitted for the 2015-2016 DHS Tanzania survey, 2016 South Africa DHS survey, and the 2016 Uganda DHS survey, using SAS. PROC SURVEYLOGISTIC was considered for this study to estimate the parameters, standard errors, and odds ratios.

South Africa

Model Checking

The AIC or SC can be used to determine if the model is a good fit or not. Table 3.17 shows that the AIC (8510.305) of the full model is smaller compared to the AIC (9922.214) of the reduced model; this indicates that the fitted model better explains the data.

Table 3.17: Model fit statistics

Criterion	Intercept only	Intercept and covariates (full model)
AIC	9922.214	8510.305
SC	9929.263	8919.174
-2logL	9920.214	8394.305

The type 3 analysis in Table 3.18 shows the variables that have been fitted into the model, with the variables that are significant with a p-value of less than 0.05. These are the variables that are associated with the response variable IPV (Intimate Partner Violence).

Table 3.18: Type 3 analysis effects

Type 3 Analysis of Effects			
Effect	**F-value**	**DF**	**P-value**
Partner drinks alcohol	111.41	2	<.0001
Woman's father ever beat her mother	47.31	2	<.0001
Access to the media	6.41	2	0.0017
Wife-beating attitude	24.43	2	<.0001
Woman's current age	33.68	1	<.0001
Region	5.19	8	<.0001
Type of residence	0.19	1	0.6661
Woman's education level	2.70	3	0.0449
Number of household members	191.19	1	<.0001
Sex of the household head	11.52	1	0.0007
Literacy	0.44	1	0.5092
Wealth index	15.62	4	<.0001
Ever had a terminated pregnancy	18.95	1	<.0001
Contraceptive method used	53.42	1	<.0001
Body Mass Index	10.46	1	0.0013
Current marital status	33.56	2	<.0001
Number of other wives	0.27	1	0.6015
Cohabitation duration	4.26	1	0.0393
Partner's desire for children	15.62	3	<.0001
Partner's education level	0.50	4	0.7364
Partner's occupation	0.24	1	0.6229
Woman's occupation	7.42	2	0.0007
Partner's age	8.74	1	0.0032
The person who usually decides how to spend a woman's earnings	0.48	1	0.4884
Woman's earnings compared to partner	3.30	4	0.0107
Knowledge of Sexually Transmitted Infection (STI)	13.16	1	0.0003
Interaction effects			
Woman's age by contraceptive use	29.10	1	<.0001
Wealth index by contraceptive use	2.78	4	0.0258

***DF=Degrees of Freedom**

Prediction Accuracy of the Model

The concordant rate was 76.1%, as shown in Table 3.19 below; this value tells how good the model was in separating 0's and 1's. The value c=0.762 is the area under the ROC curve. The area under the curve (ROC) of 0.762 implies that 76.2% of the probabilities were predicted correctly by the model, and hence the model is said to have a good prediction accuracy. The Gamma statistic has a value of 0.526, and indicates a moderate positive association between variables. The Somers' D statistic is 0.525, showing that the model is a good predictor, and that there is some association between the variables.

Table 3.19: Association of predicted probabilities and observed responses (South Africa)

Percent concordant	76.1	Somers' D	0.525
Percent discordant	23.6	Gamma	0.526
Percent tied	0.3	Tau-a	0.207
Pairs	14264753	C	0.762

Table 3.20 shows that a woman whose partner does not drink alcohol is 0.44 (OR=0.440, p-value<.0001) times less likely to experience IPV, compared to a woman whose partner does drinks it. A woman who has never witnessed her father beating her mother is 0.44 (OR=0.439, p-value<.0001) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who does not know if her father beats her mother is 0.73 (OR=0.725, p-value=0.0200) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.80 (OR=1.797, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A woman who does not know whether wife-beating is acceptable or unacceptable is 0.32 (OR=0.32, p-value<.0001) times less likely to experience IPV, compared to a woman who views wife-beating as unacceptable.

A woman with medium exposure to the media is 1.34 (OR=1.336, p-value=0.0007) times more likely to experience IPV, compared to a woman with low exposure to the media. A woman with high exposure to the media is 1.25 (OR=1.254, p-value=0.0283) times more likely to experience IPV, compared to a woman with low exposure to the media. A unit increase in the woman's age increases the chances of her experiencing IPV by 0.0122 units.

A woman from the Western Cape province is 0.65 (OR=0.645, p-value=0.0016) times less likely to experience IPV, compared to a woman from the Eastern Cape province. A woman from the Northern Cape region is 0.53 (OR=0.530, p-value<.0001) times less likely to experience IPV, compared to a woman from the Eastern Cape region. A one-member increase in the number of household members decreases a woman's chances of experiencing IPV by 0.1607 units. A woman from a household where the head of the house is male is 1.24 (OR=1.240, p-value=0.0006) times more likely to experience IPV, compared to a woman from a household where the head of the house is female.

A woman who has never terminated a pregnancy is 0.66 (OR=0.660, p-value<.0001) times less likely to experience IPV, compared to a woman who has terminated a pregnancy. A unit increase in the woman's body mass index decreases a woman's chances of experiencing IPV by 0.0022 units. A married woman is 1.70 (OR=1.704, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 1.81 (OR=1.805, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman who stayed with her partner for 5-9 years is

0.80 (OR=0.802, p-value=0.0427) times less likely to experience IPV, compared to a woman who stayed with her partner for 0-4 years. A woman with a partner who wants more children compared to her is 1.36 (OR=1.356, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as her. A woman who does not know the number of children her partner wants is 1.36 (OR=1.356, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as her.

A woman who has an employed partner is 0.84 (OR=0.842, p-value=0.0268) times less likely to experience IPV, compared to a woman who does not know if her partner is employed or not. An employed woman is 1.30 (OR=1.295, p-value<.0001) times most likely to experience IPV compared to an unemployed woman. A one-year increase in the age of a woman's partner decreases the chances of experiencing IPV by 0.0123 units. A woman who earns about the same as her partner is 1.19 (OR=1.186, p-value=0.0262) times more likely to experience IPV, compared to a woman who earns than her partner. A woman whose partner does not bring in earnings is 1.41 (OR=1.411, p-value=0.0006) times more likely to experience IPV, compared to a woman who earns less than her partner.

Table 3.20: Survey logistic regression model coefficients, standard errors, and odds ratios (South Africa)

Parameter	**Estimate**	**Standard error**	**t-value**	**P-value**	**Odds ratio**
Intercept	-0,0446	0.2857	-0,11	0,9118	0,9564
Partner drinks alcohol (ref=Yes)					
No	-0.8209	0.0572	-14.36	<.0001	0.440
Don't know	0.6884	0.5897	1.17	0.2435	1.991
Woman's father ever beat her mother (ref=Yes)					
No	-0.8233	0.0827	-9.96	<.0001	0.439
Don't know	-0.3209	0.1376	-2.33	0.0200	0.725
Access to the media (ref=Low exposure)					
Medium exposure	0.2897	0.0851	3.41	0.0007	1.336
High exposure	0.2262	0.1029	2.20	0.0283	1.254
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0.5864	0.1056	5.55	<.0001	1.797
I don't know	-1.1440	0.2797	-4.09	<.0001	0.319
Woman's current age	0.0122	0.00621	1.96	0.0499	1.012
Region (ref=Eastern Cape)					
Western Cape	-0.4393	0.1389	-3.16	0.0016	0.645
Northern Cape	-0.6347	0.1575	-4.03	<.0001	0.530
Free State	0.0550	0.1277	0.43	0.6668	1.057
Kwazulu-Natal	0.0336	0.1193	0.28	0.7783	1.034

Continued…

North West	-0.0256	0.1312	-0.20	0.8452	0.975
Gauteng	0.00839	0.1280	0.07	0.9478	1.008
Mpumalanga	0.0102	0.1081	0.09	0.9250	1.010
Limpopo	0.1858	0.1090	1.70	0.0888	1.204
Number of household members	-0.1607	0.0114	-14.07	<.0001	0.852
Sex of household head (ref=Female)					
Male	0.2149	0.0627	3.43	0.0006	1.240
Wealth index (ref=Richest)					
Poorest	0.6168	0.1568	3.93	<.0001	1.853
Poorer	0.7562	0.1478	5.12	<.0001	2.130
Middle	0.5239	0.1456	3.60	0.0003	1.689
Richer	0.4569	0.1413	3.23	0.0013	1.579
Ever had a terminated pregnancy (ref=Yes)					
No	-0.4158	0.0893	-4.66	<.0001	0.660
The contraceptive method used (ref=Yes)					
No	-1.6123	0.2623	-6.15	<.0001	0.199
Body Mass Index	-0.00220	0.00102	-2.16	0.0309	0.998
Current marital status (ref=Single)					
Married	0.5331	0.0787	6.78	<.0001	1.704
Living with partner	0.5908	0.0859	6.88	<.0001	1.805
Cohabitation duration (ref=0 to 4 years)					
5 to 9 years	-0.2212	0.1089	-2.03	0.0427	0.802
Partner's desire for children (ref=Both want same)					
Husband wants more	0.3046	0.0724	4.21	<.0001	1.356
Husband wants fewer	0.2540	0.1307	1.94	0.0523	1.289
Don't know	0.3093	0.0673	4.60	<.0001	1.362
Partner's occupation (ref=Don't know)					
Employed	-0.1725	0.0777	-2.22	0.0268	0.842
Woman's occupation (ref=Unemployed)					
Employed	0.2582	0.0622	4.15	<.0001	1.295
Don't know	0.0724	0.1378	0.53	0.5993	1.075
Partner's age	-0.0123	0.00458	-2.68	0.0076	0.988
Woman's earnings compared to partner (ref=Less compared to him)					
More compared to him	0.0220	0.0778	0.28	0.7773	1.022
About the same	0.1709	0.0764	2.24	0.0256	1.186
Partner does not bring in earnings	0.3442	0.0997	3.45	0.0006	1.411
Don't know	-0.0847	0.1670	-0.51	0.6124	0.919
Ever heard of Sexually Transmitted Infections (STIs) (ref=Yes)					
No	-0.8482	0.2218	-3.82	0.0001	0.428

Continued…

Interaction effects					
Woman's age by contraceptive use	0.0285	0.00563	5.07	<.0001	1.029
Wealth index (ref=Richest) by contraceptive use (ref=Not using)					
Poorest by not using contraceptives	0.4954	0.2074	2.39	0.0172	1.641
Poorer by not using contraceptives	0.2314	0.1997	1.16	0.2470	1.260
Middle by not using contraceptives	0.5496	0.2036	2.70	0.0071	1.733
Richer by not using contraceptives	0.2709	0.2123	1.28	0.2022	1.311

A woman who has never heard of STIs is 0.43 (OR=0.428, p-value=0.0001) times less likely to experience IPV, compared to a woman who has heard of them.

Interaction Effects

Figure 3.9 shows that IPV increases with increasing age, whether a woman is using contraceptives or not. We observe from the same figure that IPV is higher among women using contraceptives compared to women not using contraceptives.

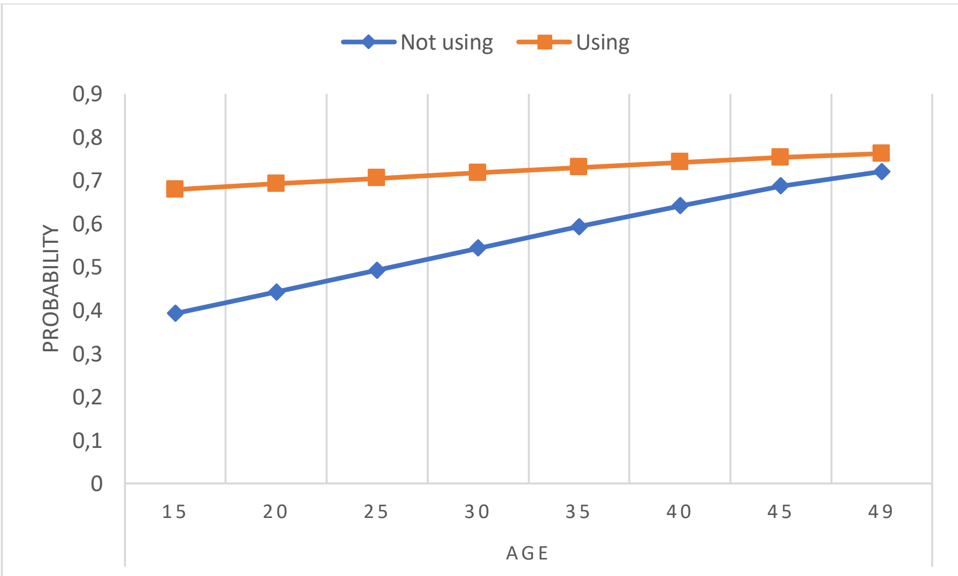

Figure 3.9: Predicted probability of experiencing IPV by woman's age and contraceptive use

Figure 3.10 shows that IPV decreases for women from the poorest to a poorer wealth index class. For a woman who is not using contraceptives, it increases from a poorer to a middle wealth index class, and decreases from a middle to a richer wealth index class. However, for a woman using contraceptives, we observe from the same figure that IPV increases for a woman from the poorest to a poorer wealth index class, and decreases for a woman from a poorer, middle, and richer wealth index class.

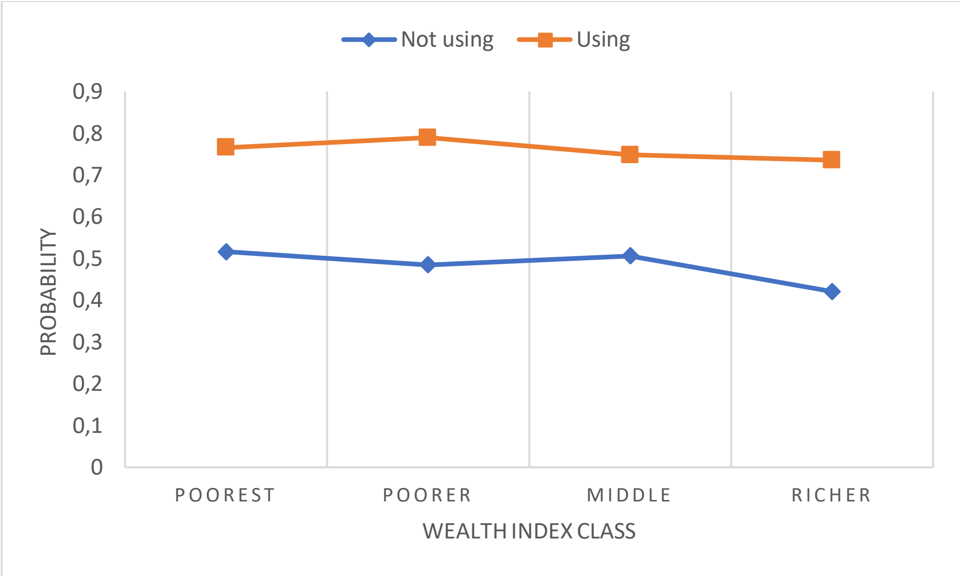

Figure 3.10: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Tanzania

Model Checking

Table 3.21 shows the AIC (13283.599) of the full model is smaller compared to the AIC (18243.286) of the reduced model; this indicates that the fitted model better explains the data.

Table 3.21: Model fit statistics

Criterion	Intercept only	Intercept and covariates (full model)
AIC	18243.286	13283.599
SC	18250.778	13868.050
2 Lo0g L	18241.286	13127.599

In Table 3.22, we see that some of the variables fitted in the model are significant at the 5% level, and some are not.

Table 3.22: Type 3 analysis effects

Type 3 Analysis of Effects			
Effect	**F-value**	**DF**	**P-value**
Partner drinks alcohol	236.44	1	<.0001
Woman's father ever beat her mother	55.69	2	<.0001
Access to the media	0.84	2	0.4314
Wife-beating attitude	55.12	2	<.0001
Woman's current age	45.97	1	<.0001
Region	4.34	29	<.0001
Type of residence	1.74	1	0.1873

Continued…

Woman's education level	8.41	3	<.0001
Number of household members	593.63	1	<.0001
Sex of the household head	6.94	1	0.0087
Literacy	0.61	1	0.4348
Wealth index	3.53	4	0.0074
Ever had a terminated pregnancy	7.24	1	0.0074
Contraceptive method used	61.09	1	<.0001
Body Mass Index	1.05	1	0.3058
Current marital status	443.26	2	<.0001
Number of other partners	11.85	1	0.0006
Cohabitation duration	70.76	1	<.0001
Partner's desire for children	1.18	3	0.3185
Partner's education level	2.07	4	0.0828
Partner's occupation	0.61	1	0.4349
Woman's occupation	34.31	1	<.0001
Partner's age	8.65	1	0.0034
The person who usually decides how to spend a woman's earnings	11.47	3	<.0001
Woman's earnings compared to partner	1.72	4	0.1432
Interaction effects			
Woman's age by contraceptive use	28.90	1	<.0001
Wealth index by contraceptive use	1.42	4	0.2252

***DF=Degrees of Freedom**

Prediction Accuracy of the Model

The concordant rate was 83.2%, as shown in Table 3.23 below; this value tells us how good the model was in separating 0's and 1's. The Gamma statistic has a value of 0.665, and it indicates a high positive association between variables. The Somers' D statistic is 0.664, showing that the model is a good predictor, and that there is a high association between the variables. The area under the curve (ROC) of 0.832 implies that 83.2% of the probabilities were predicted correctly by the model. Hence, the model is said to have a very good prediction accuracy.

Table 3.23: Association of predicted probabilities and observed responses (Tanzania)

Percent Concordant	83.2	Somers' D	0.664
Percent Discordant	16.7	Gamma	0.665
Percent Tied	0.1	Tau-a	0.329
Pairs	43502480	C	0.832

Interpretation of the results

Table 3.24 shows that a woman whose partner does not drink alcohol is 0.38 (OR=0.379, p-value<.0001) times less likely to experience IPV, compared to a woman whose partner does drink it. A woman who has not witnessed her father beating her mother is 0.60 (OR=0.596, p-value<.0001) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.48 (OR=1.483, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A woman who does not know whether wife-beating is acceptable or unacceptable is 0.65 (OR=0.650, p-value=0.0002) times less likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A unit increase in a woman's age increases the chances of her experiencing IPV by 0.0200 units.

A woman from the Kilimanjaro province is 0.49 (OR=0.485, p-value=0.0019) times less likely to experience IPV compared to a woman from Arusha province. A woman from Tanga province is 0.60 (OR=0.598, p-value=0.0063) times less likely to experience IPV compared to a woman from Arusha province. A woman from Njobe province is 0.52 (OR=0.523, p-value=0.0003) times less likely to experience IPV compared to a woman from Arusha province. A woman from Kusini Unguja province is 1.411 (OR=1.411, p-value=0.0461) times less likely to experience IPV compared to a woman from Arusha province. A woman from Kaskazini Pemba province is 0.41 (OR=0.410, p-value=0.0008) less likely to experience IPV compared to a woman from Arusha province.

A woman with secondary education is 0.71 (OR=0.714, p-value<.0001) times less likely to experience IPV, compared to a woman who has no education. A woman who has a higher education is 0.37 (OR=0.371, p-value<.0001) times less likely to experience IPV, compared to a woman who has no education. A one-member increase in the number of household members decreases the chance of a woman experiencing IPV by 0.1936 units. A woman from a household where the head of the house is male is 0.83 (OR=0.833, p-value=0.0032) times less likely to experience IPV, compared to a woman from a household where the head of the house is female.

A woman from the poorest wealth index class is 1.40 (OR=1.397, p-value<.0001) times most likely to experience IPV compared to a woman from the richest wealth index class. A woman from a poorer wealth index class is 1.28 (OR=1.284, p-value=0.0051) times more likely to experience IPV compared to a woman from the richest wealth index class. A woman from a middle wealth index class is 1.33 (OR=1.325, p-value=0.0008) times more likely to experience IPV compared to a woman from the richest wealth index class. A woman from a richer wealth index class is 1.24 (OR=1.239, p-value=0.0027) times more likely to experience IPV compared to a woman from the richest wealth index class.

A woman who has never terminated a pregnancy is 0.86 (OR=0.858, p-value=0.0119) times less likely to experience IPV, compared to a woman who has terminated a pregnancy. A married woman is 6.24 (OR=6.240, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 5.81 (OR=5.811, p-value<.0001) times more likely to experience IPV compared to a single woman. A one-wife increase in the woman’s partner increases the chances of experiencing IPV by 0.1368 units. A woman who stayed with her partner for 5-9 years is 0.51 (OR=0.511, p-value<.0001) times less likely to experience IPV, compared to a woman who stayed with her partner for 0-4 years.

A woman with a partner who wants more children compared to her is 1.18 (OR=1.177, p-value=0.0025) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as her. A woman with a partner who wants fewer children compared to her is 1.32 (OR=1.320, p-value=0.0051) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as her. A woman who does not know the number of children her partner wants is 1.29 (OR=1.293, p-value<.0001) times more likely to experience IPV, compared to a woman whose partner wants the same number of children as her.

Table 3.24: Parameter estimates of factors associated with intimate partner violence of women of reproductive age from survey logistic regression (Tanzania)

Parameter	Estimate	Standard error	t-value	P-value	Odds ratio
Intercept	0.9472	0.2520	3.76	0.0002	2.578
Husband/partner drinks alcohol (ref=Yes)					
No	-0.9695	0.0570	-17.00	<.0001	0.379
Respondent's father ever beat her mother (ref=Yes)					
No	-0.5177	0.0495	-10.46	<.0001	0.596
Don't know	-0.1222	0.0885	-1.38	0.1678	0.885
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,3942	0,0474	8,31	<.0001	1,483
I don't know	-0,4303	0,1161	-3,71	0,0002	0,650
Respondent current age	0.0200	0.00631	3.17	0.0016	1.020
Region (ref=Arusha)					
Dodoma	-0.2886	0.1654	-1.74	0.0817	0.749
Kilimanjaro	-0.7236	0.2318	-3.12	0.0019	0.485
Tanga	-0.5134	0.1873	-2.74	0.0063	0.598
Morogoro	0.00459	0.1974	0.02	0.9815	1.005
Pwari	0.0400	0.1812	0.22	0.8255	1.041
Dar Es Salaam	0.0524	0.1642	0.32	0.7498	1.054
Lindi	0.1387	0.1850	0.75	0.4540	1.149
Mtwara	0.2283	0.2262	1.01	0.3132	1.256
Ruvuma	-0.2257	0.1733	-1.30	0.1933	0.798
Iringa	-0.3506	0.2001	-1.75	0.0803	0.704
Mbeya	-0.1714	0.2130	-0.80	0.4212	0.842
Singida	-0.2440	0.2260	-1.08	0.2807	0.784
Tabora	0.1075	0.1977	0.54	0.5868	1.113
Rukwa	-0.1110	0.1925	-0.58	0.5643	0.895
Kigoma	-0.2369	0.2137	-1.11	0.2681	0.789
Shinyanga	0.1914	0.1832	1.04	0.2968	1.211
Kagera	0.1055	0.2144	0.49	0.6230	1.111

Continued…

Mwanza	0.1650	0.1882	0.88	0.3810	1.179
Mara	0.3532	0.1808	1.95	0.0513	1.424
Manyanga	0.2431	0.1882	1.29	0.1970	1.275
Njobe	-0.6485	0.1784	-3.64	0.0003	0.523
Katavi	-0.2281	0.1749	-1.30	0.1928	0.796
Simiyu	-0.00987	0.1841	-0.05	0.9573	0.990
Geita	0.0510	0.1853	0.28	0.7831	1.052
Kaskazini Unguja	0.3444	0.2003	1.72	0.0860	1.411
Kusini Unguja	0.3441	0.1721	2.00	0.0461	1.411
Mjini Magharibi	0.2741	0.1671	1.64	0.1015	1.315
Kaskazini Pemba	-0.8904	0.2650	-3.36	0.0008	0.410
Kusini Pemba	-0.4907	0.3009	-1.63	0.1036	0.612
Respondent's highest education level (ref=No education)					
Primary education	0.0522	0.0644	0.81	0.4183	1.054
Secondary education	-0.3373	0.0820	-4.11	<.0001	0.714
Higher education	-0.9905	0.1872	-5.29	<.0001	0.371
Number of household members	-0.1936	0.00791	-24.48	<.0001	0.824
Sex of household head (ref=Female)					
Male	-0.1825	0.0617	-2.96	0.0032	0.833
Wealth index (ref=Richest)					
Poorest	0.3343	0.0842	3.97	<.0001	1.397
Poorer	0.2498	0.0888	2.81	0.0051	1.284
Middle	0.2817	0.0833	3.38	0.0008	1.325
Richer	0.2145	0.0711	3.02	0.0027	1.239
Ever had a terminated pregnancy (ref=Yes)					
No	-0.1531	0.0607	-2.52	0.0119	0.858
The contraceptive method used (ref=Yes)					
No	-1.3795	0.1688	-8.17	<.0001	0.252
Current marital status (ref=Single)					
Married	1.8310	0.0622	29.42	<.0001	6.240
Living with partner	1.7597	0.0802	21.95	<.0001	5.811
Number of other wives	0.1368	0.0271	5.04	<.0001	1.147
Cohabitation duration (grouped) (ref=0 to 4 years)					
5 to 9 years	-0.6720	0.0814	-8.26	<.0001	0.511
Husband's desire for children (ref=Both want same)					
Husband wants most	0.1631	0.0538	3.03	0.0025	1.177
Husband wants fewer	0.2779	0.0989	2.81	0.0051	1.320
Don't know	0.2568	0.0550	4.67	<.0001	1.293
Respondent's occupation (ref=Unemployed)					
Employed	0.3486	0.0601	5.80	<.0001	1.417
Husband/partner's age	-0.0114	0.00354	-3.24	0.0013	0.989

Continued…

The person who usually decides how to spend a woman's earnings (ref=Woman alone)					
Woman and partner	-0,2559	0,0534	-4,79	<.0001	0,774
Partner alone	-0,4357	0,0822	-5,3	<.0001	0,647
Someone else	-0,856	1,2903	-0,66	0,5073	0,425
Interaction effects					
Respondent's age by contraceptive use	0.0302	0.00535	5.65	<.0001	1.031

A woman who earns about the same as her partner is positively associated with IPV. An employed woman is 1.42 (OR=1.417, p-value<.0001) times more likely to experience IPV compared to an unemployed woman. A woman who decides with her partner on how to spend her earnings is 0.77 (OR=0.774, p-value<.0001) times less likely to experience IPV, compared to a woman who decides how to spend her earnings alone. A woman whose partner decides how to spend her earnings is 0.65 (OR=0.647, p-value<.0001) times less likely to experience IPV, compared to a woman who decides how to spend her earnings alone.

Interaction Effects

Figure 3.11 shows that women from aged between 15 to 45 years using contraceptives, are at a high risk of experiencing IPV, compared to women not using contraceptives. There is a slight decrease of IPV for women between about 46 to 50 years of age from the same figure. From about the ages of 46 to 50 years, women who are not using contraceptives are at a high risk of experiencing IPV.

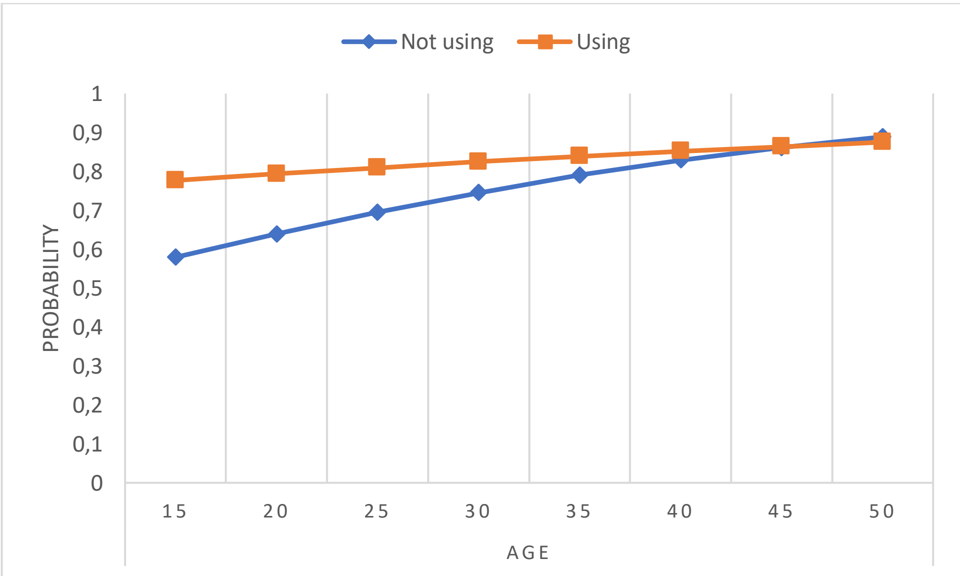

Figure 3.11: Predicted probability of experiencing IPV by woman's age and contraceptive use

Uganda

Model Checking

Table 3.25 shows that the full model has a smaller AIC (19093.573) compared to the AIC (23272.633) of the reduced model; this indicates that the fitted model better explains the data.

Table 3.25: Model fit statistics

Criterion	Intercept only	Intercept and covariates
AIC	23272.633	19093.573
SC	23280.459	19617.905
-2 Log L	23270.633	18959.573

The explanatory variables used in the model are shown in Table 3.26 with their p-values; this indicates their influence on the response variable.

Table 3.26: Type 3 analysis of effects (Uganda)

Type 3 Analysis of Effects			
Effect	**F-value**	**DF**	**P-value**
Partner drinks alcohol	417.33	1	<.0001
Woman's father ever beat her mother	131.25	2	<.0001
Access to the media	17.99	2	<.0001
Wife-beating attitude	2.87	2	0.0473
Woman's current age	37.06	1	<.0001
Region	8.06	14	<.0001
Type of residence	2.40	1	0.1215
Woman's education level	2.21	3	0.0856
Number of household members	365.34	1	<.0001
Sex of the household head	16.73	1	<.0001
Literacy	1.62	1	0.2036
Wealth index	10.37	4	<.0001
Ever had a terminated pregnancy	9.42	1	0.0022
Contraceptive method used	55.81	1	<.0001
Body Mass Index	0.00	1	0.9598
Current marital status	329.13	2	<.0001
Number of other partners	23.18	1	<.0001
Cohabitation duration	42.71	1	<.0001
Partner's desire for children	5.22	3	0.0014
Partner's education level	8.22	4	<.0001
Partner's occupation	2.46	1	0.1176
Woman's occupation	41.92	4	<.0001
Partner's age	1.08	1	0.2991

Continued…

The person who usually decides how to spend a woman's earnings	14.93	3	<.0001
Woman's earnings compared to partner	15.45	4	<.0001
Knowledge of Sexually Transmitted Infections	0.78	1	0.3777
Interaction effects			
Woman's age by contraceptive use	42.78	1	<.0001
Wealth index by contraceptive use	3.66	4	0.0058

***DF=Degrees of Freedom**

Prediction Accuracy of the Model

The concordant rate was 77.8%, as shown in Table 3.27 below; this value tells us how good the model was in separating 0's and 1's. The Gamma statistic has a value of 0.559, and it indicates a high positive association between variables. The Somers' D statistic is 0.560, showing that the model is a good predictor, and that there is a high association between the variables. The area under the curve (ROC) of 0.779 implies that 77.9% of the probabilities were predicted correctly by the model, and hence the model is said to have a good prediction accuracy.

Table 3.27: Association of predicted probabilities and observed responses (Uganda)

Percent concordant	77.8	Somers' D	0.559
Percent discordant	22.0	Gamma	0.560
Percent tied	0.2	Tau-a	0.244
Pairs	74826784	C	0.779

Table 3.28 shows that a woman with a partner who does not drink alcohol is 0.43 (OR=0.428, p-value<.0001) times less likely to experience IPV, compared to a woman with a partner who does drink it. A woman who has never witnessed her father beating her mother is 0.54 (OR=0.536, p-value<.0001) times less likely to experience IPV compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.22 (OR=1.217, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable.

A woman with high exposure to the media is 0.72 (OR=0.722, p-value=0.0160) times less likely to experience IPV compared to a woman with low exposure to the media. A woman from the Kampala province is 1.32 (OR=1.322, p-value=0.0310) times more likely to experience IPV compared to a woman from South Buganda. A woman from the Busoga region is 1.25 (OR=1.250, p-value=0.0421) times more likely to experience IPV compared to a woman from South Buganda. A woman from the Bukedi province is 1.59 (OR=1.586, p-value<.0001) times more likely to experience IPV compared to a woman from South Buganda. A woman from the Busigu province is 1.44 (OR=1.437, p-value=0.0006) times more likely to experience IPV compared to a woman from South Buganda. A woman from the Bunyoro province is 1.29 (OR=1.292, p-value=0.0225) times more likely to experience IPV compared to a woman from South Buganda. A woman from the Kigezi province is 1.50 (OR=1.504, p-value=0.0002) times more likely to experience IPV compared to a woman from South Buganda. A one-member increase in the number of household members reduces a woman's chances of experiencing IPV by 0.1366 units.

A woman from a house where the household head is male is 1.13 (OR=1.130, p-value=0.0081) times more likely to experience IPV, compared to a woman from a house where the household head is female. A woman who has never had a terminated pregnancy is 0.88 (OR=0.875, p-value=0.0031) times less likely to experience IPV, compared to a woman who has had one. A married woman is 3.54 (OR=3.541, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 3.60 (OR=3.598, p-value<.0001) times more likely to experience IPV compared to a single woman. A unit increase in the number of other partners a man has, increases the woman's chances of experiencing IPV by 0.0555 units. A woman who has stayed with her partner for 5-9 years is 0.67 (OR=0.666, p-value<.0001) times less likely to experience IPV, compared to a woman who has stayed with her partner for 0-4 years.

A woman who does not know the number of children her partner wants is 0.85 (OR=0.850, p-value=0.0020) times less likely to experience IPV, compared to a woman who wants the same number of children as him. A woman who has a partner with a secondary level of education is 0.78 (OR=0.781, p-value=0.0015) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who has a partner with a higher level of education is 0.62 (OR=0.624, p-value<.0001) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who does not know her partner's level of education is 0.50 (OR=0.502, p-value<.0001) times less likely to experience IPV, compared to a woman who has a partner with no education. A woman who has an employed partner is 0.77 (OR=0.772, p-value=0.0102) times less likely to experience IPV, compared to a woman whose partner is unemployed.

Table 3.28: Survey logistic regression model coefficients, standard errors, and odds ratios (Uganda)

Parameter	**Estimate**	**Standard error**	**t-value**	**P-value**	**Odds ratio**
Intercept	-0,4834	0.2005	-2,16	0,0311	0,6166831
Partner drinks alcohol (ref=Yes)					
No	-0.8475	0.0423	-20.02	<.0001	0.428
Woman's father ever beat her mother (ref=Yes)					
No	-0.6237	0.0376	-16.59	<.0001	0.536
Don't know	-0.0601	0.1050	-0.57	0.5672	0.942
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,1963	0,0379	26,7825	<.0001	1,2169
I don't know	-0,4259	0,1714	6,1756	0,013	0,6532

Continued…

Access to the media (ref=Low exposure)					
Medium exposure	-0.0151	0.0401	-0.38	0.7065	0.985
High exposure	-0.3255	0.1347	-2.42	0.0160	0.722
Woman's current age	0.0114	0.00461	2.47	0.0139	1.011
Region (ref=South Buganda)					
Kampala	0.2789	0.1290	2.16	0.0310	1.322
North Buganda	0.1529	0.1085	1.41	0.1592	1.165
Busoga	0.2228	0.1094	2.04	0.0421	1.250
Bukedi	0.4611	0.1073	4.30	<.0001	1.586
Bugisu	0.3628	0.1053	3.44	0.0006	1.437
Teso	0.0714	0.1193	0.60	0.5496	1.074
Karamoja	0.0951	0.1155	0.82	0.4109	1.100
Lango	-0.1844	0.1264	-1.46	0.1449	0.832
Acholi	-0.0473	0.1139	-0.41	0.6783	0.954
West Nile	-0.0930	0.1186	-0.78	0.4335	0.911
Bunyoro	0.2561	0.1120	2.29	0.0225	1.292
Tooro	-0.1905	0.1200	-1.59	0.1127	0.827
Ankole	0.1478	0.1113	1.33	0.1845	1.159
Kigezi	0.4082	0.1077	3.79	0.0002	1.504
Number of household members	-0.1366	0.00747	-18.28	<.0001	0.872
Sex of household head (ref=Female)					
Male	0.1218	0.0459	2.66	0.0081	1.130
Wealth index (ref=Richest)					
Poorest	0.4180	0.1222	3.42	0.0007	1.519
Poorer	0.1097	0.1083	1.01	0.3116	1.116
Middle	-0.0210	0.1051	-0.20	0.8419	0.979
Richer	-0.0260	0.0928	-0.28	0.7792	0.974
Ever had a terminated pregnancy (ref=Yes)					
No	-0.1331	0.0448	-2.97	0.0031	0.875
The contraceptive method used (ref=Yes)					
No	-1.1418	0.1571	-7.27	<.0001	0.319
Current marital status (ref=Single)					
Married	1.2644	0.0565	22.36	<.0001	3.541
Living with partner	1.2805	0.0527	24.28	<.0001	3.598
Number of other partners	0.0555	0.0113	4.89	<.0001	1.057
Cohabitation duration (ref=0 to 4 years)					
5 to 9 years	-0.4061	0.0623	-6.52	<.0001	0.666
Partner's desire for children (ref=Both want same)					
Partner wants more	0.0238	0.0458	0.52	0.6030	1.024
Partner wants fewer	-0.0882	0.0673	-1.31	0.1905	0.916
Don't know	-0.1624	0.0525	-3.10	0.0020	0.850

Continued…

Partner's education level					
Primary education	-0.1173	0.0714	-1.64	0.1011	0.889
Secondary education	-0.2470	0.0774	-3.19	0.0015	0.781
Higher education	-0.4721	0.0924	-5.11	<.0001	0.624
Don't know	-0.6900	0.1280	-5.39	<.0001	0.502
Partner's occupation (ref=Unemployed)					
Employed	-0.2583	0.1003	-2.58	0.0102	0.772
The person who usually decides what to do with woman's earnings (ref=Woman alone)					
Woman and partner	-0.2170	0.0410	-5.30	<.0001	0.805
Partner alone	-0.3914	0.0689	-5.68	<.0001	0.676
Someone else	-0.1932	0.5123	-0.38	0.7062	0.824
Woman's earnings compared to her partner (ref=Less compared to him)					
More compared to him	0.2141	0.0654	3.28	0.0011	1.239
About the same	-0.4210	0.0495	-8.51	<.0001	0.656
Partner does not bring in	-0.2745	0.1266	-2.17	0.0305	0.760
Don't know	-0.5402	0.0956	-5.65	<.0001	0.583
Interaction effects					
Woman's age by contraceptive use	0.0320	0.00454	7.04	<.0001	1.032
Wealth index (ref=Richest) by contraceptive use (ref=Not using)					
Poorest by not using contraceptives	-0.2709	0.1220	-2.22	0.0267	0.763
Poorer by not using contraceptives	0.1093	0.1173	0.93	0.3519	1.115
Middle by not using contraceptives	0.1517	0.1136	1.34	0.1823	1.164
Richer by not using contraceptives	0.1522	0.1163	1.31	0.1911	1.164

A woman who decides with her partner on how to spend her earnings is 0.81 (OR=0.805, p-value<.0001) times less likely to experience IPV, compared to a woman who decides how to spend them alone. A woman whose partner decides how to spend her earnings is 0.68 (OR=0.676, p-value<.0001) times less likely to experience IPV, compared to a woman who decides how to spend them alone. A woman who earns more than her partner is 1.24 (OR=1.239, p-value=0.0011) times more likely to experience IPV, compared to a woman who earns less than him. A woman who earns about the same as her partner is 0.66 (OR=0.656, p-value<.0001) times less likely to experience IPV, compared to a woman who earns less than him. A woman who has a partner who does not bring in earnings is 0.76 (OR=0.760, p-value=0.0305) times less likely to experience IPV, compared to a woman who earns less than him. Finally, a woman who does not know how much her partner earns is 0.58 (OR=0.583, p-value<.0001) times less likely to experience IPV, compared to a woman who earns less than him.

Interaction Effects

Figure 3.12 shows that IPV increases with increasing age, whether a woman is using contraceptives or not. We observe from the same figure that in for 15 to 35 years-old age range, IPV is higher among women using contraceptives compared to women not using them. Women who are 35 years old have an equal chance of experiencing IPV, whether using contraceptives or not. From about 37 years of age, the IPV for women not using contraceptives is higher compared to those using them.

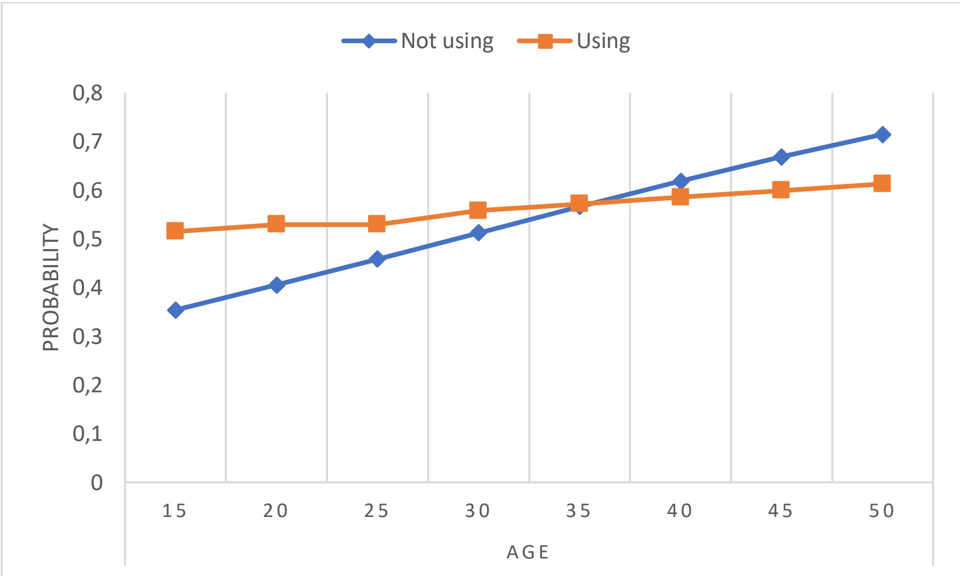

Figure 3.12: Predicted probability of experiencing IPV by woman's age and contraceptive use

It can be observed from Figure 3.13 that IPV is higher among women using contraceptives compared to women not using any. It further shows that IPV decreases with increasing socio-economic status of women using contraceptives from the poorest households. It can also be seen that IPV increases slightly with increasing wealth index, among the women using contraceptives from the most impoverished households, and for those from a poorer household. It decreases slightly with increasing wealth index among women from poorer households to those from richer ones.

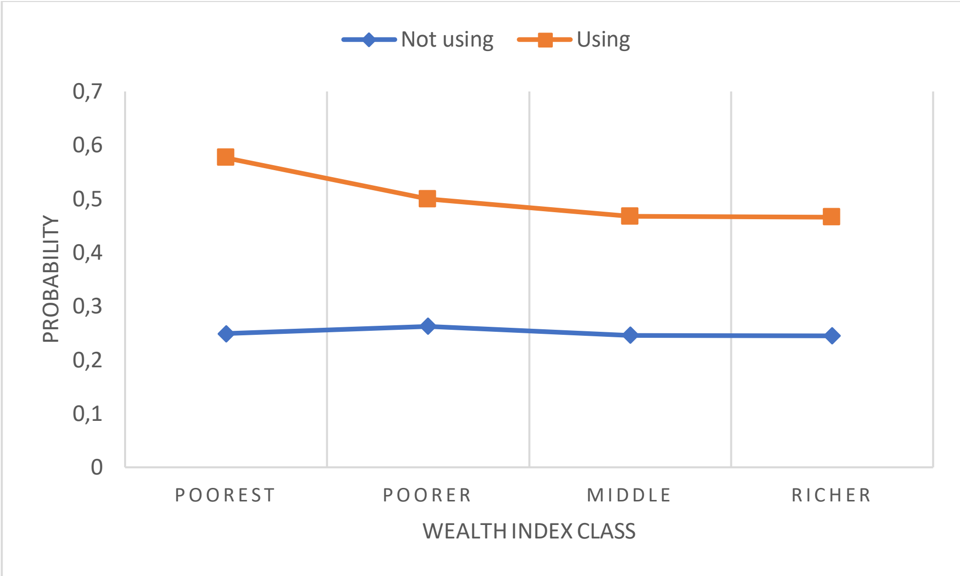

Figure 3.13: Predicted probability of experiencing IPV by Wealth index class and contraceptive use

3.14 Results Interpretation of Design Effect

The variance for variables that produced less than one for DEFF values was overestimated while using a logistic regression model, implying that the variance values were too large compared to those found when using the complex design. The variance for variables that produced a value greater than zero for DEFF was underestimated while using the logistic regression model. The implication of this is that the variance values were too small compared to those found when using the complex design. Therefore, using survey logistic regression is good since it considers the survey design features.

South Africa

Table 3.6 and Table 3.20 contain the odds ratios for logistic and survey logistic regression, respectively. The models fitted by both methods produced the area under the curve (ROC), which is between 0.7 and 0.8. The ROC suggests that both the models had a good prediction accuracy. Table 3.29 shows the DEFF and DEFT (square root of DEFF) for each significant estimated coefficient. The effects of a partner's alcohol consumption status give a DEFF=1.0699 and DEFT=1.0344; the standard error and confidence interval would be 1.0344 times greater than they would be for simple random sampling. The effects of a woman's father ever beating her mother have a DEFF=1.1446 and DEFT=1.0699; the standard error and confidence interval would be 1.0699 times greater than they would be for simple random sampling.

The effects of an 'acceptable' wife-beating attitude have a DEFF=1.0115 and DEFT=1.0057; the standard error and confidence interval would be 1.0057 times greater than they would be for simple random sampling. The effects of an 'unknown' (Don't know) wife-beating attitude have a DEFF=1.0811 and DEFT=1.0398; the standard error and confidence interval would be 1.0398 times greater than they would be for simple random sampling. The effects of a 'medium' media exposure have a DEFF=0.9815 and DEFT=0.9907; the standard error and confidence interval would be 0.9907 times less compared to those for simple random sampling. The effects of a 'high' media exposure have a DEFF=1.0000 and DEFT=1.000; the standard error and confidence interval would be equal for a complex design to those for simple random sampling.

Table 3.29: Design effect, comparison of logistic and survey logistic regression (South Africa)

Parameter	Estimate	P-value	Var (CSD)	Var (SRS)	DEFF	DEFT
Intercept	0,5656	0.0481	0,0816	0,0792	1,0301	1,0149
Partner drinks alcohol (ref=Yes)						
No	-0.8209	<.0001	0,0033	0,0031	1,0699	1,0344
Woman's father ever beat her mother (ref=Yes)						
No	-0.8233	<.0001	0,0068	0,0060	1,1446	1,0699
Don't know	-0.3209	0.0200	0,0189	0,0185	1,0252	1,0125
Wife-beating attitude (ref=Unacceptable)						
Acceptable	0.5864	<.0001	0.0112	0.0110	1.0115	1.0057
I don't know	-1.1440	<.0001	0.0782	0.0724	1.0811	1.0398
Access to the media (ref=Low exposure)						
Medium exposure	0.2897	0.0007	0,0072	0,0074	0,9815	0,9907
High exposure	0.2262	0.0283	0,0106	0,0106	1,0000	1,0000
Woman's current age	0,0122	0.0499	0,0000386	0,0000384	1,0032	1,0016
Region (ref=Eastern Cape)						
Western Cape	-0.4393	0.0016	0,0193	0,0173	1,1157	1,0563
Northern Cape	-0.6347	<.0001	0,0248	0,0167	1,4815	1,2172
Number of household members	-0.1607	<.0001	0,000130	0,000132	0,9827	0,9913
Sex of household head (ref=Female)						
Male	0.2149	0.0006	0,0039	0,0042	0,9391	0,9691
Wealth index (ref=Richest)						
Poorest	0,6168	<.0001	0,0246	0,0228	1,0797	1,0391
Poorer	0,7562	<.0001	0,0218	0,0205	1,0668	1,0328
Middle	0,5239	0.0003	0,0212	0,0194	1,0941	1,0460
Richer	0,4569	0.0013	0,0200	0,0197	1,0114	1,0057
Ever had a terminated pregnancy (ref=Yes)						
No	-0.4158	<.0001	0,0080	0,0074	1,0757	1,0372
The contraceptive method used (ref=Yes)						
No	-1,6123	<.0001	0,0688	0,0673	1,0225	1,0112
Body Mass Index	-0.00220	0.0309	0,00000104	0,00000097	1,0723	1,0355
Current marital status (ref=Single)						
Married	0.5331	<.0001	0,0062	0,0061	1,0077	1,0038
Living with partner	0.5908	<.0001	0,0074	0,0077	0,9638	0,9817
Cohabitation duration (ref=0 to 4 years)						
5 to 9 years	-0.2212	0.0427	0,0119	0,0107	1,1114	1,0542

Continued…

Partner's desire for children (ref=Both want same)						
Partner wants more	0.3046	<.0001	0,0052	0,0054	0,9677	0,9837
Don't know	0.3093	<.0001	0,0045	0,0044	1,0181	1,0090
Partner's occupation (ref=Don't know)						
Employed	-0.1725	0.0268	0,0060	0,0059	1,0316	1,0157
Woman's occupation (ref=Unemployed)						
Employed	0.2582	<.0001	0,0039	0,0037	1,0329	1,0163
Partner's age	-0.0123	0.0076	0,0000210	0,0000208	1,0088	1,0044
Woman's earnings compared to partner (ref=Less compared to him)						
About the same	0.1709	0.0256	0,0058	0,0059	0,9896	0,9948
Partner does not bring in earnings	0.3442	0.0006	0,0099	0,0103	0,9687	0,9842
Ever heard of a Sexually Transmitted Infections (STIs) (ref=Yes)						
No	-0.8482	0.0001	0,0492	0,0519	0,9472	0,9732
Interaction effects						
Woman's age by contraceptive use	0,0285	<.0001	0,0000317	0,0000335	0,9455	0,9724
Wealth index (ref=Richest) by contraceptive use (ref=Not using)						
Poorest by not using contraceptives	0,4954	0.0172	0,0430	0,0414	1,0387	1,0192
Middle by not using contraceptives	0,5496	0.0071	0,0415	0,0391	1,0606	1,0298

The effects of living in the Western Cape region have a DEFF=1.1157 and DEFT=1.0563; the standard error and confidence interval would be 1.0563 times greater than they would be for simple random sampling. The effects of living in the Northern Cape region have a DEFF=1.4815 and DEFT=1.2172; the standard error and confidence interval would be 1.2172 times greater than they would be for simple random sampling. The effects of the number of household members have a DEFF=0.9827 and DEFT=0.9913; the standard error and confidence interval would be 0.9913 times less than they would be for simple random sampling. The effects of the household head have a DEFF=0.9391 and DEFT=0.9691; the standard error and confidence interval would be 0.9691 times less compared to those for simple random sampling.

The effects of a woman not terminating a pregnancy have a DEFF=1.0757 and DEFT=1.0372; the standard error and confidence interval would be 1.0372 times greater than they would be for simple random sampling. The effects of the body mass index have a DEFF=1.0723 and DEFT=1.0355; the standard error and confidence interval would be 1.0355 times greater than for simple random sampling. The effects of a woman’s current marital status as ‘married have a DEFF=1.0077 and DEFT=1.0038; the standard error and confidence interval would be 1.0038 times greater than for simple random sampling. The effects of a woman’s current marital status as ‘living with a partner’ have a DEFF=0.9638 and DEFT=0.9817; the standard error and confidence interval would be 0.9817 times less than they would be for simple random sampling. The effects of a cohabitation period of 5-9 years have a

DEFF=1.1114 and DEFT=1.0542; the standard error and confidence would be 1.0542 times greater than they would be for simple random sampling.

The effects of a woman’s partner's desire for more children have a DEFF=0.9677 and DEFT=0.9677; the standard error and confidence interval would be 0.9677 times less than they would be for simple random sampling. The effects of a woman’s partner's ‘unknown’ desire for children have a DEFF=1.0181 and DEFT=1.0090; the standard error and confidence interval would be 1.0090 times greater than for simple random sampling. The effects of a woman’s partner occupation status being ‘employed’ have a DEFF=1.0316 and DEFT=1.0157; the standard error and confidence interval would be 1.0157 times greater than they would be for simple random sampling. The effects of woman’s occupation status being ‘employed’ have a DEFF=1.0329 and DEFT=1.0163; the standard error and confidence interval would be 1.0163 times greater than they would be for simple random sampling. The effects of a woman’s partner’s age have a DEFF=1.0088 and DEFT=1.0044; the standard error and confidence interval would be 1.0044 times greater than for simple random sampling. The effects of a woman’s earnings being ‘about the same as’ her partner’s, have a DEFF=0.9896 and DEFT=0.9948; the standard error and confidence interval would be 0.9948 times less than for simple random sampling. The effects of a woman’s partner ‘not bringing in earnings’ have a DEFF=0.9687 and DEFT=0.9842; the standard error and confidence interval would be 0.9842 times less than they would be for simple random sampling.

The effects of a woman having ‘no knowledge’ of STIs have a DEFF=0.9472 and DEFT=0.9732; the standard error and confidence interval would be 0.9732 times less than they would be for simple random sampling. The effects of a woman’s knowledge of whether she can refuse sex from her partner have a DEFF=1.0658 and DEFT=1.0324; the standard error and confidence interval would be 1.0324 times greater than they would be for simple random sampling. The effects of a woman’s knowledge of whether she can or cannot refuse sex from her partner being ‘unknown’ (don’t know) have a DEFF=0.8744 and DEFT=0.9351; the standard error and confidence interval would be 0.9351 times less than they would be for simple random sampling.

Interaction Effects

The effects of age and contraceptive use have a DEFF=0.9455 and DEFT=0.9724; the standard error and confidence interval would be 0.9724 times less than they would be for simple random sampling. The effects of the ‘poorest’ wealth index and ‘not using’ contraceptives have a DEFF=1.0387 and DEFT=1.0192; the standard error and confidence interval would be 1.0192 times greater than they would be for simple random sampling. The effects of a ‘middle’ wealth index and ‘not using’ contraceptives have a DEFF=1.0606 and DEFT=1.0298; the standard error and confidence interval would be 1.0298 times greater than they would be for simple random sampling.

Tanzania

Table 3.11 and Table 3.24 contain the odds ratios for logistic and survey logistic regression, respectively. The models fitted by both methods produced the area under the curve (ROC), which is between 0.8 and 0.9. The ROC suggests that both the models had a very good prediction accuracy. Table 3.30 shows the DEFF and DEFT (square root of DEFF) for each significant estimated coefficient. The effects of the alcohol consumption status of a woman's partner have a DEFF=1.0859 and DEFT=1.0420; the standard error and confidence interval would be 1.0420 times greater than they would be for simple random sampling. The effects of a woman's father 'not' ever beating her mother have a DEFF=1.0122 and DEFT=1.0061; the standard error and confidence interval would be 1.0061 times greater than for simple random sampling. The effects of an 'acceptable' wife-beating attitude have a DEFF=1.0128 and DEFT=1.0064; the standard error and confidence interval would be 1.0064 times greater than for simple random sampling. The effects of a wife-beating attitude being 'unknown' (I don't know) have a DEFF=0.8982 and DEFT=0.9478; the standard error and confidence interval would be 0.9478 times less than they would be for simple random sampling.

Table 3.30: Design effect, comparison of logistic, and survey logistic regression (Tanzania)

Parameter	**Estimate**	**P-value**	**Var (CSD)**	**Var (SRS)**	**DEFF**	**DEFT**
Intercept	0.9472	0.0002	0,063504	0,059049	1,0754	1,0370
Partner drinks alcohol (ref=Yes)						
No	-0.9695	<.0001	0,003249	0,002992	1,0859	1,0420
Woman's father ever beat her mother (ref=Yes)						
No	-0.5177	<.0001	0,00245	0,002421	1,0122	1,0061
Wife-beating attitude (ref=Unacceptable)						
Acceptable	0.3942	<.0001	0.00225	0.00222	1.01278	1.00637
I don't know	-0.4303	0.0002	0.01348	0.01501	0.89824	0.94776
Woman's current age	0.0200	0.0016	3,98E-05	3,84E-05	1,036	1,018
Region (ref=Arusha)						
Dodoma	-0.2886	0.0817	0,027357	0,03052	0,8964	0,9468
Kilimanjaro	-0.7236	0.0019	0,053731	0,031827	1,6883	1,2993
Tanga	-0.5134	0.0063	0,035081	0,02706	1,2964	1,1386
Njobe	-0.6485	0.0003	0,031827	0,030625	1,0392	1,0194
Kusini Unguja	0.3441	0.0461	0,029618	0,0324	0,9141	0,9561
Kaskazini Pemba	-0.8904	0.0008	0,070225	0,037056	1,8951	1,3766
Woman's highest education level (ref=No education)						
Secondary education	-0.3373	<.0001	0,006724	0,006773	0,9927	0,9964
Higher education	-0.9905	<.0001	0,035044	0,051665	0,6783	0,8236
Number of household members	-0.1936	<.0001	6,26E-05	5,61E-05	1,1153	1,0561
Sex of household head (ref=Female)						
Male	-0.1825	0.0032	0,003807	0,003745	1,0164	1,0082

Continued…

Wealth index (ref=Richest)						
Poorest	0.3343	<.0001	0,00709	0,007157	0,9906	0,9953
Poorer	0.2498	0.0051	0,007885	0,006496	1,2138	1,1017
Middle	0.2817	0.0008	0,006939	0,005898	1,1764	1,0846
Richer	0.2145	0.0027	0,005055	0,004747	1,0649	1,0319
Ever had a terminated pregnancy (ref=Yes)						
No	-0.1531	0.0119	0,003684	0,003226	1,1420	1,0687
The contraceptive method used (ref=yes)						
No	-1.3795	<.0001	0,028493	0,029344	0,9710	0,9854
Current marital status (ref=Single)						
Married	1.8310	<.0001	0,003869	0,003894	0,9936	0,9968
Living with partner	1.7597	<.0001	0,006432	0,00558	1,1527	1,0736
Number of other partners	0.1368	<.0001	0,000734	0,00059	1,2437	1,1152
Cohabitation duration (grouped) (ref=0 to 4 years)						
5 to 9 years	-0.6720	<.0001	0,006626	0,00555	1,1938	1,0926
Partner's desire for children (ref=Both want same)						
Partner wants more	0.1631	0.0025	0,002894	0,003114	0,9296	0,9642
Partner wants fewer	0.2779	0.0051	0,009781	0,009584	1,0205	1,0102
Don't know	0.2568	<.0001	0,003025	0,002927	1,0335	1,0166
The person who usually decides how to spend a woman's earnings (ref=Woman alone)						
Woman and partner	-0.2559	<.0001	0.00285	0.00251	1.13608	1.06587
Partner alone	-0.4357	<.0001	0.00676	0.00681	0.99274	0.99636
Woman's occupation (ref=Unemployed)						
Employed	0.3486	<.0001	0,003612	0,003588	1,0067	1,0033
Partner's age	-0.0114	0.0013	1,25E-05	1,18E-05	1,0652	1,0321
Interaction effects						
Woman's age by contraceptive use	0.0302	<.0001	2,86E-05	2,77E-05	1,0345	1,0171

The effects of living in Dodoma region have a DEFF=0.8964 and DEFT=0.9468; the standard error and confidence interval would be 0.9468 times less compared to simple random sampling. The effects of living in Kilimanjaro region have a DEFF=1.6883 and DEFT=1.2993; the standard error and confidence interval would be 1.2993 times greater than they would be for simple random sampling. The effects of living in Tanga region have a DEFF=1.2964 and DEFT=1.1386; the standard error and confidence interval would be 1.1386 times greater than they would be for simple random sampling. The effects of living in Njobe region have a DEFF=1.0392 and DEFT=1.0194; the standard error and confidence interval would be 1.0194 times greater than they would be for simple random sampling. The effects of living in Kusini Uguja region have a DEFF=0.9141 and DEFT=0.9561; the standard error and confidence interval would be 0.9561 times less compared to what they would be for simple random sampling. The effects of living in Kaskazini Pemba region have a DEFF=1.8951 and DEFT=1.3766; the standard error and confidence interval would be 1.3766 times greater than they would be for simple random sampling. The effects of a woman attaining a secondary level of education have a DEFF=0.9927 and DEFT=0.9964; the standard error and confidence interval would be 0.9964 times less than they would be for simple random sampling. The effects of a woman attaining higher level of education have a DEFF=0.6783 and DEFT=0.8236; the standard error and confidence interval would be 0.8236 times less than they would be for simple random sampling.

The effects of the number of household members have a DEFF=1.1153 and DEFT=1.0561; the standard error and confidence interval would be 1.0561 times greater than they would be for simple random sampling. The effects of the household head have a DEFF=1.0164 and DEFT=1.0082; the standard error and confidence interval would be 1.0082 times greater than for simple random sampling.

The effects of belonging to the 'poorest' wealth index class have a DEFF=0.9906 and DEFT=0.9953; the standard error and confidence interval would be 0.9953 times less compared to what they would be for simple random sampling. The effects of belonging to a 'poorer' wealth index class have a DEFF=1.2138 and DEFT=1.1017; the standard error and confidence interval would be 1.1017 times greater than they would be for simple random sampling. The effects of belonging to a 'middle' wealth index class have a DEFF=1.1764 and DEFT=1.0846; the standard error and confidence interval would be 1.0846 times greater than they would be for simple random sampling. The effects of belonging to a 'richer' wealth index class have a DEFF=1.0649 and DEFT=1.0319; the standard error and confidence interval would be 1.0319 times greater than they would be for simple random sampling.

The effects of whether a woman has not terminated a pregnancy have a DEFF=1.1420 and DEFT=1.0687; the standard error and confidence interval would be 1.0687 times less compared to simple random sampling. The effects of a woman's current marital status being 'married' have a DEFF=0.9936 and DEFT=0.9968; the standard error and confidence interval would be 0.9968 times less than for simple random sampling. The effects of a woman's current marital status being 'living with a partner' have a DEFF=1.1527 and DEFT=1.0736; the standard error and confidence interval would be 1.0736 times greater than they would be for simple random sampling. The effects of the cohabitation period being between 5-9 years have a DEFF=1.1938 and DEFT=1.0926; the standard error and confidence would be 1.0926 times greater they would be for simple random sampling.

The effects of a woman's partner's wanting more children have a DEFF=0.9296 and DEFT=0.9642; the standard error and confidence interval would be 0.9642 times less than they would be for simple random sampling. The effects of a woman's partner wanting fewer children have a DEFF=1.0205 and DEFT=1.0102; the standard error and confidence interval would be 1.0102 times greater than they would be for simple random sampling. The effects of a woman's partner's desire for children being 'unknown' (don't know) have a DEFF=1.0335 and DEFT=1.0166; the standard error and confidence interval would be 1.0166 times greater than they would be for simple random sampling. The effects of a woman's occupation status being 'employed' have a DEFF=1.0067 and DEFT=1.0033; the standard error and confidence interval would be 1.0033 times greater than for simple random sampling. The effects of a woman and her partner usually deciding how to spend her earnings have a DEFF=1.1361 and DEFT=1.0659; the standard error and confidence interval would be 1.0659 times greater than they would be for simple random sampling. The effects of the partner alone usually deciding how to spend the woman's earnings have a DEFF=0.9927 and DEFT=0.9964; the standard error and confidence interval would be 0.9964 times less than they would be for simple random sampling. The effects of a woman's partner's age have a DEFF=1.0652 and DEFT=1.0321; the standard error and confidence interval would be 1.0321 times greater than they would be for simple random sampling.

Interaction Effects

The effects of age and contraceptive use have a DEFF=1.0345 and DEFT=1.0171; the standard error and confidence interval would be 1.0171 times greater than they would be for simple random sampling.

Uganda

Table 3.16 and Table 3.28 contain the odds ratios for logistic and survey logistic regression, respectively. The models fitted by both methods produced the area under the curve (ROC), which is between 0.8 and 0.9. The ROC suggests that both the models had a very good prediction accuracy. Table 3.31 shows the DEFF and DEFT (square root of DEFF) for each significant estimated coefficient.

The effects of a partner's alcohol consumption status have a DEFF=1.2391 and DEFT=1.1132; the standard error and confidence interval would be 1.1132 times greater than they would be for simple random sampling. The effects of a woman's father ever beating her mother have a DEFF=0.9588 and DEFT=0.9792; the standard error and confidence interval would be 0.9792 times less than they would be for simple random sampling. The effects of an 'acceptable' wife-beating attitude have a DEFF=1.00 and DEFT=1.00; the standard error and confidence interval would be the same as for simple random sampling. The effects of an 'unknown' (don't know) wife-beating attitude have a DEFF=1.00 and DEFT=1.00; the standard error and confidence interval would be the same as they would be for simple random sampling. The effects of a 'high' media exposure have DEFF=1.1954 and DEFT=1.0933; the standard error and confidence interval would be 1.0933 times greater than they would be for simple random sampling.

Table 3.31: Design effect, comparison of logistic and survey logistic regression (Uganda)

Parameter	Estimate	Pr > ChiSq	Var (CSD)	Var (SRS)	DEFF	DEFT
Partner drinks alcohol (ref=Yes)						
No	-0.8475	<.0001	0,0018	0,0014	1,2391	1,1132
Woman's father ever beat her mother (ref=Yes)						
No	-0.6237	<.0001	0,0014	0,0015	0,9588	0,9792
Wife-beating attitude (ref=Unacceptable)						
Acceptable	0.1963	<.0001	0.00144	0.00144	1.00	1.00
Don't know	-0.4259	0.0130	0.02938	0.02938	1.00	1.00

Continued…

Access to the media (ref=Low exposure)						
High exposure	-0.3255	0.0160	0,0181	0,0152	1,1954	1,0933
Woman's current age	0,0114	0.0139	0,0000213	0,0000212	1,0044	1,0022
Region (ref=South Buganda)						
Kampala	0.2789	0.0310	0,0166	0,0131	1,2693	1,1266
Busoga	0.2228	0.0421	0,0120	0,0106	1,1303	1,0632
Bukedi	0.4611	<.0001	0,0115	0,0100	1,1467	1,0709
Bugisu	0.3628	0.0006	0,0111	0,0109	1,0173	1,0086
Bunyoro	0.2561	0.0225	0,0125	0,0107	1,1687	1,0811
Kigezi	0.4082	0.0002	0,0116	0,0101	1,1439	1,0695
Number of household members	-0.1366	<.0001	0,000056	0,000054	1,0273	1,0136
Sex of household head (ref=Female)						
Male	0.1218	0.0081	0,00211	0,00208	1,0132	1,0066
Ever had a terminated pregnancy (ref=Yes)						
No	-0.1331	0.0031	0,00201	0,00198	1,0135	1,0067
The contraceptive method used (ref=Yes)						
No	-1,1418	<.0001	0,0247	0,0202	1,2205	1,1048
Current marital status (ref=Single)						
Married	1.2644	<.0001	0,003192	0,003014	1,0591	1,0291
Living with partner	1.2805	<.0001	0,002777	0,002830	0,9813	0,9906
Number of other partners	0.0555	<.0001	0,000128	0,000110	1,1582	1,0762
Cohabitation duration (ref=0 to 4 years)						
5 to 9 years	-0.4061	<.0001	0,00388	0,00398	0,9748	0,9873
Partner's desire for children (ref=Both want same)						
Don't know	-0.1624	0.0020	0,0028	0,0026	1,0808	1,0396
Partner's education level						
Primary education	-0.1173	0.1011	0,0051	0,0054	0,9386	0,9688
Secondary education	-0.2470	0.0015	0,0060	0,0062	0,9648	0,9822
Higher education	-0.4721	<.0001	0,0085	0,0086	0,9935	0,9968
Don't know	-0.6900	<.0001	0,0164	0,0187	0,8780	0,9370
Partner's occupation (ref=Unemployed)						
Employed	-0.2583	0.0102	0,0101	0,0095	1,0561	1,0277
Woman's earnings compared to partner (ref=Less compared to him)						
More compared to him	0.2141	0.0011	0,0043	0,0043	0,9939	0,9970
About the same	-0.4210	<.0001	0,0025	0,0024	1,0122	1,0061
Partner does not bring in	-0.2745	0.0305	0,0160	0,0175	0,9185	0,9584
Don't know	-0.5402	<.0001	0,0091	0,0094	0,9673	0,9835
The person who usually decides how to spend the woman's earnings (ref=Woman alone)						
Woman and partner	-0.2170	<.0001	0.0017	0.0016	1.0402	1.0199
Partner alone	-0.3914	<.0001	0.0047	0.0048	0.9828	0.9914
Interaction effects						
Woman's age by contraceptive use	0,032	<.0001	0,000020612	0,000020340	1,0133	1,0066

The effects of living in Kampala region have a DEFF=1.2693 and DEFT=1.1266; the standard error and confidence interval would be 1.1266 times greater than they would be for simple random sampling. The effects of living in Busoga region have a DEFF=1.1303 and DEFT=1.0632; the standard error and confidence interval would be 1.0632 times greater than they would be for simple random sampling. The effects of living in Bukedi region have a DEFF=1.1467 and DEFT=1.0709; the standard error and confidence interval would be 1.0709 times greater than they would be for simple random sampling. The effects of living in Busigu region have a DEFF=1.0173 and DEFT=1.0086; the standard error and confidence interval would be 1.0086 times greater than for simple random sampling. The effects of living in Bunyoro region have a DEFF=1.1687 and DEFT=1.0811; the standard error and confidence interval would be 1.0811 times greater than they would be for simple random sampling. The effects of living in Kigezi region have a DEFF=1.1439 and DEFT=1.0695; the standard error and confidence interval would be 1.0695 times greater than they would be for simple random sampling. The effects of the number of members in a household have a DEFF=1.0273 and DEFT=1.0136; the standard error and confidence interval would be 1.0136 times greater than they would be for simple random sampling.

The effects of the household head being male have a DEFF=1.0132 and DEFT=1.0066; the standard error and confidence interval would be 1.0066 times greater compared to simple random sampling. The effects of a woman not having terminated a pregnancy have a DEFF=1.0135 and DEFT=1.0067; the standard error and confidence interval would be 1.0067 times greater than they would be for simple random sampling. The effects of a woman's current marital status being 'married' have a DEFF=1.0591 and DEFT=1.0291; the standard error and confidence interval would be 1.0291 times greater than they would be for simple random sampling. The effects of a woman's current marital status being 'living with a partner' have a DEFF=0.9813 and DEFT=0.9906; the standard error and confidence interval would be 0.9906 times less compared to what they would be for simple random sampling. The effects of a partner having other wives have a DEFF=1.1582 and DEFT=1.0762; the standard error and confidence would be 1.0762 times greater than they would be for simple random sampling. The effects of a cohabitation period of 5-9 years have a DEFF=0.9748 and DEFT=0.9873; the standard error and confidence would be 0.9873 times less than they would be for simple random sampling.

The effects of a woman's partner's desire for children being 'unknown' (don't know) have a DEFF=1.0808 and DEFT=1.0396; the standard error and confidence interval would be 1.0396 times greater than they would be for simple random sampling. The effects of a partner having attained a primary level of education have a DEFF=0.9386 and DEFT=0.9688; the standard error and confidence interval would be 0.9688 times less than for simple random sampling. The effects of a partner having attained a secondary level of education have a DEFF=0.9648 and DEFT=0.9822; the standard error and confidence interval would be 0.9822 times less than they would be for simple random sampling. The effects of a partner having attained a higher level of education have a DEFF=0.9935 and DEFT=0.9968; the standard error and confidence interval would be 0.9968 times less than for simple random sampling.

The effects of a partner's level of education being 'unknown' (don't know) have a DEFF=0.8780 and DEFT=0.9370; the standard error and confidence interval would be 0.9370 times less than they would be for simple random sampling.

The effects of the partner's occupation status being 'employed' have a DEFF=1.0561 and DEFT=1.0277; the standard error and confidence interval would be 1.0277 times greater than they would be for simple random sampling. The effects of a woman earning more compared to her partner have a DEFF=0.9939 and DEFT=0.9970; the standard error and confidence interval would be 0.9970 times less than they would be for simple random sampling. The effects of a woman earning about the same as her partner have a DEFF=1.0122 and DEFT=1.0061; the standard error and confidence interval would be 1.0061 times greater than for simple random sampling. The effects of a woman's partner not bringing home earnings have a DEFF=0.9185 and DEFT=0.9584; the standard error and confidence interval would be 0.9584 times less than they would be for simple random sampling. The effects of a woman's earnings compared to her partner being unknown (don't know) have a DEFF=0.9673 and DEFT=0.9673; the standard error and confidence interval would be 0.9673 times less than they would be for simple random sampling. The effects of both a woman and her partner usually deciding how to spend her earnings have a DEFF=1.0402 and DEFT=1.0199; the standard error and confidence interval would be 1.0199 times greater than they would be for simple random sampling. The effects of a woman's partner alone usually deciding how to spend her earnings have a DEFF=0.9828 and DEFT=0.9914; the standard error and confidence interval would be 0.9914 times less than they would be for simple random sampling.

Interaction Effects

The effects of age and contraceptive use have a DEFF=1.0133 and DEFT=1.0066. The standard error and confidence interval would be 1.0066 times greater than for simple random sampling.

3.13 Summary

The section above was based on the 2016 South Africa Demographic and Health Survey, 2015-2016 Tanzania Demographic and Health Survey, and the 2016 Uganda Demographic and Health Survey data. In it, survey logistic regression was used to identify the critical determinants of intimate partner violence amongst women of reproductive age in South Africa, Tanzania, and Uganda. The study's findings showed that in the three different countries, some determinants are common: partner's alcohol drinking status, whether the woman's father ever beat her mother, wife-beating attitudes, woman's current age, region where the woman resides, number of household members, sex of the household head, wealth index, pregnancy termination status, contraceptive method used, current marital status, cohabitation duration, and woman's occupation status.

Some determinants are only significant in two of the countries and insignificant in the other. South Africa and Tanzania have the following determinants in common: woman's education level. For South Africa and Uganda: media exposure, partner's desire for children, and woman's earnings compared to her partner. For Tanzania and Uganda: number of other partners and the person who usually decides how to spend the woman's earnings. South Africa has determinants that are significant to intimate partner violence, present in that country alone: body mass index, partner's age, and the knowledge of sexually transmitted infections. Uganda is the only country where the partner's education is significant to intimate partner violence. In the following chapter, we will look at how the generalized linear mixed models (GLMMs) accounts for variability within the sampling units.

Chapter 4

Generalized Linear Mixed Models

In ., we looked at logistic and survey logistic regression models, and how these models assume all the variables effects to be fixed effects. The weakness of generalized linear models is that random effects are not included, although, there are circumstances in which the effect of the predictor is random (Habyarimana, 2016). For instance, in this study we have worked with the DHS dataset in which the clusters are deemed to be a random effect. Therefore, in this chapter we use GLMMs which gives an option to incorporate random effects.

4.1 Introduction

The Generalized Linear Models (GLMs) explored in . might not be appropriate enough for the data we interested in. Logistic regression models are in the family of GLMs in which the complexity of the survey design is disregarded, in the sense that random effects are ignored (Dlamini, 2016). The generalized linear mixed models (GLMMs), however, include random effects in the analysis. These models are helpful since the models combine linear mixed models (fixed and random effects inclusive) and generalized linear mixed models. They also handle a range of response distributions and data with values sampled in some group structures, instead of entirely independently (Waagepetersen, 2007; Dlamini, 2016). GLM allows the modeling of different kinds of responses such as binary ones (McCullagh & Nelder, 1989; Dlamini, 2016), and linear mixed models (LMMs) are known to incorporate random effects.

Generalized linear mixed models are an extension of linear mixed models, with a relaxing of some of the assumptions of LMMs. They present all the advantages of logistic regression such as information on sample size (Dlamini, 2016). Furthermore, they are able to do one analysis with all random effects, and cater for the binary response variable (Dlamini, 2016). The greatest benefit of GLMMs is their capability to handle unbalanced data due to missing values (Manning, 2007; Dlamini, 2016).

4.2 Review of Linear Mixed Models

The generalized linear models discussed in . do not account for the random effect (Dlamini, 2016). It is therefore essential to expand the model from equation (4.1):

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{4.1}$$

to yield equation (4.2):

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{ZU} + \boldsymbol{\epsilon} \tag{4.2}$$

where $\boldsymbol{Y}$ is the $n \times 1$ vector of responses,
$\boldsymbol{X}$ is a $n \times (p+1)$ design matrix for fixed effects,
$\boldsymbol{\beta}$ is a $(p \times 1) \times 1$ vector of unknown fixed-effects parameters,
$\boldsymbol{Z}$ is a $n \times q$ design matrix for random effects,
$\boldsymbol{U}$ is a $q \times 1$ vector of unknown random-effects parameters, and
$\boldsymbol{\epsilon}$ is a $n \times 1$ vector of error terms which have a multivariate normal distribution, with mean vector **0,** and variance-covariance matrix **R,** i.e., $\boldsymbol{\epsilon} \sim N_{\boldsymbol{n}}(\boldsymbol{0}, \boldsymbol{R})$ (Dlamini, 2016)**.**

Given the nature of random effect hypothesis, $\boldsymbol{U}$ is treated differently from $\boldsymbol{\beta}$ (Dlamini, 2016). Statistical linear mixed models state that recorded data consist of two parts: fixed and random effects (Littell, et al., 2000; Dlamini, 2016). Fixed effects are described as the expected value of the observation, and the random effects are described as the variance and covariance of the observation (Littell, et al., 2000; Dlamini, 2016). The assumption made is that observations from the same unit are correlated; hence linear mixed models tackle covariation between measures on the same unit (Kincaid, 2005; Dlamini, 2016). The variance $V(\boldsymbol{Y})$ of the model is represented by equation (4.3) which is known as the modeling covariance structure (Dlamini, 2016). The specification of the covariance structure for a mixed model is done through $\boldsymbol{G}$ and $\boldsymbol{R}$ (Dlamini, 2016).

$$V(\boldsymbol{Y}) = \boldsymbol{ZGZ}' + \boldsymbol{R} \tag{4.3}$$

where $\boldsymbol{ZGZ'}$ constitutes the between-subject portion of the covariance structure, and $\boldsymbol{R}$ constitutes the within-subject portion (Dlamini, 2016).

In linear mixed models with greater than one random effect, this is assumed to come from a multivariate normal distribution with mean **0,** and variance-covariance matrix $\boldsymbol{G}$ (Dlamini, 2016). Instead, the variance elements are estimated. The diagonal elements of matrix $\boldsymbol{G}$ are the variance component for each random effect, while the off-diagonal components are the covariance between different dimensions (Dlamini, 2016). Suppose there is one random effect in the model, then $\boldsymbol{G}$ will have only one element:

the variance element of the random effects (Dlamini, 2016). For greater than one random effect, $\boldsymbol{G}$ will be a $k \times k$ matrix for k random effects (Dlamini, 2016).

4.3 Generalized Linear Mixed Models

4.3.1 Model Formulation

Suppose the normality assumption of $f(\boldsymbol{Y}|\boldsymbol{\theta})$ is now relaxed. Assume that $\boldsymbol{Y}$ and $\boldsymbol{\theta}$ are independent and $f(\boldsymbol{Y}|\boldsymbol{\theta})$ is a member of the exponential family of distribution (McCullagh & Nelder, 1989; Dlamini, 2016).

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = exp\left\{\frac{y_t\theta_t - b(\theta_t)}{\phi} - c(y_t, \phi)\right\} \tag{4.4}$$

where ϕ is the scale parameter. Based on the model, the conditional y related to θ_t is given as follows (Dlamini, 2016):

$$E(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{\partial b(\theta_t)}{\partial \theta_t}.$$

The model with both effects is given by:

$$g(\theta_t) = \boldsymbol{X}_t'\boldsymbol{\beta} + \boldsymbol{Z}_t'\boldsymbol{U}_t \tag{4.5}$$

where, $\eta_t = g(\theta_t)$, g is the link function, and $\boldsymbol{U}_t$ a vector of random effects (Dlamini, 2016). In this study, intimate partner violence status is either 0 (have not experienced IPV) or 1 (have experienced IPV). Thus, we utilize the logistic regression where we regard $g(.)$ as the logit link, with $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t (t = 1,2, \dots, n)$ being p-dimension and q-dimension vectors of known covariate values, while $\boldsymbol{\beta}$ is a p-dimension vector of unknown fixed effects regression coefficients (Dlamini, 2016).

4.3.2 Inverse Link Function

To map the value of the linear predictor for observations t, η_t to the conditional mean of observation t, μ_t, we need to use the link function. Both μ_t and η_t are the scalars for inverse link function on a one-to-one basis. The binomial distribution inverse link function is given by:

$$h(\eta) = \frac{e^{\eta}}{1 + e^{\eta}}.$$

Univariate link and inverse link functions are increasing monotonic functions (Kachman, 2000; Moeti, 2010). Increasing the linear predictor results in an increase in the conditional mean, but it is not at a constant rate (Kachman, 2000; Moeti, 2010). The selection of an inverse link function is based on the error distribution (Kachman, 2000; Moeti, 2010).

Mean of Y

The mean of Y is given by: $E[y_t] = E\big[E[y_t|b]\big] = E[\mu_t] = E[g^{-1}(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})]$. Since $g^{-1}(.)$ is a non-linear function, the mean cannot be simplified. For a certain $g(.)$, if we have a log link $g(\mu) = \log(\mu)$ and $g^{-1}(x) = \exp(x)$. Then:

$$E[y_t] = E[exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})] = exp(\boldsymbol{x}_t'\boldsymbol{\beta})\, E[exp(\boldsymbol{z}_t'\boldsymbol{b})] = exp(\boldsymbol{x}_t'\boldsymbol{\beta})\, M_u(\boldsymbol{z}_t), \qquad (4.3.1)$$

where $M_u(\boldsymbol{z}_t)$ is the moment generating function of **b,** evaluated at $\boldsymbol{z}_t$ (Mculloch, et al., 2008). Suppose $\mu_t \sim N(0, \sigma_\mu^2)$, and that each of the rows of **Z** only has a single non-zero entry equal to 1. Then $M_u(\boldsymbol{z}_t) = \exp\left(\frac{\sigma_\mu^2}{2}\right)$ and $E[y_t] = \exp(\boldsymbol{x}_t'\boldsymbol{\beta})\exp\left(\frac{\sigma_\mu^2}{2}\right)$ or $logE[y_t] = \boldsymbol{x}_t'\boldsymbol{\beta} + \frac{\sigma_\mu^2}{2}$ (Mculloch, et al., 2008).

Variances

The marginal variance of Y is derived by:

$$var(y_t) = var(E[y_t|b]) + E[var(y_t|b)],$$

$$= var(\mu_t) + E[\tau^2 v(\mu_t)],$$

$$= var(g^{-1}[\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b}]) + E[\tau^2 v(g^{-1}[\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b}])].$$

We have to make assumptions about $g(\cdot)$ or the conditional distribution of y. Therefore, we assume that we have a log link and that **y** is conditionally independent, given **b** with a Poisson distribution. Applying these to equation (4.3.1) for the mean, we get:

$$var(y_t) = var(\mu_t) + E[\mu_t],$$

$$var(y_t) = var[\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})] + E[\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})],$$

$$= E\big[\big(\exp\big(2(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})\big)\big)\big] - [E(\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b}))]^2 + E[\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b})],$$

$$= \exp(2\boldsymbol{x}_t'\boldsymbol{\beta})(\, M_u(2\boldsymbol{z}_t)) - [M_u(2\boldsymbol{z}_t)]^2 + \exp(-\boldsymbol{x}_t'\boldsymbol{\beta})\, M_u(\boldsymbol{z}_t).$$

Furthermore, if we assume that $\mu_t \sim N(0, \sigma_\mu^2)$, and that each row of **Z** has only a single non-zero entry equal to 1 then:

$$var(y_t) = \exp(2\boldsymbol{x}_t'\boldsymbol{\beta})\left(\exp(2\sigma_\mu^2) - \exp(\sigma_\mu^2)\right) + \exp(\boldsymbol{x}_t'\boldsymbol{\beta})\exp\left(\frac{\sigma_\mu^2}{2}\right),$$

$$= E[y_t](\exp(\boldsymbol{x}_t'\boldsymbol{\beta})\left[\exp\left(\frac{3\sigma_\mu^2}{2}\right) - \left(\exp\left(\frac{\sigma_\mu^2}{2}\right)\right] + 1\right).$$

If the value of $var(y_t)$ is greater than 1, the variance will be larger than the mean (Moeti, 2010). The marginal distribution will always be over-dispersed, compared to the conditional distribution of y_t given **b,** which is Poisson. Thus, random effects can model over-dispersion to a particular source (Mculloch, et al., 2008).

Variance Function

A function that is used to model non-systematic variability is the variance function. The two sources in which residual variability arises, are the sampling distribution or over-dispersion described in the previous section. An example of the sampling distribution is a Poisson random variable with mean μ and variance μ (Kachman, 2000; Moeti, 2010). The variance function of a binomial distribution is $\frac{\mu(1-\mu)}{n}$. A first approach used to model the over-dispersion is to scale the residual variability as $var(y_t|\boldsymbol{\mu}) = \phi v(\mu_t)$, where ϕ is an over-dispersion parameter. Secondly, we can add a random effect, that is: $e_t \sim N(0, \phi)$, to the linear predictor for each value. As a third approach, we can choose another distribution: a two-parameter (μ, ϕ) negative binomial distribution, which can be utilized instead of a one parameter μ Poisson distribution for count data (Moeti, 2010). These approaches all use the estimation of an additional parameter ϕ (Kachman, 2000; Moeti, 2010). The variance function $v(\mu, \phi)$ is utilized to model the residual variability (Moeti, 2010). The choosing of the variance function is influenced by the error distribution which was selected (Moeti, 2010). There is a need to account for an over-dispersion parameter, since the observed residual variability is usually more significant than expected, due to sampling (Kachman, 2000; Moeti, 2010).

Covariance and Correlations

Utilizing the random effects brings in a correlation among values that have any random effect in common (Moeti, 2010). If we assume that the elements of y are conditionally independent, then:

$$cov(y_t, y_m) = cov(E[y_t|b], [y_m|b]) + E[cov(y_t, y_m|\boldsymbol{b}],$$

$$= cov(\mu_t, \mu_m) + E[0],$$

$$= cov(g^{-1}[\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b}], g^{-1}[\boldsymbol{x}_m'\boldsymbol{\beta} + \boldsymbol{z}_m'\boldsymbol{b}]).$$

Introducing the log link, we evaluate the equation as:

$$cov(y_t, y_m) = cov(\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{b}), \exp(\boldsymbol{x}_m'\boldsymbol{\beta} + \boldsymbol{z}_m'\boldsymbol{b})),$$

$$= \exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_m'\boldsymbol{\beta})\, cov(\exp(\boldsymbol{z}_t'\boldsymbol{b}), \exp(\boldsymbol{z}_m'\boldsymbol{b})),$$

$$= \exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_m'\boldsymbol{\beta})\, [M_u(\boldsymbol{z}_t + \boldsymbol{z}_m) - M_u(\boldsymbol{z}_t)M_u(\boldsymbol{z}_m)].$$

Assuming $\boldsymbol{b} \sim N\left(0, \boldsymbol{I}\sigma_\mu^2\right)$ and that each row of $\mathbf{Z}$ has a single non-zero entry equal to 1. Then, $cov(y_t, y_m) = \exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_m'\boldsymbol{\beta})\left[\exp\left(\sigma_\mu^2\right)\left(\exp\left(\boldsymbol{z}_t'\boldsymbol{z}_m\sigma_\mu^2\right) - 1\right)\right] cov(y_t, y_m) = 0$ if $\boldsymbol{z}_t'\boldsymbol{z}_m = 0$. This is only possible if the two observed values do not share a random effect, and are positive otherwise, that is: $\boldsymbol{z}_t'\boldsymbol{z}_m = 1$ (Mculloch, et al., 2008; Moeti, 2010). When $\boldsymbol{z}_t'\boldsymbol{z}_m = 1$, we can calculate the correlation. We need to cancel $\exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_m'\boldsymbol{\beta})$ in the numerator and denominator so that:

$$corr(y_t, y_m) = \frac{e^{2\sigma_\mu^2} - e^{\sigma_\mu^2}}{\sqrt{\left(e^{2\sigma_\mu^2} - e^{\sigma_\mu^2} + e^{-\boldsymbol{x}_t'\boldsymbol{\beta} + \frac{\sigma_\mu^2}{2}}\right)\left(e^{2\sigma_\mu^2} - e^{\sigma_\mu^2} + e^{-\boldsymbol{x}_m'\boldsymbol{\beta} + \frac{\sigma_\mu^2}{2}}\right)}},$$

$$= \frac{1}{\sqrt{(1 + \eta e^{-\boldsymbol{x}_t'\boldsymbol{\beta}})(1 + \eta e^{-\boldsymbol{x}_m'\boldsymbol{\beta}})}},$$

where η is given by $\frac{1}{\left(e^{3\frac{\sigma_\mu^2}{2}} - e^{\frac{\sigma_\mu^2}{2}}\right)}$ (Mculloch, et al., 2008).

4.4 Maximum Likelihood Estimation

In linear mixed models, the marginal distribution of Y can be computed as the multivariate normal, meaning $f(Y)$ is a density function of a multivariate normal distribution (Bolker, et al., 2009; Vittinghoff, et al., 2011; Dlamini, 2016). For generalized linear mixed models, it is not easy to evaluate the integral because N is a q-dimensional integral over the random effects (Vittinghoff, et al., 2011). The random effect model could be fitted by maximization of the marginal likelihood, and that is obtained by integrating out the random effects (Vittinghoff, et al., 2011). The likelihood is given by

$$\begin{aligned} L(\beta, G, \phi) &= \prod_{t=1}^{N} f_t(\boldsymbol{Y}_t | \boldsymbol{\beta}, \boldsymbol{G}, \phi) \\ &= \prod_{t=1}^{N} f_t(\boldsymbol{Y}_t | \boldsymbol{\beta}, \boldsymbol{G}, \phi) \,.\, f(\boldsymbol{U}_t, \boldsymbol{G}) d\boldsymbol{u}_t \end{aligned} \tag{4.6}$$

where $f_t(\boldsymbol{Y}_t | \boldsymbol{\beta}, \boldsymbol{G}, \phi) = \int \prod_{m=1}^{n_t} f_{tm}(\boldsymbol{Y}_{tm} | \boldsymbol{\beta}, \boldsymbol{G}, \boldsymbol{\phi}).f(\boldsymbol{U}_t, \boldsymbol{G}) d\boldsymbol{u}_t$ (Molenberghs & Verbeke, 2006). In general, numerical approximations have to be used in evaluating the likelihood of GLMMs.

4.4.1 Estimation: Approximation of the Integrand

The Laplace method approximates the integrand and is one of the natural alternatives, when the exact likelihood function is hard to compute (Molenberghs & Verbeke, 2006; Dlamini, 2016) . When the integrands are approximated, the objective is to get traceable integrals such that closed-form expressions can be acquired (Dlamini, 2016), making numerical maximization of the approximated likelihood feasible (Molenberghs & Verbeke, 2006; Dlamini, 2016). Let us assume that we want to approximate the integral of the form:

$$I = \int e^{(-q(x))} dx \tag{4.7}$$

where $q(.)$ is a well-behaved function, in a way that its minimum value is at $x = \bar{x}$ with $q'(\bar{x}) = 0$ and $q''(\bar{x}) > 0$, so we can consider the Taylor expansion about $\bar{x}$ given by:

$$q(x) \approx q(\bar{x}) + \frac{1}{2} q''(\bar{x})(x - \bar{x}) + \cdots \tag{4.8}$$

This gives an approximation to (4.7) as:

$$\int e^{(-q(x))} dx \approx \sqrt{\frac{2\pi}{q''(\bar{x})}} e^{(-q(\bar{x}))}. \tag{4.9}$$

We may also have the multivariate extension of (4.9), which is often helpful. Let $q(\alpha)$ be a well-behaved function with its minimum at $\alpha = \bar{\alpha}$ with $q'(\bar{\alpha}) = 0$ and $q''(\bar{\alpha}) > 0,$ where q' and q'' are the gradient and Hessian of *q* respectively. We have:

$$\int e^{(-q(x))} dx \approx c |q''(\bar{x})|^{-\frac{1}{2}} e^{(-q(\bar{x}))} \tag{4.10}$$

where c is the constant that depends on the dimension of the integral, $|q''(\bar{x})|$ is the determinant of the matrix $q''(\bar{x})$ and in which $q''(\bar{x}) > 0$ implies matrix $q''(\bar{x})$ is positive definite (Dlamini, 2016).

Laplace Approach

The Laplace approximation is considered to be the greatest convenient approach to approximate integrals (Tierny & Kadane, 1986; Habyarimana, 2016) and is of the form:

$$I = \int e^{K(a)} db \tag{4.11}$$

where $K(a)$ is a known unimodal and bounded function of a *k*-dimensional variable *a* (Habyarimana, 2016). Consider $\hat{a}$ as the value of *a* for which *K* is minimized (Habyarimana, 2016). The second-order Taylor series expansion of $K(a)$ around $\hat{a}$ and can be written as follows:

$$K(a) \approx K(\hat{a}) + \frac{1}{2}(a - \hat{a})' K''(\hat{a})(a - \hat{a}) \tag{4.12}$$

where $K''(\hat{a})$ equals the Hessian of *K* (Habyarimana, 2016). That means the matrix of the second-order derivative of K, evaluated at $\hat{a}$ (Habyarimana, 2016). ***I*** can be approximated by replacing *K(a)* in (4.*11*) by its value in (4.*12*), and it becomes:

$$I \approx (2\pi)^{\frac{1}{2}} |-K''(\hat{a})|^{-\frac{1}{2}} e^{K(\hat{a})} \tag{4.13}$$

The Laplace approximation is exact if *K(a)* is a quadratic function of *a* (Habyarimana, 2016).

When the Laplace approximation method fails, we revert to numerical integration, which proves to be useful. The Gaussian and adaptive Gaussian quadrature are largely designed for the approximation of integrals of the following form (Habyarimana, 2016):

$$\int f(u)c(u)du \tag{4.14}$$

for known function *f(u)* and *c(u)*, the density of multivariate or univariate standard normal distribution (Habyarimana, 2016). Random effects need to be standardized to get an identity covariance matrix (Habyarimana, 2016). If we let δ_t be equal to $\delta_t = D^{-\frac{1}{2}}b_t$, then δ_t is normally distributed with mean equal to zero, and covariance I; the linear predictor then becomes $\theta_{tj} = x'_{tj}\beta + z_{tj}D^{-\frac{1}{2}}\delta_t$. The variance component in D is now in the linear predictor (Habyarimana, 2016). The likelihood contribution for subject *t* is given by the following:

$$f_t(y_t|\beta, D, \phi) = \int \prod_{t=1}^{n_t} f_{tj}\left(y_{tj}|b_t, \beta, \phi\right) f(b_t, D) db_t \tag{4.15}$$

$$= \int \prod_{t=1}^{n_t} f_{tj}\left(y_{tj}|\delta_t, \beta, \phi\right) f(\delta_t, D) d\delta_t \tag{4.16}$$

where the random effects b_t, with mean 0 and covariance D, is assumed to be normally distributed (Molenberghs & Verbeke, 2005; Habyarimana, 2016). Expression (4.16) is of the form (4.14), required to apply the Gaussian quadrature (Molenberghs & Verbeke, 2005; Habyarimana, 2016).

Gaussian Quadrature

A classical Gaussian quadrature approximates an integral of the form (4.14) by the weighted sum (Habyarimana, 2016):

$$\int f(u)c(u)du \approx \sum_{k=1}^{K} w_k f(z_k) \tag{4.17}$$

where K is the order of the approximation; as K increases, the accuracy of the approximation also increases (Habyarimana, 2016). Additionally, solutions of the K^{th} order Hermite polynomial are z_k, and the corresponding weights are w_k. In case of univariate integration, the approximation involves subdividing the integration region into intervals and approximating rectangles (Habyarimana, 2016).

Adaptive Gaussian Quadrature

The quadrature points are scaled and centered as if *f(u)c(u)* were a normal distribution (Habyarimana, 2016). The mean of the distribution would then be $\hat{u}$ of $ln[f(u)c(u)]$, and the corresponding variance would be:

$$\left[-\frac{\partial^2}{\partial z^2} ln[f(u)c(u)]|_{z=\hat{z}}\right]^{-1} \tag{4.18}$$

From there, the new quadrature points are given by:

$$z_k^* = \hat{z} + \left[-\frac{\partial^2}{\partial z^2} ln[f(u)c(u)]|_{z=\hat{z}}\right]^{-\frac{1}{2}} z_k \tag{4.19}$$

with corresponding weights:

$$w_k^* = \left[-\frac{\partial^2}{\partial z^2} ln[f(u)c(u)]|_{z=\hat{z}}\right]^{-\frac{1}{2}} \times \frac{c(z_k^*)}{z(z_k)} w_k \tag{4.20}$$

The integral is now approximated by:

$$\int f(u)c(u)du \approx \sum_{k=1}^{K} w_k^* f(z_k^*) \tag{4.21}$$

Adaptive Gaussian quadrature requires fewer quadrature points compared to the classical Gaussian quadrature (Molenberghs & Verbeke, 2005; Habyarimana, 2016). The adaptive Gaussian quadrature requires the calculation of $\hat{z}$ for each unit in the dataset, but the numerical maximization of N functions of the form (4.12) makes Gaussian quadrature time consuming (Molenberghs & Verbeke, 2005; Habyarimana, 2016).

4.4.2 Estimation: Approximate of Data

Another class of estimation approach is based on decomposing the data into the mean and error terms (Dlamini, 2016). The Taylor series expansion of the mean is a non-linear function of predictors (Dlamini, 2016). The method in this class differs in the order of the Taylor approximation (Dlamini, 2016). The decomposition is considered as:

$$Y_{tm} = \mu_{tm} + \epsilon_{tm} = h(X'_{tm}\beta + Z'_{tm}U) + \epsilon_{tm} \tag{4.22}$$

where $h(.)$ is the inverse link function, and the error term has an appropriate distribution with variance equal to $var(Y_{tm}|U_t) = \phi V(\mu_{tm})$. Here, $V(.)$ is the usual variance function in the exponential family (Molenberghs & Verbeke, 2006; Dlamini, 2016). Consider a binary outcome with a logit link function (Dlamini, 2016). One has:

$$\mu_{tm} = h(X'_{tm}\beta + Z'_{tm}U) = P_{tm} = \frac{\exp(X'_{tm}\beta + Z'_{tm}U)}{1 + \exp(X'_{tm}\beta + Z'_{tm}U)} \tag{4.23}$$

where $h(X'_{tm}\beta + Z'_{tm}U)$ is the inverse of the logit link function, which is the logistic function. $x_t$and z_t are as in the definition of the generalized linear mixed model. This is considered as a particular case of GLMM, where the exponential family in Bernoulli and corresponding link function is $g(\mu) = logit(\mu)$ (Dlamini, 2016).

4.4.3 Penalized Quasi-Likelihood

Penalized quasi-likelihood (PQL) is one of the methods that approximates data by mean plus error term, with variance equals to $Var(Y_{tm}|U_t)$ (Moeti, 2010; Dlamini, 2016). This method uses the Taylor expansion around estimates $\hat{\beta}$ and $\widehat{U}$ of fixed effects and random effects, respectively (Bolker, et al., 2009; Moeti, 2010; Dlamini, 2016). Then we have the following:

$$\begin{aligned} Y_{tm} &= \mu_{tm} + \epsilon_{tm} = h(X'_{tm}\beta + Z'_{tm}U) + \epsilon_{tm} \\ &\approx h\left(X'_{tm}\hat{\beta} + Z'_{tm}\widehat{U}\right) + h\left(X'_{tm}\hat{\beta} + Z'_{tm}\widehat{U}\right)X'_{tm}\left(\beta - \hat{\beta}\right) \\ &\quad + h\left(X'_{tm}\hat{\beta} + Z'_{tm}\widehat{U}\right)Z'_{tm}\left(U - \widehat{U}\right) + \epsilon_{tm} \\ &= \hat{\mu}_{tm}V(\hat{\mu}_{tm})X'_{tm}\left(\beta - \hat{\beta}\right) + V(\hat{\mu}_{tm})Z'_{tm}\left(U - \widehat{U}\right) + \epsilon_{tm}, \end{aligned} \tag{4.24}$$

and:

$$\boldsymbol{Y}_t = \widehat{\boldsymbol{\mu}}_t + \widehat{\boldsymbol{V}}_t\boldsymbol{X}_t(\beta - \hat{\beta}) + \widehat{\boldsymbol{V}}_t\boldsymbol{Z}_t\left((U) - \widehat{U}\right) + \epsilon_t$$

where $\widehat{\boldsymbol{\mu}}_t$ contains values of $\widehat{\boldsymbol{\mu}}_{tm} = h\left(X'_{tm}\hat{\beta} + Z'_{tm}\widehat{U}\right)$, $\boldsymbol{V}_t$ is the diagonal matrix with elements $V(\hat{\mu}_{tm}) = h\left(X'_{tm}\hat{\beta} + Z'_{tm}\widehat{U}\right)$, and $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$ contain the X'_{tm} and Z'_{tm}, respectively. Re-ordering the above expression and pre-multiplying by $\widehat{\boldsymbol{V}}_t^{-1}$, we obtain:

$$\begin{aligned} Y_t^* &= \widehat{\boldsymbol{V}}_t^{-1}\left(\boldsymbol{Y}_t - \widehat{\boldsymbol{\mu}}_t\right) + \boldsymbol{X}_t\hat{\beta} + \boldsymbol{Z}_t\widehat{\boldsymbol{U}} \\ &\approx \boldsymbol{X}_t\hat{\beta} + \boldsymbol{Z}_t\widehat{\boldsymbol{U}} + \boldsymbol{\epsilon}_t^* \end{aligned} \tag{4.25}$$

where $\boldsymbol{\epsilon}_t^*$ is equal to $\widehat{\boldsymbol{V}}_t^{-1}\boldsymbol{\epsilon}_t^*$ and has a zero mean (Dlamini, 2016). This can be viewed as a linear mixed model for a pseudo data $\boldsymbol{Y}_t^*$ with error term $\boldsymbol{\epsilon}_t^*$ (Dlamini, 2016).

4.4.4 Marginal Quasi-Likelihood

Marginal quasi-likelihood (MQL) is an approximation method that is similar to the PQL method (Dlamini, 2016). However, it is based on a linear Taylor expansion of the mean around the current estimate $\hat{\beta}$ for fixed effects, and around $\boldsymbol{U} = \boldsymbol{0}$ for random effects (Bolker, et al., 2009; Moeti, 2010; Dlamini, 2016). This gives the same expansion as shown for PQL, but now the current predictor is of the form $h(X'_{tm}\hat{\beta})$ (Dlamini, 2016). The pseudo data is now of the form:

$$\boldsymbol{Y}_t^* = \widehat{\boldsymbol{V}}_t^{-1}\left(\boldsymbol{Y}_t - \widehat{\boldsymbol{\mu}}_t\right) + \boldsymbol{X}_t\hat{\beta} \tag{4.26}$$

and satisfy the approximate linear mixed model:

$$\boldsymbol{Y}_t^* \approx \boldsymbol{X}_t\hat{\beta} + \boldsymbol{Z}_t\widehat{\boldsymbol{U}} + \boldsymbol{\epsilon}_t^* \tag{4.27}$$

The model fitting is also done by iteration between the calculation of the pseudo data, and fitting of the approximate linear mixed model for this pseudo data (Molenberghs & Verbeke, 2006; Dlamini, 2016). The resulting estimates are known as quasi-likelihood estimates (MQL) (Dlamini, 2016).

4.4.5 Discussion of MQL and PQL

There is no principal difference between penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL); they both do not include the random $\boldsymbol{U}_t$ in the linear predictor (Bolker, et al., 2009; Dlamini, 2016). Both of these methods are based on similar ideas and will have almost similar properties. However, the accuracy of both models depends on the accuracy of the linear mixed model for pseudo data $\boldsymbol{Y}_t^*$ (Dlamini, 2016). The Laplace method, PQL, and MQL perform poorly in binary cases with a small number of repeated observations available (Molenberghs & Verbeke, 2006). The MQL completely ignores the variability of the random effect in the linearization of the mean (Dlamini, 2016). The Laplace method is more accurate compared to penalized quasi-likelihood. However, Laplace is slower and less flexible compared to penalized quasi-likelihood (Bolker, et al., 2009). The MQL remains biased, while PQL will be consistent using an increased number of measurements.

4.5 Inference

Fitting the GLMM is largely based on maximum likelihood principles. Therefore, inferences for the parameters are obtained from standard maximum likelihood theory (Molenberghs & Verbeke, 2005; Habyarimana, 2016). If the model fitted is appropriate, then the obtained estimators are asymptotically normally distributed with the correct values as means, and with the inverse Fisher information matrix as a covariance matrix (Habyarimana, 2016). Therefore, the Wald-type test, comparing standardized estimates to the standard normal distribution, can be used (Habyarimana, 2016). Alternatively, the likelihood ratio test and score tests can also be used (Habyarimana, 2016).

The inference on the fixed effects is made using the Wald test, the approximate t-tests, and F-tests (Verbeke & Molenberghs, 2000; Habyarimana, 2016). The approximate Wald test is acquired from approximating the distribution of $\frac{(\hat{\beta}_m-\beta_m)}{s.e(\hat{\beta}_m)}$ by a standard univariate normal distribution of each parameter β_m in $\beta, m = 1,2,\dots,p$ (Habyarimana, 2016). Generally, it may be of interest to construct confidence intervals and tests of hypotheses about certain linear combinations of the component β (Habyarimana, 2016). The likelihood ratio test results are valid if the model is fitted using ML, and not valid when REML is used (Habyarimana, 2016). The REML log-likelihood functions are based on observations, making them no longer comparable (Verbeke & Molenberghs, 2000; Habyarimana, 2016).

4.6 Generalized Linear Models Applied to Binary Outcomes

The mixed-effects logistic regression model is commonly chosen for analyzing multilevel dichotomous data and is the most used in GLMM (Habyarimana, 2016). In the GLMM setting, this model uses the logit link and is given by (Habyarimana, 2016):

$$g(\mu_{tmk}) = logit(\mu_{tmk}) = log\left[\frac{\mu_{tmk}}{1-\mu_{tmk}}\right] = \eta_{tmk} \tag{4.28}$$

The conditional expectation $\mu_{tmk} = E(Y_{tmk}|b_t, x_t)$ equals $P(Y_{tmk}|b_t, x_{tmk})$, namely, the conditional probability of the response given the random effects and the covariate values, with Y_{tmk} the t^{th} response in the m^{th} household, with k^{th} primary sampling unit (Habyarimana, 2016). This model can also be written as:

$$P(Y_{tmk}|b_t, x_{tmk}, z_{tmk}) = g^{-1}(\eta_{tmk}) \tag{4.29}$$

where $g^{-1}(\eta_{tmk})$ is commonly known as the logistic cumulative distribution function and is given by (Habyarimana, 2016):

$$g^{-1}(\eta_{tmk}) = [1 + \exp(-\eta_{tmk})]^{-1}.$$

4.7 Application of Generalized Linear Mixed Models to the Data from South Africa, Tanzania, and Uganda

After checking for collinearity, the parameters were found to not have a significant correlation and hence the application of GLMM. The other covariates are controlled, using the DHS survey clusters as random effects. GLMM allows us to model a binary response variable, the IPV, and take random effects such as a survey cluster. To check relationships between predictors, we assessed the covariates used in the logistic regression model, together with the IPV response in a GLMM. The statistical models are fitted well by the GLIMMIX procedure applied to the data, with correlations or non-constant variability. In this procedure, the response does not need to be normally distributed, and allows different estimation methods to be specified. The model fitted used the Laplace estimation method. The random effect was the clusters.

4.7.1 Model Fitting

South Africa

In Table 4.1, the ratio of generalized chi-square statistic and its degrees of freedom is 0.94, which is close to 1. This value is the measure of the residual variability in the marginal distribution of the data. The value indicates that variability has adequately been modeled, and there is no residual over-dispersion.

Table 4.1: Fit statistics for conditional distribution

-2 log L (IPV | r. effects)	8141.35
Pearson chi-square	7986.04
Pearson chi-square / DF	0.94

The type 3 tests of fixed effects for the model fitted using the Laplace method in GLMMs, are shown in Table 4.2. The F-statistics used for the significant test for the fixed effects and corresponding p-value, show that some effects are essential. Some are not in the fitted model at a 5% significance level: type of residence, woman's education level, literacy of the woman, number of other partners, partner's education level, partner's occupation, and the person who usually decides how to spend a woman's earnings.

Table 4.2: Type III tests of fixed effects

Effect	F-value	DF	P-value
Partner drinks alcohol	128.09	2	<.0001
Woman's father ever beat her mother	55.99	2	<.0001
Access to the media	5.34	2	0.0048
Wife-beating attitude	24.29	2	<.0001
Woman's current age	34.16	1	<.0001
Region	5.59	8	<.0001
Type of residence	0.15	1	0.6972
Woman's education level	2.34	3	0.0713
Number of household members	171.51	1	<.0001
Sex of the household head	10.96	1	0.0009
Literacy	0.44	1	0.5066
Wealth index	17.05	4	<.0001
Ever had a terminated pregnancy	19.52	1	<.0001
Contraceptive method used	48.40	1	<.0001
Body Mass Index	10.88	1	0.0010
Current marital status	31.64	2	<.0001
Number of other partners	0.17	1	0.6826
Cohabitation duration	4.79	1	0.0286
Partner's desire for children	15.86	3	<.0001
Partner's education level	0.45	4	0.7750
Partner's occupation	0.25	1	0.6187
Woman's occupation	6.82	2	0.0011
Partner's age	8.90	1	0.0029
The person who usually decides on how to spend a woman's earnings	0.53	1	0.4665
Woman's earnings compared to partner	3.21	4	0.0120
Knowledge of Sexually Transmitted Infections (STIs)	11.83	1	0.0006

Continued…

Interaction effects			
Woman's age by contraceptive use	27.14	1	<.0001
Wealth index by contraceptive use	3.07	4	0.0154

***DF=Degrees of Freedom**

Table 4.3 shows that a woman whose partner does not drink alcohol is 0.44 (OR=0.435, p-value<.0001) times less likely to experience IPV, compared to a woman whose partner does drink it. A woman who has never witnessed her father beating her mother is 0.43 (OR=0.432, p-value<.0001) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who does not know if her father beats her mother or not is 0.73 (OR=0.727, p-value=0.0221) times less likely to experience IPV, compared to a woman who has witnessed her father beating her mother. A woman who views wife-beating as acceptable is 1.83 (OR=1.833, p-value<.0001) times more likely to experience IPV, compared to a woman who views wife-beating as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.32 (OR=0.319, p-value<.0001) times less likely to experience IPV, compared to a woman who views wife-beating as unacceptable.

A woman with medium exposure to the media is 1.29 (OR=1.2912, p-value=0.0043) times more likely to experience IPV, compared to a woman with low exposure to the media. A woman from the Western Cape province is 0.64 (OR=0.643, p-value=0.0032) times less likely to experience IPV compared to a woman from the Eastern Cape province. A woman from the Northern Cape region is 0.53 (OR=0.532, p-value<.0001) times less likely to experience IPV compared to a woman from the Eastern Cape region. A one-member increase in the number of household members decreases a woman's chances of experiencing IPV by 0.1652 units. A woman from a household where the head of the house is male is 1.25 (OR=1.253, p-value=0.0007) times more likely to experience IPV, compared to one from a household where the head of the house is female.

A woman who has never terminated a pregnancy is 0.66 (OR=0.663, p-value<.0001) times less likely to experience IPV, compared to a woman who has terminated one. A unit increase in a woman's body mass index decreases her chances of experiencing IPV by 0.0023 units. A married woman is 1.72 (OR=1.715, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 1.86 (OR=1.816, p-value<.0001) times more likely to experience IPV compared to a single woman.

A woman with a partner who wants more children compared to her is 1.38 (OR=1.384, p-value<.0001) times more likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman with a partner with fewer children compared to her is 1.29 (OR=1.286, p-value=0.0488) times more likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman who does not know the number of children her partner wants is

1.38 (OR=1.378, p-value<.0001) times more likely to experience IPV, compared to one whose partner wants the same number of children as her.

A woman who has an employed partner is 0.83 (OR=0.833, p-value=0.0198) times less likely to experience IPV, compared to one who does not know if her partner is employed or not. An employed woman is 1.27 (OR=1.266, p-value=0.0002) times more likely to experience IPV compared to an unemployed woman. A one-year increase in the age of a woman's partner decreases the chances of experiencing IPV by 0.0114 units. A woman who earns about the same as her partner is 1.20 (OR=1.196, p-value=0.0233) times more likely to experience IPV, compared to one who earns less than her partner. A woman whose partner does not bring in earnings is 1.41 (OR=1.414, p-value=0.0009) times more likely to experience IPV, compared to one who earns less than her partner.

Table 4.3: Solutions for fixed effects

Effect	**Estimate**	**Standard error**	**t value**	**Pr > |t|**	**Odds ratio**
Intercept	0,1092	0.3949	0.28	0.7823	1,1154
Partner drinks alcohol (ref=Yes)					
No	-0,8319	0.05682	-14.64	<.0001	0,4352
Don't know	0,7004	0.6409	1.09	0.2746	2,0146
Woman's father ever beat her mother (ref=Yes)					
No	-0,8394	0.07941	-10.57	<.0001	0,432
Don't know	-0,3185	0.1391	-2.29	0.0221	0,7272
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,5896	0.1077	5.47	<.0001	1,8033
I don't know	-1,1429	0.2789	-4.10	<.0001	0,3189
Access to the media (ref=Low exposure)					
Medium exposure	0,2556	0.08952	2.86	0.0043	1,2912
High exposure	0,1819	0.1086	1.68	0.0939	1,1995
Woman's current age	0,014	0.006355	2.20	0.0276	1,0141
Region (ref=Eastern Cape)					
Western Cape	-0,4414	0.1495	-2.95	0.0032	0,6431
Northern Cape	-0,6303	0.1459	-4.32	<.0001	0,5324
Free State	0,04441	0.1334	0.33	0.7392	1,0454
Kwazulu-Natal	0,02263	0.1204	0.19	0.8510	1,0229
North West	-0,01214	0.1292	-0.09	0.9252	0,9879
Gauteng	-0,01433	0.1331	-0.11	0.9142	0,9858
Mpumalanga	0,005698	0.1248	0.05	0.9636	1,0057
Limpopo	0,1596	0.1265	1.26	0.2073	1,173
Number of household members	-0,1652	0.01203	-13.72	<.0001	0,8477
Sex of household head (ref=Female)					
Male	0,2256	0.06656	3.39	0.0007	1,2531

Continued…

Wealth index (ref=Richest)					
Poorest	0,6659	0.1658	4.02	<.0001	1,9462
Poorer	0,7936	0.1556	5.10	<.0001	2,2113
Middle	0,5477	0.1486	3.69	0.0002	1,7293
Richer	0,4667	0.1462	3.19	0.0014	1,5947
Ever had a terminated pregnancy (ref=Yes)					
No	-0,4104	0.08808	-4.66	<.0001	0,6634
Contraceptive use (ref=Using)					
Not using	-1,6414	0.2660	-6.17	<.0001	0,1937
Body mass index	-0,00226	0.001013	-2.23	0.0257	0,9977
Current marital status (ref=Single)					
Married	0,5395	0.08087	6.67	<.0001	1,7151
Living with partner	0,6211	0.09070	6.85	<.0001	1,861
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	-0,188	0.1071	-1.76	0.0792	0,8286
Partner's desire for children (ref=Both want same)					
Partner wants more	0,3252	0.07558	4.30	<.0001	1,3843
Partner wants fewer	0,2517	0.1277	1.97	0.0488	1,2862
Don't know	0,3208	0.06865	4.67	<.0001	1,3782
Woman's occupation (ref=Unemployed)					
Employed	0,236	0.06339	3.72	0.0002	1,2662
Partner's age	-0,01139	0.004691	-2.43	0.0152	0,9887
Woman's earnings compared to partner (ref=Less compared to him)					
More compared to him	0,02596	0.07410	0.35	0.7261	1,0263
About the same	0,1786	0.07868	2.27	0.0233	1,1955
Partner does not bring in earnings	0,3463	0.1040	3.33	0.0009	1,4138
Don't know	-0,07424	0.1796	-0.41	0.6793	0,9284
Knows STI's (ref=Yes)					
No	-0,8308	0.2313	-3.59	0.0003	0,4357
Interaction effects					
Age of respondent by contraceptive use	0,02955	0.005948	4.97	<.0001	1,03
Wealth index by contraceptive use (ref=Richest by not using contraceptives)					
Poorest by not using contraceptives	0,5064	0.2083	2.43	0.0151	1,6593
Poorer by not using contraceptives	0,2409	0.2045	1.18	0.2389	1,2724
Middle by not using contraceptives	0,5577	0.2022	2.76	0.0058	1,7467
Richer by not using contraceptives	0,2711	0.2088	1.30	0.1941	1,3114

A woman who does not know about STIs is 0.44 (OR=0.436, p-value=0.0003) times less likely to experience IPV, compared to one who does not know about them.

Interaction Effects

Figure 4.1 shows that IPV increases with increasing age, whether a woman is using contraceptives or not. We observe from the same figure that IPV is higher among women using contraceptives compared to women not using any.

Figure 4.2 shows that, for women who are not using contraceptives, IPV decreases for a woman from the poorest to a poorer wealth index class; it then increases from a poorer to a middle wealth index class, and decreases from a middle to a richer wealth index class. We observe from the same figure that, for women using contraceptives, IPV increases for women from the poorest to a poorer wealth index class, and decreases for women from a poorer, middle, and richer wealth index class.

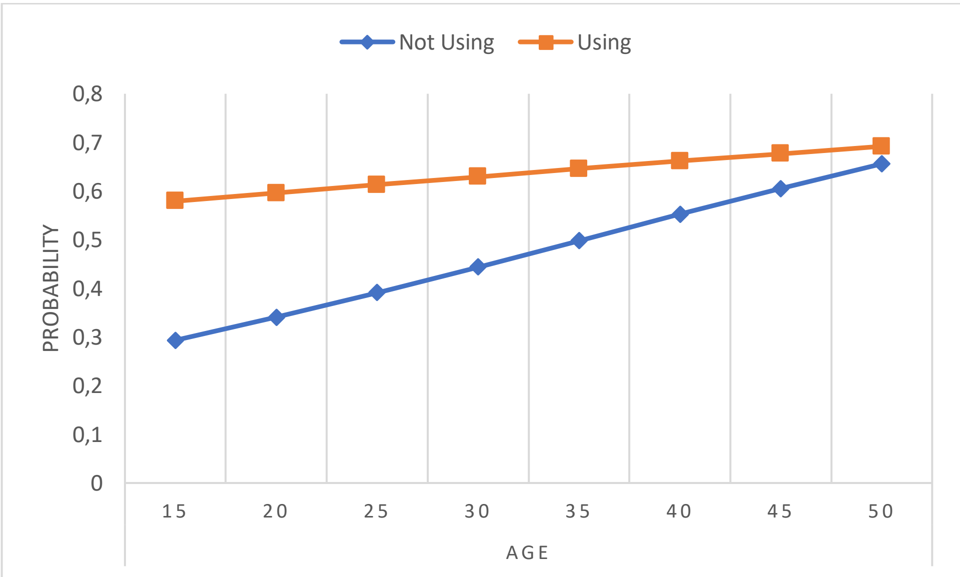

Figure 4.1: Predicted probability of experiencing IPV by woman's age and contraceptive use

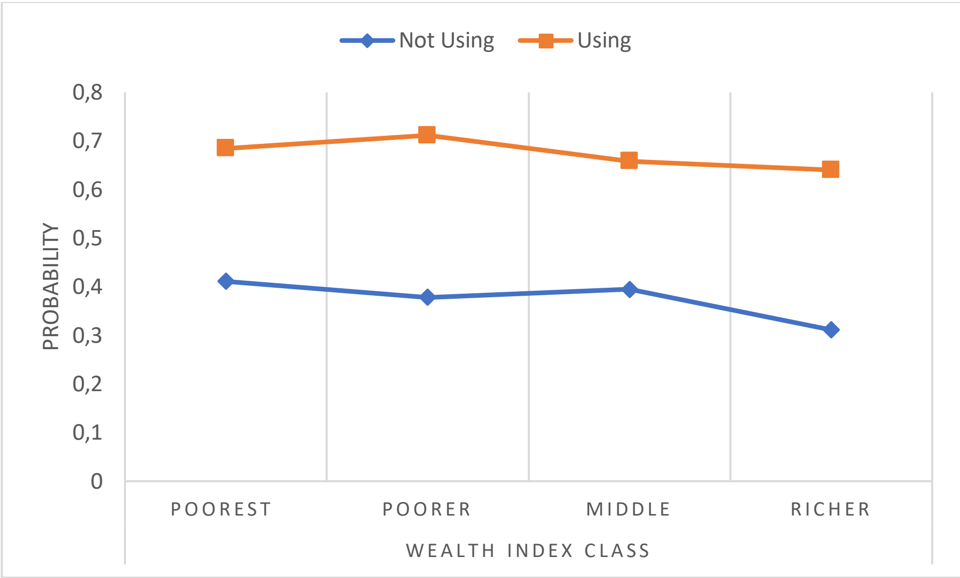

Figure 4.2: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Tanzania

In Table 4.4, the ratio of generalized chi-square statistic and its degrees of freedom is 1.00. This value is the measure of the residual variability in the marginal distribution of the data. Since the value is equal to 1, this indicates that the variability in the data has been adequately modeled, and then there is no residual over-dispersion.

Table 4.4: Fit statistics for conditional distribution

-2 log L (IPV | r. effects)	12838.36
Pearson chi-qquare	13244.59
Pearson chi-qquare / DF	1.00

The type 3 tests of fixed effects for the model fitted using the Laplace method in GLMMs, are shown in Table 4.5 The F-statistics used for the significant test for the fixed effects and corresponding p-value, show that some effects are essential. Some are not in the fitted model at a 5% level of significance: access to media, type of residence, literacy of the woman, body mass index, partner’s desire for children, partner’s education level, partner’s occupation, and woman’s earnings compared to her partner, which were all found to be insignificant.

Table 4.5: Type III tests of fixed effects

Effect	F-value	DF	P-value
Partner drinks alcohol	277.09	1	<.0001
Woman's father ever beat her mother	60.67	2	<.0001
Access to the media	0.57	2	0.5629
Wife-beating attitude	48.58	2	<.0001
Woman's current age	53.74	1	<.0001

Continued…

Region	4.64	29	<.0001
Type of residence	1.61	1	0.2038
Woman's education level	9.14	3	<.0001
Number of household members	656.61	1	<.0001
Sex of the household head	7.08	1	0.0078
Literacy	0.58	1	0.4468
Wealth index	2.90	4	0.0205
Ever had a terminated pregnancy	7.04	1	0.0080
Contraceptive method used	60.78	1	<.0001
Body Mass Index	1.16	1	0.2815
Current marital status	441.25	2	<.0001
Number of other wives	13.13	1	0.0003
Cohabitation duration	85.88	1	<.0001
Partner's desire for children	1.40	3	0.2406
Partner's education level	1.71	4	0.1447
Partner's occupation	0.59	1	0.4414
Woman's occupation	32.46	1	<.0001
Partner's age	9.48	1	0.0021
The person who usually decides how to spend a woman's earnings	12.09	3	<.0001
Woman's earnings compared to partner	1.64	4	0.1606
Interaction effects			
Woman's age by contraceptive use	29.96	1	<.0001
Wealth index by contraceptive use	1.29	4	0.2697

***DF=Degrees of Freedom**

Interpretation of the Results

Table 4.6 shows that a woman whose partner does not drink alcohol is 0.39 (OR=0.368, p-value<.0001) times less likely to experience IPV, compared to one whose partner does drinks it. A woman who has never witnessed her father beating her mother is 0.60 (OR=0.597, p-value<.0001) times less likely to experience IPV, compared to one who has witnessed her father beating hermother. A woman who views wife-beating as acceptable is 1.49 (OR=1.486, p-value<.0001) times more likely to experience IPV, compared to one who views it as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.66 (OR=0.655, p-value=0.0007) times less likely to experience IPV, compared to one who views it as unacceptable. A woman from Kilimanjaro province is 0.47 (OR=0.469, p-value=0.0002) times less likely to experience IPV compared to a woman from Arusha province. A woman from Tanga province is 0.58 (OR=0.581, p-value=0.0047) times less likely to experience IPV compared to a woman from Arusha province. A woman from Njobe province is 0.50 (OR=0.503, p-value=0.0006) times less likely to experience IPV compared to a woman from Arusha province. A woman from Kaskazin Pemba province is 0.41 (OR=0.408, p-value<.0001) times less likely to experience IPV compared to a woman from Arusha province. A woman from Kusini Pemba province is 0.61 (OR=0.609, p-value=0.0230) less likely to experience IPV compared to a woman from Arusha province.

A woman with secondary education is 0.79 (OR=0.793, p-value=0.0306) times less likely to experience IPV, compared to one who has none. A woman who has a higher education is 0.52 (OR=0.524, p-value=0.0108) times less likely to experience IPV, compared to one who has none. A one-member increase in the number of members in a household decreases the chance of a woman experiencing IPV by 0.1971 units. A woman from a household where the head of the house is male is 0.82 (OR=0.820, p-value=0.0015) times less likely to experience IPV, compared to one from a household where the head is female. A woman who has never terminated a pregnancy is 0.86 (OR=0.863, p-value=0.0107) times less likely to experience IPV, compared to a woman who has terminated one. A married woman is 6.41 (OR=6.41, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 6.04 (OR=6.042, p-value<.0001) times more likely to experience IPV compared to a single woman. A one-partner increase by the woman's partner increases the chances of experiencing IPV by 0.1318 units. A woman who stayed with her partner for 5-9 years is 0.49 (OR=0.494, p-value<.0001) times less likely to experience IPV, compared to one who stayed with him for 0-4 years.

A woman with a partner who wants more children than her is 1.18 (OR=1.182, p-value=0.0033) times more likely to experience IPV, compared one whose partner wants the same number of children as her. A woman with a partner who wants fewer children than her is 1.34 (OR=1.343, p-value=0.0030) times more likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman who does not know the number of children her partner wants is 1.30 (OR=1.299, p-value<.0001) times more likely to experience IPV, compared to one whose partner wants the same number of children as her.

Table 4.6: Solutions for fixed effects

Effect	**Estimate**	**Standard error**	**t value**	**Pr > |t|**	**Odds ratio**
Intercept	0,9269	0.2938	3.16	0.0017	2,5267
Partner drinks alcohol (ref=Yes)					
No	-0,9988	0.05605	-17.82	<.0001	0,3683
Woman's father ever beat her mother (ref=Yes)					
No	-0,5157	0.05019	-10.27	<.0001	0,5971
Don't know	-0,1216	0.08787	-1.38	0.1665	0,8855
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,3959	0.04812	8.23	<.0001	1,4857
I don't know	-0,4225	0.1248	-3.38	0.0007	0,6554

Access to the media (ref=Low exposure)					
Medium exposure	0,01008	0.05065	0.20	0.8422	1,0101
High exposure	-0,2314	0.1471	-1.57	0.1158	0,7934
Woman's current age	0,02265	0.006348	3.57	0.0004	1,0229
Region (ref=Arusha)					
Dodoma	-0,3165	0.2018	-1.57	0.1169	0,7287
Kilimanjaro	-0,7581	0.2045	-3.71	0.0002	0,4686
Tanga	-0,5436	0.1922	-2.83	0.0047	0,5807
Morogoro	-0,00712	0.2026	-0.04	0.9720	0,9929
Pwari	0,01124	0.2014	0.06	0.9555	1,0113
Dar Es Salaam	0,03426	0.1754	0.20	0.8451	1,0349
Lindi	0,1203	0.1988	0.61	0.5451	1,1278
Mtwara	0,2202	0.2030	1.08	0.2782	1,2463
Ruvuma	-0,2647	0.1981	-1.34	0.1816	0,7674
Iringa	-0,3842	0.2067	-1.86	0.0630	0,681
Mbeya	-0,2046	0.2026	-1.01	0.3127	0,815
Singida	-0,2462	0.1959	-1.26	0.2090	0,7818
Tabora	0,08292	0.1881	0.44	0.6593	1,0865
Rukwa	-0,1402	0.1957	-0.72	0.4735	0,8692
Kigoma	-0,2451	0.1896	-1.29	0.1961	0,7826
Shinyanga	0,1967	0.1895	1.04	0.2992	1,2174
Kagera	0,1109	0.1995	0.56	0.5781	1,1173
Mwanza	0,127	0.1943	0.65	0.5134	1,1354
Mara	0,3582	0.1886	1.90	0.0575	1,4308
Manyanga	0,2395	0.1961	1.22	0.2219	1,2706
Njobe	-0,6878	0.2016	-3.41	0.0006	0,5027
Katavi	-0,2537	0.1931	-1.31	0.1888	0,7759
Simiyu	-0,01514	0.1880	-0.08	0.9358	0,985
Geita	0,04888	0.1902	0.26	0.7971	1,0501
Kaskazini Unguja	0,3553	0.2085	1.70	0.0883	1,4266
Kusini Unguja	0,3385	0.2113	1.60	0.1092	1,4028
Mjini Magharibi	0,2766	0.1895	1.46	0.1443	1,3186
Kaskazini Pemba	-0,8975	0.2222	-4.04	<.0001	0,4076
Kusini Pemba	-0,4955	0.2179	-2.27	0.0230	0,6093
Woman's highest education level (ref=No education)					
Primary education	0,1054	0.08439	1.25	0.2117	1,1112
Secondary education	-0,2325	0.1075	-2.16	0.0306	0,7925
Higher education	-0,6469	0.2537	-2.55	0.0108	0,5237
Number of household members	-0,1971	0.007722	-25.52	<.0001	0,8211
Sex of household head (ref=Female)					
Male	-0,199	0.06272	-3.17	0.0015	0,8195

Continued…

Wealth index (ref=Richest)					
Poorest	0,3986	0.1582	2.52	0.0117	1,4897
Poorer	0,3269	0.1424	2.30	0.0217	1,3867
Middle	0,4995	0.1339	3.73	0.0002	1,6479
Richer	0,3796	0.1121	3.39	0.0007	1,4617
Ever had a terminated pregnancy (ref=Yes)					
No	-0,1474	0.05777	-2.55	0.0107	0,8629
Contraceptive use (ref=Using)					
Not using	-1,2285	0.1905	-6.45	<.0001	0,2927
Current marital status (ref=Single)					
Married	1,858	0.06387	29.09	<.0001	6,4109
Living with partner	1,7987	0.07676	23.43	<.0001	6,0418
Number of other partners	0,1318	0.02474	5.32	<.0001	1,1409
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	-0,7058	0.07600	-9.29	<.0001	0,4937
Woman's occupation (ref=Unemployed)					
Employed	0,3748	0.06167	6.08	<.0001	1,4547
Partner's age	-0,01171	0.003501	-3.34	0.0008	0,9884
The person who usually decides how to spend a woman's earnings (ref=Woman alone)					
Woman and partner	-0,2576	0.05141	-5.01	<.0001	0,773
Partner alone	-0,4346	0.08404	-5.17	<.0001	0,648
Someone else	-0,856	1.3511	-0.63	0.5264	0,425
Interaction effects					
Woman's age by contraceptive use	0,03002	0.005362	5.60	<.0001	1,0305
Wealth index (ref=Richest) by contraceptive use (ref=Not using)					
Poorest by not using contraceptives	-0,1354	0.1613	-0.84	0.4015	0,8734
Poorer by not using contraceptives	-0,1361	0.1487	-0.92	0.3601	0,8728
Middle by not using contraceptives	-0,3093	0.1434	-2.16	0.0310	0,734
Richer by not using contraceptives	-0,2649	0.1309	-2.02	0.0431	0,7673

An employed woman is 1.46 (OR=1.455, p-value<.0001) times more likely to experience IPV compared to an unemployed woman. A woman who decides with her partner how to spend her earnings is 0.77 (OR=0.773, p-value<.0001) times less likely to experience IPV, compared to one who decides how to spend her earnings alone. A woman whose partner alone decides how to spend her earnings is 0.65 (OR=0.648, p-value<.0001) times less likely to experience IPV, compared to one who decides how to spend her earnings alone. A one-year increase in the age of a woman's partner decreases the chances of experiencing IPV by 0.0117 units.

Interaction Effects

Figure 4.3 shows that women aged between 15 to about 40 years and who are using contraceptives, are at higher risk of experiencing IPV, compared to those who are not using them. From the same figure, women aged between about 41 to 44 years and who are not using contraceptives, are at a higher risk of experiencing IPV compared to those using them.

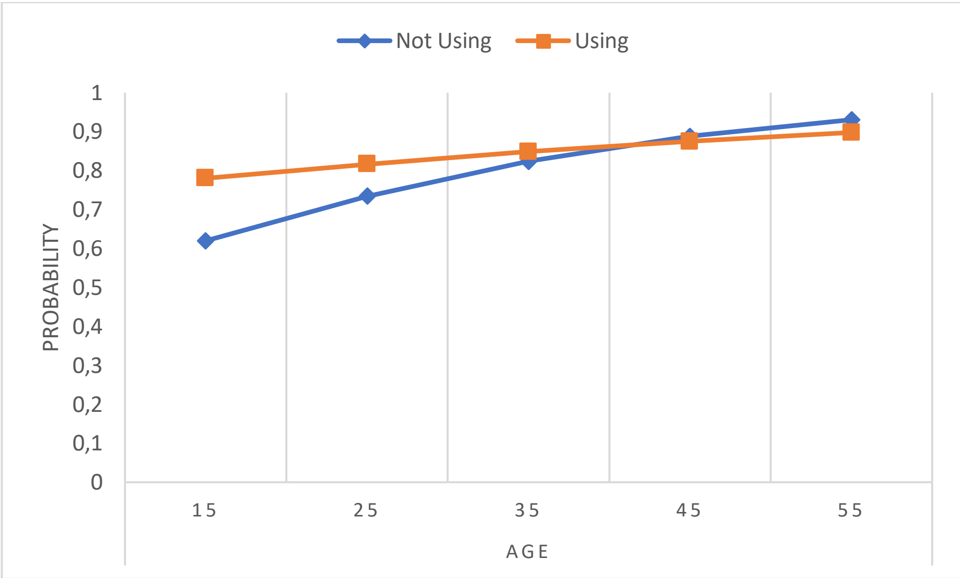

Figure 4.3: Predicted probability of experiencing IPV by woman's age and contraceptive use

Figure 4.4 shows that women using contraceptives from all the different wealth index classes are at a high risk of experiencing IPV, compared to those not using them from all these same classes. From the same figure, women who use contraceptives and come from a middle wealth index class are at the highest risk of experiencing IPV, followed by women from the poorest and richer classes, and finally those from a poorer wealth index class. Women who are not using contraceptives from the poorest wealth index class are at the highest risk of experiencing IPV, followed by the poorer, middle, and richer wealth index classes.

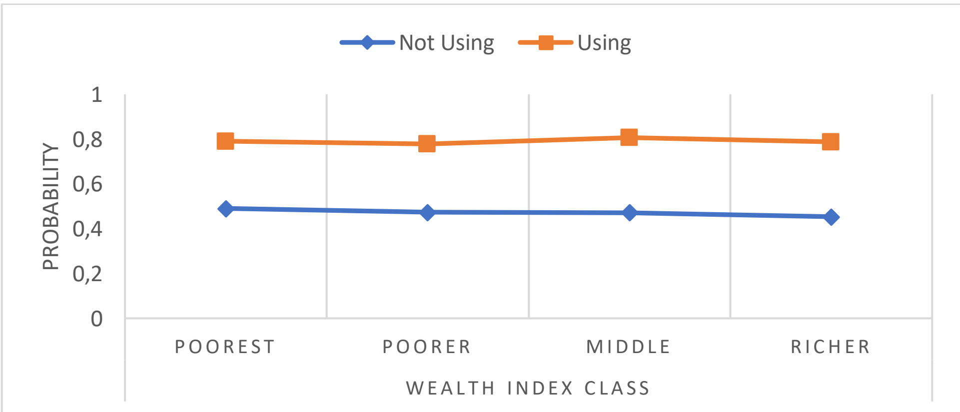

Figure 4.4: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Uganda

In Table 4.7, the ratio of generalized chi-square statistic and its degrees of freedom is 0.93. This value is the measure of the residual variability in the marginal distribution of the data. This value indicates that the variability in the data has been adequately modeled, but there is some residual over-dispersion.

Table 4.7: Fit statistics for conditional distribution

-2 log L (IPV | r. effects)	18959.58
Pearson chi-square	17209.88
Pearson chi-square / DF	0.93

The type 3 tests of fixed effects for the model fitted using the Quadrature method with two quadrature points in GLMMs, are shown in Table 4.8. The F-statistics, which are used for the significant test for the fixed effects, and the corresponding p-value, show that some effects are essential, and some are not in the fitted model, at 5% level of significance. The categories of type of residence, woman's level of education, literacy of the woman, partner's occupation, partner's age, and knowledge of STIs, were all found to be insignificant.

Table 4.8: Type III tests of fixed effects

Effect	F value	DF	P-value
Partner drinks alcohol	521.43	1	<.0001
Woman's father ever beat her mother	128.55	2	<.0001
Access to the media	17.31	2	<.0001
Wife-beating attitude	3.42	2	0.0328
Woman's current age	36.52	1	<.0001
Region	7.30	14	<.0001
Type of residence	2.20	1	0.1381
Woman's education level	2.39	3	0.0665
Number of household members	357.64	1	<.0001
Sex of the household head	16.42	1	<.0001
Literacy	1.83	1	0.1764
Wealth index	9.66	4	<.0001
Ever had a terminated pregnancy	9.96	1	0.0016
Contraceptive method used	52.34	1	<.0001
Body Mass Index	0,96	1	<.0001
Current marital status	326.23	2	<.0001
Number of other wives	27.04	1	<.0001
Cohabitation duration	41.45	1	<.0001
Partner's desire for children	5.54	3	0.0008
Partner's education level	8.16	4	<.0001

Continued…

Partner's occupation	2.62	1	0.1057
Woman's occupation	4.18	4	0.0022
Partner's age	1.11	1	0.2912
The person who usually decides how to spend a woman's earnings	15.88	3	<.0001
Woman's earnings compared to partner	15.01	4	<.0001
Knowledge of Sexually Transmitted Infections (STIs)	0.72	1	0.3957
Interaction effects			
Woman's age by contraceptive use	41.53	1	<.0001
Wealth index by contraceptive use	3.50	4	0.0073

Table 4.9 shows that a woman with a partner who does not drink alcohol is 0.43 (OR=0.427, p-value<.0001) times less likely to experience IPV, compared to one with a partner who does drinks it. A woman who has never witnessed her father beating her mother is 0.54 (OR=0.536, p-value<.0001) times less likely to experience IPV, compared to one who has witnessed him beating her mother. A woman who views wife-beating as acceptable is 1.21 (OR=1.211, p-value<.0001) times more likely to experience IPV, compared to one who views it as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.66 (OR=0.656, p-value=0.0142) times less likely to experience IPV, compared to one who views it as unacceptable.

A woman with high exposure to the media is 0.75 (OR=0.746, p-value=0.0241) times less likely to experience IPV, compared to one with low exposure to it. A woman from the Busoga region is 1.24 (OR=1.243, p-value=0.0354) times more likely to experience IPV compared to one from South Buganda. A woman from the Bukedi province is 1.59 (OR=1.589, p-value<.0001) times more likely to experience IPV compared to one from South Buganda. A woman from the Busigu province is 1.40 (OR=1.403, p-value=0.0014) times more likely to experience IPV compared to one from South Buganda. A woman from the Bunyoro province is 1.28 (OR=1.277, p-value=0.0223) times more likely to experience IPV compared to one from South Buganda. A woman from the Kigezi province is 1.50 (OR=1.501, p-value<.0001) times more likely to experience IPV compared to one from South Buganda. A woman with primary education is 1.14 (OR=1.141, p-value=0.0427) times more likely to experience IPV compared to a woman with no education.

A one-member increase in the number of household members reduces a woman's chances of experiencing IPV by 0.1365 units. A woman from a house where the household head is male is 1.13 (OR=1.129, p-value=0.0081) times more likely to experience IPV, compared to one from a house where the household head is female. A woman who has never had a terminated pregnancy is 0.88 (OR=0.878, p-value=0.0036) times less likely to experience IPV, compared to a woman who has had one. A married woman is 3.53 (OR=3.533, p-value<.0001) times more likely to experience IPV compared to a single woman. A woman living with her partner is 3.58 (OR=3.575, p-value<.0001) times more likely to experience IPV compared to a single woman. A unit increase in the number of other wives a woman's partner has, increases her chances of experiencing IPV by 0.0547 units. A woman who has stayed with her partner for 5-9 years is 0.66 (OR=0.664, p-value<.0001) times less likely to experience IPV, compared to one who has stayed with him for 0-4 years.

A woman who does not know the number of children her partner wants is 0.85 (OR=0.851, p-value<.0001) times less likely to experience IPV, compared to one who wants the same number of children as him. A woman who has a partner with a secondary level of education is 0.77 (OR=0.770, p-value=0.0014) times less likely to experience IPV, compared to one who has a partner with no education. A woman who has a partner with a higher level of education is 0.63 (OR=0.627, p-value<.0001) times less likely to experience IPV, compared to one who has a partner with no education. A woman who does not know her partner's level of education is 0.49 (OR=0.489, p-value<.0001) times less likely to experience IPV, compared to one who has a partner with no education. A woman who has an employed partner is 0.78 (OR=0.781, p-value=0.0128) times less likely to experience IPV, compared to one who has an unemployed partner.

Table 4.9: Solutions for fixed effects

Effect	Estimate	Standard error	t-value	P-value	Odds ratio
Intercept	-0,3744	0.2250	-1.66	0,032	0,6877
Partner drinks alcohol (ref=Yes)					
No	-0,8513	0.03819	-22.29	<.0001	0,4269
Woman's father ever beat her mother (ref=Yes)					
No	-0,6231	0.03874	-16.09	<.0001	0,5363
Don't know	-0,05887	0.09378	-0.63	0.5301	0,9428
Wife-beating attitude(ref=Unacceptable)					
Acceptable	0,1917	0.03813	5.03	<.0001	1,2113
I don't know	-0,4209	0.1717	-2.45	0.0142	0,6565
Access to the media (ref=Low exposure)					
Medium exposure	-0,0152	0.04053	-0.38	0.7076	0,9849
High exposure	-0,293	0.1299	-2.26	0.0241	0,746
Woman's current age	0,008664	0.005351	1.62	0.1054	1,0087
Region (ref=South Buganda)					
Kampala	0,2297	0.1223	1.88	0.0602	1,2582
North Buganda	0,1514	0.1031	1.47	0.1419	1,1635
Busoga	0,2171	0.1032	2.10	0.0354	1,2425
Bukedi	0,463	0.1013	4.57	<.0001	1,5888
Bugisu	0,3385	0.1059	3.20	0.0014	1,4028
Teso	0,04367	0.1116	0.39	0.6957	1,0446
Karamoja	0,08403	0.1078	0.78	0.4359	1,0877
Lango	-0,1659	0.1307	-1.27	0.2042	0,8471
Acholi	-0,06881	0.1069	-0.64	0.5196	0,9335
West Nile	-0,1166	0.1130	-1.03	0.3020	0,8899
Bunyoro	0,2444	0.1069	2.29	0.0223	1,2769
Tooro	-0,1918	0.1099	-1.74	0.0811	0,8255
Ankole	0,1363	0.1033	1.32	0.1872	1,146
Kigezi	0,4059	0.1011	4.02	<.0001	1,5007
Number of household members	-0,1365	0.007448	-18.33	<.0001	0,8724
Sex of household head (ref=Female)					
Male	0,1211	0.04575	2.65	0.0081	1,1287
Wealth index (ref=Richest)					
Poorest	0,4487	0.1254	3.58	0.0003	1,5663
Poorer	0,1386	0.1150	1.21	0.2281	1,1487
Middle	0,008578	0.1102	0.08	0.9380	1,0086
Richer	-0,00009	0.1008	-0.00	0.9993	0,9999
Ever had a terminated pregnancy (ref=Yes)					
No	-0,1299	0.04460	-2.91	0.0036	0,8782

Continued…

Contraceptive use (ref=Using)					
Not using	-1,1367	0.1599	-7.11	<.0001	0,3209
Body mass index	-0,00007	0.001756	0.98	<.0002	0,9999
Current marital status (ref=Single)					
Married	1,2622	0.05535	22.81	<.0001	3,5332
Living with partner	1,2739	0.05355	23.79	<.0001	3,5748
Number of other partners	0,05467	0.01057	5.17	<.0001	1,0562
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	-0,409	0.06368	-6.42	<.0001	0,6643
Husband's desire for children (ref=Both want same)					
Husband wants more	0,02303	0.04327	0.53	0.5945	1,0233
Husband wants fewer	-0,08677	0.06462	-1.34	0.1794	0,9169
Don't know	-0,1619	0.05068	-3.19	0.0014	0,8505
Partner's education level (ref=Secondary education)					
No education	-0,133	0.07520	-1.77	0.0770	0,8755
Primary education	-0,2617	0.08195	-3.19	0.0014	0,7697
Higher	-0,4672	0.09962	-4.69	<.0001	0,6268
Don't know	-0,7155	0.1381	-5.18	<.0001	0,4889
The person who usually decide to do with woman's earnings (ref=Woman alone)					
Woman and partner	-0,2171	0.04034	-5.38	<.0001	0,8048
Partner alone	-0,3917	0.06976	-5.61	<.0001	0,6759
Someone else	-0,1933	0.4914	-0.39	0.6941	0,8242
Woman's earnings compared to partner (ref=Less compared to him)					
More compared to him	0,2157	0.06588	3.27	0.0011	1,2407
About the same	-0,4149	0.05055	-8.21	<.0001	0,6604
Partner does not bring in earnings	-0,262	0.1330	-1.97	0.0488	0,7695
Don't know	-0,518	0.09916	-5.22	<.0001	0,5957
Interaction effects					
Woman's age by contraceptive use	0,03194	0.004535	7.04	<.0001	1,0325
Wealth index (ref=Richest) by contraceptive use (ref=Not using)					
Poorest by not using contraceptives	-0,2762	0.1236	-2.24	0.0254	0,7587
Poorer by not using contraceptives	0,1055	0.1181	0.89	0.3717	1,1113
Middle by not using contraceptives	0,1471	0.1188	1.24	0.2158	1,1585
Richer by not using contraceptives	0,147	0.1183	1.24	0.2142	1,1584

A woman who decides with her partner how to spend her earnings is 0.81 (OR=0.805, p-value<.0001) times less likely to experience IPV, compared to one who decides how to spend them alone. A woman whose partner alone decides how to spend her earnings is 0.68 (OR=0.676, p-value<.0001) times less likely to experience IPV, compared to one who decides how to spend them alone. A woman who earns more compared to her partner is 1.24 (OR=1.241, p-value=0.0011) times more likely to experience IPV, compared to one who earns less than him. A woman who earns about the same as her partner is 0.66 (OR=0.660, p-value<.0001) times less likely to experience IPV, compared to one who earns less than him. A woman who has a partner who does not bring in earnings is 0.77 (OR=0.770, p-value=0.0488) times less likely to experience IPV, compared to one who earns less than him. Lastly, a woman who does not know how much her partner earns is 0.60 (OR=0.596, p-value<.0001) times less likely to experience IPV, compared to one who earns less than him.

Interaction Effects

Figure 4.5 shows that women using contraceptives from all the different wealth index classes are at a higher risk of experiencing IPV, compared to women not using contraceptives from all the same classes. From the same figure, women who use contraceptives and come from the poorest wealth index class are at the highest risk of experiencing IPV, followed by women from the poorer, middle, and richer wealth index classes. Women who are not using contraceptives from a poorer wealth index class are at highest risk of experiencing IPV, followed by women from the poorest, middle, and richer wealth index classes.

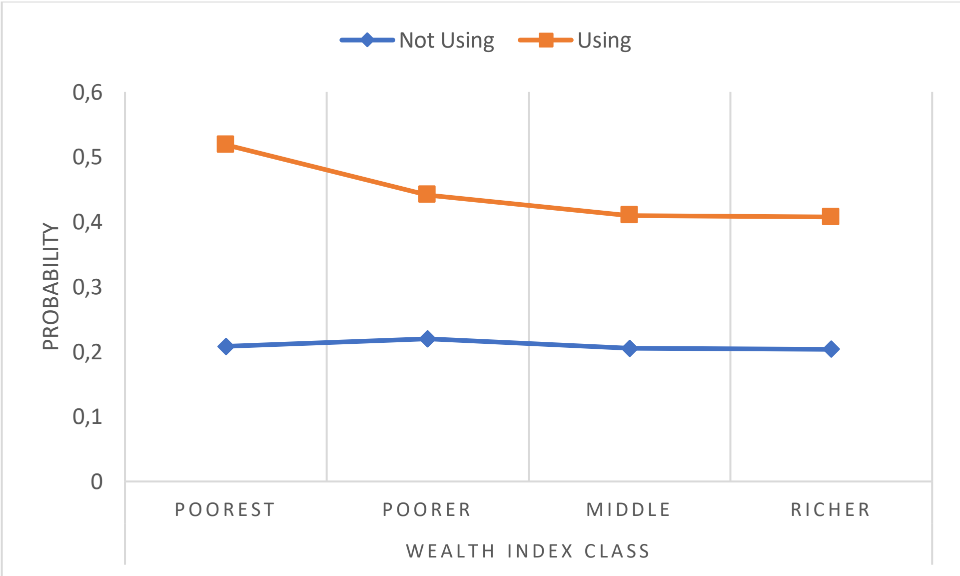

Figure 4.5: Predicted probability of experiencing IPV by woman's age and contraceptive use

Figure 4.6 shows that women who use contraceptives and are aged between 15 to 35 years have a higher risk of experiencing IPV, compared to those not using them. From the same figure, the risk of IPV for both women using and not using contraceptives is equal. Lastly, we observe from Figure 4.6 that between the ages of 35 to 50 years, the risk of experiencing IPV for women who are not using contraceptives is higher than for those using them.

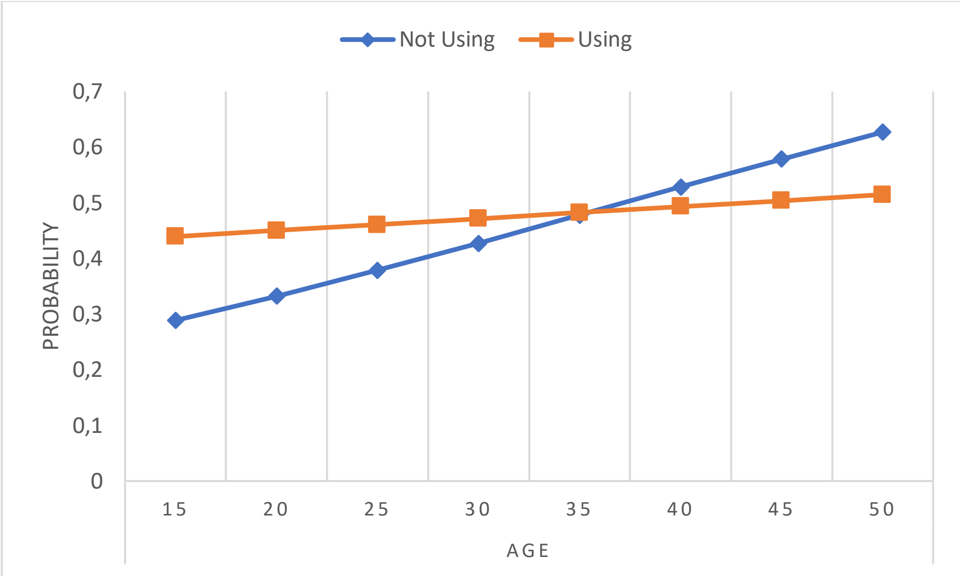

Figure 4.6: Predicted probability of experiencing IPV by wealth index class and contraceptive use

4.8 Summary

The section above was based on the 2016 South Africa Demographic and Health Survey, 2015-2016 Tanzania Demographic and Health Survey, and the 2016 Uganda Demographic and Health Survey data. In it, we used the generalized linear mixed model to identify the critical determinants of intimate partner violence amongst women of reproductive age in South Africa, Tanzania, and Uganda. The study's findings showed that in the three different countries, some determinants are common: partner's alcohol drinking status, whether the woman's father ever beat her mother, wife-beating attitude, woman's current age, the region where the woman resides, number of household members, the sex of the household head, wealth index, pregnancy termination status, contraceptive method used, current marital status, and cohabitation durations.

Some determinants are only significant in two of the countries and insignificant in the other. South Africa and Tanzania have the following determinants in common: woman's occupation and partner's age. South Africa and Uganda: access to the media, partner's desire for children, and the woman's earnings compared to her partner's. Tanzania and Uganda: number of other partners, and the person who usually decides how to spend the woman's earnings. South Africa, Tanzania, and Uganda all have significant determinants, specifically for one of them. In Uganda, South Africa, and Tanzania, the partner's education level, the woman's knowledge of STIs, and the woman's education level, respectively, were significant to intimate partner violence. In the following chapter, the generalized additive mixed model (GAMM) is used.

Chapter 5

Generalized Additive Mixed Models

5.1 Introduction

In previous chapters, we modeled the individual data using various statistical models such as generalized linear models, through logistic regression and survey logistic regression (Habyarimana, et al., 2014; Habyarimana, 2016), and generalized linear mixed models. All the models used are parametric (Habyarimana, 2016). The parametric models offer a powerful tool for modeling the relationship between the outcome and predictor variables (Habyarimana, 2016). However, some of the models discussed above may suffer from inflexibility in modeling complicated relationships between the outcome and predictor variables, in some applications (Habyarimana, 2016). The parametric mean assumption may not be desirable at times, as the suitable functional form of predictor variables may not be known in advance (Habyarimana, 2016). The response variables may depend on the covariates in a complicated manner (Lin & Zhang, 1999; Habyarimana, et al., 2018; Habyarimana, 2016). The linearity inherent in GLM and the assumption of normality, are easily relaxed by using a generalized additive mixed model. The flexibility of nonparametric regression for continuous predictor variables, coupled with GLM for predictor variables, offer ways to reveal structure within the data, that may miss GLM assumptions (Habyarimana, 2016). The flexibility of GAMM motivated the current research to use a semiparametric logistic mixed model, to assess the determinants of intimate partner violence. In other studies, (Hastie & Tibshirani, 1990; Green & Silverman, 1993; Habyarimana, 2016), many nonparametric regression models and smoothing methods for independent data were used (Habyarimana, 2016). The most commonly used are kernel smoothers, splines smoothers, locally weighted running-line smoothers, and running-mean smoothers (Habyarimana, 2016).

5.2 Generalized Additive Mixed Model

The generalized additive mixed model (GAMM) is perceived as an extension of generalized additive models, to include random effects or to be an extension of generalized linear mixed models (Breslow & Clayton, 1993; Habyarimana, 2016). These models allow the parametric fixed effects to be modeled non-parametrically using smooth additive functions in a similar spirit to Hastie & Tibshirani (1990), and Habyarimana (2016). Suppose that the observations of the m^{th} of k units consist of an outcome variable y_m and p covariates $x_m = (1, x_{m1}, \dots, x_{mp})^T$ associated with fixed effects, and $q \times 1$ of covariates z_m associated with random effects (Habyarimana, 2016). Therefore, Lin & Zhang (1999) formulated GAMM as follows (Habyarimana, 2016)

$$g(\mu_m) = \beta_0 + f_1(x_{m1}) + \cdots + f_p(x_{mp}) + z_t b \qquad (5.1)$$

where $g(.)$ is a monotonic differentiable link function, $\mu_m = E(y_m|b)$, $f_m(.)$ is a centered twice-differentiable smooth function, the random effect b is assumed to be distributed as $N\{0, K(\vartheta)\}$, and ϑ is a $c \times 1$ vector of variance elements (Habyarimana, 2016). A fundamental advantage of GAMM (5.1) over GAM is that the nonparametric additive functions are used to model covariate effects. Random effects are utilized to model the correlations between values (Lin & Zhang, 1999; Habyarimana, 2016). If $f_m(.)$ is a linear function, then (5.1) reduces to a generalized linear mixed model (GLMM) (Breslow & Clayton, 1993; Habyarimana, 2016). For a given variance elements ϑ, the log-quasi-likelihood function of $\beta_0, f_m, \vartheta, m = 1,2, \dots, k)$ is given (Lin & Zhang, 1999; Habyarimana, 2016) by:

$$exp\left[l\{\beta_0, f_1(.), \dots, f_k(.), \vartheta\}\right] \propto |K|^{\frac{-1}{2}} \int exp\left\{\frac{-1}{2\emptyset} \sum_{m=1}^{k} d_m(y_m; \mu_m) - \frac{1}{2} b' K^{-1} b\right\} db \qquad (5.2)$$

where $y_m = (y_1, y_2, \dots, y_k)$, and $d_m(y_m; \mu_m) \propto -2 \int_{y_m}^{\mu_m} m_m(y_m - u)/v(u)du$ defines the conditional deviance function of $\{\beta_0, f_m(.), \vartheta\}$, given **b**. Statistical inference in a generalized additive mixed model includes inference on the nonparametric functions $f_m(.)$, which requires the estimation of smoothing parameters, as well as inference on the variance elements ϑ (Habyarimana, 2016).

5.2.1 Natural Cubic Smoothing Spline Estimation

Following the derivation of Lin & Zhang (1999), with a given λ and ϑ, the natural cubic smoothing spline estimators of the $f_m(.)$ maximize the penalized log-quasi-likelihood as follows:

$$\begin{aligned} & l\{\beta_0, f_1(.), \dots, f_k(.), \vartheta\} - \frac{1}{2} \sum_{t=1}^{k} \lambda_t \int_{s_m}^{t_t} f_t^{''}(x^2) dx \\ & = l\{\beta_0, f_1(.), \dots, f_k(.), \vartheta\} - \frac{1}{2} \sum_{t=1}^{k} \lambda_t f_t^T H_t f_t \end{aligned} \qquad (5.3)$$

where (s_t, t_t) defines the range of the t^{th} covariates, and λ_t are smoothing parameters that regulate tradeoff between the goodness-of-fit and smoothness of the estimated functions (Habyarimana, 2016). In addition, $f_m(.)$ is an $r_t \times 1$ unknown vector of the values of $f_m(.)$, calculated at the r_t ordered distinct values of the $x_{tm}(t = 1,2, \dots, m)$, and H_t is the corresponding non-negative definite smoothing matrix (Green & Silverman, 1993; Habyarimana, 2016). GAMM, given in equation (5.1), can be formulated in matrix form as:

$$g(\mu_t) = 1\beta_0 + M_1 f_1 + M_2 f_2 + \cdots + M_k f_k + Z_b \qquad (5.4)$$

where $g(\mu_t) = \{g(\mu_1), g(\mu_2), \dots, g(\mu_m)\}$, 1 is an $m \times 1$ vector of 1's, M_t is a $k \times r_t$ incident matrix defined similarly to the one given in previous studies (Greenland, et al., 1994; Habyarimana, 2016), such that the t^{th} component of $M_m f_m$ is $f_m(x_{tm})$ and $Z_t = (z_1, z_2, \dots, z_m)^T$ (Habyarimana, 2016). The numerical integration is needed to estimate equation (5.2), except for the Gaussian outcome (Habyarimana, 2016). The natural cubic smoothing spline estimators of $f_t(.)$, evaluated by explicit maximization of equation (5.4), are sometimes challenging (Habyarimana, 2016). To solve this problem, Lin & Zhang (1999) proposed the double penalized quasi-likelihood approach as an alternative approximation approach, discussed in subsection 5.2.2 Double Penalized Quasi-likelihood below (Habyarimana, 2016).

5.2.2 Double Penalized Quasi-likelihood

Since f_t is a centered parameter vector, it can be parametrized in terms of β_t and $a_t((r_t - 2) \times 1)$ in a one-to-one transformation as (Habyarimana, 2016):

$$f_t = X_t \beta_t + \beta_t a_t \tag{5.5}$$

where X_t is an $r_t \times 1$ vector containing the $r_h$centered ordered distinct values of the $x_{tm}(t = 1,2, \dots, m)$, $\beta_t = L_t(L_t^T L_t)^{-1}$, and L_t is an $r_t \times (r_t - 2)$ full rank matrix satisfying $H_t = L_t L_t^T$ and $L_t^T X_t = 0$ using the identity $f_t^T H_t f_t = a_t^T a_t$ (Habyarimana, 2016), the double penalized quasi-likelihood concerning (β_0, f_t) and b is given by:

$$-\frac{1}{2\varphi}\sum_{t=1}^{m} d_t(y; \mu_t) - \frac{1}{2} b^T K^{-1} b - \frac{1}{2} a^T \Gamma^{-1} a \tag{5.6}$$

where $a = (a_1^T, a_2^T, \dots, a_k^T)^T$ and $\Gamma = daig(\tau_1 I, \tau_2 I, \dots, \tau_k I)$, with $\tau_t = \frac{1}{\lambda_t}$. A small value of $\tau = (\tau_1, \tau_2, \dots, \tau_k)^T$ corresponds to over-smoothing. Plugging equation (5.5) into (5.4), expression (5.4) suggests that given ϑ and τ, the DPQL estimators $\widehat{f_t}$ can be obtained by fitting the following GLMM, using (Breslow & Clayton, 1993; Habyarimana, 2016) the penalized quasi-likelihood approach:

$$g(\mu) = X\beta + Ba + zb \tag{5.7}$$

where $X = (1, M_1 X_1, M_2 X_2, \dots, M_k X_k), B = (M_1 B_1, M_2 B_2, \dots, M_k B_k), \beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$ is a $(k+1) \times 1$ vector of regression coefficients, and **a** and **b** are independent random effects with distributions $a \sim N(0, \Gamma)$ and $b \sim N(0, K)$ (Habyarimana, 2016). Therefore DPQL estimator $\widehat{f_m}$ is calculated as $\widehat{f_t} = X_t \widehat{\beta_t} + \beta_t \widehat{a_t}$, which is the linear combination of the penalized quasi-likelihood estimators of the fixed $\widehat{\beta_t}$ (Breslow & Clayton, 1993; Habyarimana, 2016), and the random effects $\widehat{a_t}$ in the working GLMM (5.7) (Habyarimana, 2016). The maximization of the expression (5.6) concerning (β, a, b) can proceed by using the Fisher scoring algorithm to solve (Habyarimana, 2016):

$$\begin{pmatrix} X^TWX & X^TWB & X^TWZ \\ B^TWX & B^TWB+\Gamma^{-1} & B^TWZ \\ Z^TWX & Z^TWB & Z^TWZ+K^{-1} \end{pmatrix}\begin{pmatrix} \beta \\ a \\ b \end{pmatrix} = \begin{pmatrix} X^TWY \\ B^TWY \\ Z^TWY \end{pmatrix}, \quad (5.8)$$

where the working vector Y is defined as $Y = \beta_0 1 + \sum_{m=1}^{p} M_t f_t + Zb + \Delta(Y-\mu)$ and $\Delta = diag[g'(\mu_t)], W = diag[\{\vartheta v(\mu_t) g'(\mu_t)^2\}^{-1}]$. An examination of the equation (5.8) shows that it corresponds to the standard equation of the best linear unbiased predictors of β and (a,b), under linear mixed model (Habyarimana, 2016):

$$Y = X\beta_0 + Ba + Zb + \epsilon \quad (5.9)$$

such that a and b are independent random effects, where $a \sim N(0,\Gamma)$ and $b \sim N(0,K)$ and $\epsilon \sim N(0,W^{-1})$ (Habyarimana, 2016). This suggests that the DPQL estimators $\widehat{f_J}$ and the random effects estimators $\hat{b}$ can be easily obtained using BLUPs, by iteratively fitting model (5.9) to the working vector Y (Lin & Zhang, 1999; Habyarimana, 2016). To compute the covariance matrix of $\widehat{f_J}$, it is more convenient to calculate β and a using:

$$\begin{pmatrix} X^TR^{-1}X & X^TR^{-1}B \\ B^TR^{-1}X & B^TR^{-1}B+\Gamma^{-1} \end{pmatrix}\begin{pmatrix} \beta \\ a \end{pmatrix} = \begin{pmatrix} X^TR^{-1}Y \\ B^TR^{-1}Y \end{pmatrix}, \quad (5.10)$$

where $R = W^{-1} + ZKZ^T$ (Habyarimana, 2016). Denoting H as the coefficient matrix on the left-hand side of equation (5.10) and $H_0 = (X,B)^T R^{-1}(X,B)$, the approximate covariance matrix of $\hat{\beta}$ and $\hat{a}$ is $cov(\hat{\beta},\hat{a}) = H^{-1}H_0H^{-1}$ (Habyarimana, 2016). It follows that the approximate covariance matrix of $\widehat{f_J}$ is $(X_m, B_m)cov(\hat{\beta},\hat{a})(X_m,B_m)^T$, where $cov(\hat{\beta},\hat{a})$ can be easily found from corresponding blocks of $H^{-1}H_0H^{-1}$ (Habyarimana, 2016). It is assumed that the $\widehat{f_J}(.)$ are smooth functions in calculating the covariance of the $\widehat{f_J}$ (Habyarimana, 2016).

5.3 Estimating Parameters and Variance Components

In the previous sections, we assumed that the smoothing parameters λ and the variance component ϑ are known, when estimating nonparametric function f_m (Habyarimana, 2016). However, the parameters need to be estimated from the data. Under the classical nonparametric regression model (Habyarimana, 2016):

$$y = f(X) + \epsilon \quad (5.11)$$

where ϵ are independent random errors distributed as $N(0,\sigma^2)$; Kohn, et al. (1991) proposed a way to estimate the smoothing parameter λ by maximizing a marginal likelihood (Habyarimana, 2016). The marginal likelihood of $\tau = \frac{1}{\lambda}$ is constructed by assuming that $f(X)$ has a prior, specified in the form of equation (5.5) with $a \sim N(0,\tau I)$, and a flat prior for β; therefore integrating out a and β as follows (Habyarimana, 2016):

$$exp\{\iota_M(y;\tau,\sigma^2)\} \propto \tau^{\frac{1}{2}} \int exp\left\{\iota(y;\beta,a,\sigma^2) - \frac{1}{2\tau}a^T a\right\} dad\beta \tag{5.12}$$

where $\iota(y;\beta,a,\tau^2)$ is the log-likelihood of f under model (5.*11)*. Robinson, (1991) pointed out that the marginal likelihood (5.12) of τ is indeed the restricted maximum likelihood (REML) under the linear mixed model (Habyarimana, 2016):

$$y = 1\beta_0 + X\beta_1 + Ba + \epsilon \tag{5.13}$$

where $a \sim N(0,\tau I)$, $\epsilon \sim N(0,\sigma^2 I)$, and B being defined as earlier (Habyarimana, 2016). τ is regarded as a covariance component, hence the marginal estimator of τ is a REML estimator. Kohn et al. (1991) found that the maximum marginal likelihood estimator of τ can at times perform better, compared to the generalized cross-validation estimator, for estimating a nonparametric function (Habyarimana, 2016). Zhang et al. (1998) extended the results to estimate the smoothing parameter λ and variance component ϑ jointly using REML (Habyarimana, 2016). In case of longitudinal data with a normally distributed outcome and a nonparametric mean function, their model is formulated as follows (Habyarimana, 2016):

$$y = f(X) + Zb + \epsilon \tag{5.14}$$

where $f(X)$ denotes the values of nonparametric function $f(.)$, evaluated at the design points of $X_{(m\times 1)}, b \sim N\big(0, K(\vartheta)\big)$ and $\epsilon \sim \mathrm{n}(0, \mathrm{V}(\vartheta))$. When $f(.)$ is estimated using a cubic smoothing spline (5.5), Zhang et al. (1998) rewrote the model (5.14) as a linear mixed model:

$$y = 1\beta_0 + X\beta_1 + Ba + Zb + \epsilon \tag{5.15}$$

where $a \sim N(0,\tau I)$, and the distribution of **b** and ϵ are the same as those in the model (5.14) (Habyarimana, 2016). Therefore, τ is proposed as an extra variance component in addition to ϑ in model (5.15), to estimate ϑ and τ jointly by using REML (Habyarimana, 2016). In this case, REML corresponds to the marginal likelihood of (τ,ϑ) constructed by assuming that f takes the form of (5.5) with $a \sim N(0,\tau I)$, and a flat prior for β; therefore integrating out a and β as follows (Habyarimana, 2016):

$$\exp\{\iota_M(y;\tau,\vartheta)\} \propto K^{\frac{-1}{2}}\tau^{\frac{-1}{2}} \int exp\left\{\iota(y;\beta,a,b) - \frac{1}{2}b^T K^{-1} b - \frac{1}{2\tau}a^T a\right\} dbdad\beta \tag{5.16}$$

where $\iota(y;\beta,a,b) = \iota(y;f,b)$ is the conditional likelihood (standard) of f, given the random effects **b** under the model (5.14) (Habyarimana, 2016). Note that the marginal log-likelihood $\iota_M(y;\tau,\vartheta)$ in (5.16) has closed form. Zhang et al. (1998) proposed an extension to the marginal likelihood approach to GAMM (5.4), and then estimating τ and ϑ jointly by maximizing a marginal quasi-likelihood (Habyarimana, 2016). Specifically, the GLMM representation of GAMM in (5.7) suggests that τ may be treated as extra variance components in addition to ϑ. In a similar way to REML (5.16), the marginal quasi-likelihood of (τ, ϑ) can be constructed under the GAMM (5.4), by assuming that f_m takes the form (5.5) with $a_m \sim N(0, \tau_m I)(m = 1,2, \dots, p)$, and integrating a_m and β out as follows (Habyarimana, 2016):

$$\exp\{\iota_M(y; \tau, \vartheta)\} \propto |\Lambda|^{\frac{-1}{2}} \int exp\left\{\iota(y; \beta, a, \vartheta) - \frac{1}{2} a^T \Gamma^{-1} a\right\} da d\beta \tag{5.17}$$

$$\propto |K|^{\frac{-1}{2}} |\Gamma|^{\frac{-1}{2}} \int \left\{\sum_{t=1}^{n} -\frac{1}{2\phi} d_t(y_t; \mu_t) - \frac{1}{2} b^T K^{-1} b - \frac{1}{2} a^T \Gamma^{-1} a\right\}$$

where $\iota(y; \beta, a, \vartheta) = \iota(y; \beta_0, f_1, f_2, \dots, f_k, \vartheta)$ was defined in (5.2) (Habyarimana, 2016). Based on the Gaussian nonparametric mixed model (5.14), the marginal quasi-likelihood reduces Gaussian REML (5.16) (Habyarimana, 2016). An evaluation of the marginal quasi-likelihood reducing (5.16) for non-Gaussian outcomes, is obtained after intractable numerical integration (Habyarimana, 2016). Laplace's approximation method is considered as an alternative method used to circumvent the problem (Habyarimana, 2016). Specifically, in taking the quadratic expansion exponent of the integrand of the deviance statistic $d_t(y; \mu_t)$ by the Pearson χ^2-statistic (Breslow & Clayton, 1993; Habyarimana, 2016), the approximate marginal log-quasi-likelihood is then given by:

$$\iota_M(y; \tau, \vartheta) \approx -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^T V^{-1} X| - \frac{1}{2}\left(Y - X\hat{\beta}^T V^{-1}\right)\left(Y - X\hat{\beta}\right), \tag{5.18}$$

where $V = B\Gamma B^T + ZKZ^T + W^{-1}$. Equation (5.18) corresponds to the REML log-likelihood of working vector y under the linear mixed model (5.9), with both **a** and **b** as random effects, and τ and ϑ as variance components (Habyarimana, 2016). Then τ and ϑ can be estimated by iteratively fitting model (5.9) using REML (Habyarimana, 2016).

5.4 Application to the Determinants of Intimate Partner Violence in South Africa, Tanzania, and Uganda

In Chapter 4, we used GLMM, however, the modeling was based on parametric models. The aim of this study was to model the effects of a woman's age, use of contraceptives, the interaction between age and use of contraceptives, non-parametrically, and the interaction between wealth index and use of contraceptives while other covariates remain parametric utilizing generalized additive mixed models.

5.4.1 Model Fitting and Interpretation of the Results

The various procedures for estimation discussed for fitting GAMM, can be used when fitting the semiparametric logistic mixed model (Habyarimana, 2016). The library mgcv from the R package was used to fit the data. R package has many options for controlling the model smoothness, using splines as cubic smoothing splines, locally weighted running line smoothers, and kernel smoothers (Habyarimana, 2016). Many authors have addressed the above-mentioned in detail such as: Green & Silverman (1993), Hardle (1999), Hastie & Tibshirani (1990), and Habyarimana (2016). The shrinkage smoothers have several advantages, for instance, helping to circumvent the knot placement (Habyarimana, 2016). Additionally, this method is constructed to smooth any number of covariates (Habyarimana, 2016). The study's main effect is considered for possible two-way interaction effects, where the AIC of each model, the inference of smooth function, and the p-value of the individual smooth term are considered (Habyarimana, 2016). Finally, the model with smaller AIC and high statistical significance was selected as follows (Habyarimana, 2016):

$$g(\mu_m) = \beta_0 + (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p) + f_1(Age\ of\ the\ woman_m) + b_{0m} \quad (5.19)$$

where $g(\mu_m)$ is the logit link function, $\beta's$ are parametric regression coefficients, $f_m's$ are centered smooth functions, and b_{0m} is the random effect distributed as $N(0, K(\vartheta))$ (Habyarimana, 2016). The standard, widely used methods for estimating additive models include cubic smoothing splines, locally-weighted running line smoothers, and kernel smoothers (Hardle, 1999; Hastie & Tibshirani, 1990; Habyarimana, 2016). The results from the equation above are presented under each country section and are interpreted.

South Africa

Table 5.1 shows that a woman whose partner drinks alcohol is 2.49 (OR=2.4905, p-value<2e-16***) times more likely to experience IPV, compared to one whose partner does not drink it. A woman who does not know if her partner drinks alcohol or not is 5.90 (OR=5.9000, p-value=0.004250**) times more likely to experience IPV, compared to one whose partner does not drink it. A woman who has witnessed her father beating her mother is 2.19 (OR=2.1884, p-value<2e-16***) times more likely to experience IPV, compared to one who has never witnessed her father beating her mother. A woman who does not know if her father beats her mother is 1.97 (OR=1.9691, p-value=1.52e-07***) times more likely to experience IPV, compared to one whose father has never beaten her mother. A woman who sees wife-beating as an acceptable act is 1.90 (OR=1.9049, p-value=5.64e-09***) times more likely to experience IPV, compared to a woman who sees it as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.34 (OR=0.3400, p-value=0.000142***) times less likely to experience IPV, compared to one who sees it as unacceptable. A woman from a household with more than five members is 0.45 (OR=0.4532, p-value<2e-16***) times less likely to experience IPV, compared to a woman from a household with five or fewer members. A woman with medium exposure to the media is 1.37 (OR=1.3745, p-value=0.000327***) times more likely to experience IPV, compared to one with low exposure to the media.

A woman from the Eastern Cape province is 1.51 (OR=1.5099, p-value=0.002482**) times more likely to experience IPV compared to one from the Western Cape province. A woman from the Free State province is 1.52 (OR=1.5218, p-value=0.002389**) times more likely to experience IPV compared to one from the Western Cape province. A woman from Kwazulu-Natal province is 1.58 (OR=1.5806, p-value=0.000628***) times more likely to experience IPV compared to one from the Western Cape province. A woman from North West province is 1.50 (OR=1.5024, p-value=0.003626**) times more likely to experience IPV compared to one from the Western Cape province. A woman from the Gauteng province is 1.58 (OR=1.5841, p-value=0.000774***) times more likely to experience IPV compared to one from the Western Cape province. A woman from Mpumalanga province is 1.48 (OR=1.4842, p-value=0.004168**) times more likely to experience IPV compared to one from the Western Cape province. A woman from Limpopo province is positively associated with IPV. A woman from Limpopo province is 1.74 (OR=1.7437, p-value=0.000101***) times more likely to experience IPV compared to one from the Western Cape province.

A woman from a household where the head of the house is female is 0.80 (OR=0.8030, p-value=0.000934***) times more likely to experience IPV, compared to one from a household where the head of the house is male. A woman who has terminated a pregnancy is 1.37 (OR=1.3695, p-value=0.000277***) times more likely to experience IPV, compared to one who has never terminated a pregnancy. A woman who is obese (BMI≥30) is 0.67 (OR=0.6743, p-value=0.018388*) times less likely to experience IPV compared to an underweight woman (BMI<18). A married woman is 1.47 (OR=1.4650, p-value=2.44e-06***) times more likely to experience IPV compared to a single woman. A woman living with her partner is 1.57 (OR=1.5717, p-value=7.58e-07***) times more likely to experience IPV compared to a single woman.

A woman with a partner who wants more children than her is 1.45 (OR=1.4473, p-value=1.82e-06***) times more likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman with a partner who wants fewer children than her is 1.51 (OR=1.5140, p-value=0.001279***) times more likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman who does not know the number of children her partner wants is 1.40 (OR=1.3964, p-value=5.77e-07***) times more likely to experience IPV, compared to one whose partner wants the same number of children as her.

A woman with an employed partner is 2.17 (OR=2.1655, p-value=0.000799***) times more likely to experience IPV, compared to one whose partner is unemployed. An employed woman is 1.21 (OR=1.2122, p-value=0.001994**) times more likely to experience IPV compared to an unemployed woman. A woman with a partner aged between 25 to 34 years is 0.77 (OR=0.7650, p-value=0.036368*) times less likely to experience IPV, compared to one whose partner is less than 25 years old. A woman with a partner who is aged 35 years or older, is 0.67 (OR=0.6553, p-value=0.003263**) times less likely to experience IPV, compared to one whose partner is less than 25 years old. A woman who decides with her partner how to spend her earnings is 0.87 (OR=0.8566, p-value=0.018921*) times less likely to experience IPV, compared to one who solely decides how to spend her earnings. A woman who earns about the same as her partner is 1.30 (OR=1.2988, p-value=0.007599**) times more likely to experience IPV, compared to one who earns more than him. A woman whose partner does not bring in earnings is 1.34 (OR=1.3494, p-value=0.013337*) times more likely to experience IPV, compared to one who earns more than him. A woman who does not know how much her partner earns is 1.43 (OR=1.4279, p-value=0.024263*) times more likely to experience IPV, compared to one who earns more than him.

Table 5.1: The parameter estimates of the IPV for the fixed part of GAMM

Effect	Estimate	Standard error	t-value	P-value	Odds ratio
Intercept	-3,315692	0,345442	-9,598	<2e-16***	0,0363
Partner drinks alcohol (ref=No)					
Yes	0,912471	0,056567	16,131	<2e-16***	2,4905
Don't know	1,774949	0,620662	2,86	0,004250**	5,9000
Woman's father ever beat her mother (ref=No)					
Yes	0,783176	0,079147	9,895	<2e-16***	2,1884
Don't know	0,677596	0,128967	5,254	1,52e-07***	1,9691
Number of household members (ref=Less than 5)					
More than or equal to 5	-0,791347	0,064592	-12,251	<2e-16***	0,4532
Wife beating attitude (ref=Unacceptable)					
Acceptable	0,644407	0,110471	5,833	5,64e-09***	1,9049
I don't know	-1,078901	0,28348	-3,806	0,000142***	0,3400
Access to the media (ref=Low exposure)					
Medium exposure	0,318069	0,088488	3,594	0,000327***	1,3745
High exposure	0,158442	0,10711	1,479	0,139113	1,1717
Region (ref=Western Cape)					
Eastern Cape	0,412044	0,136149	3,026	0,002482**	1,5099
Northern Cape	-0,241707	0,150897	-1,602	0,109236	0,7853
Free State	0,419888	0,138212	3,038	0,002389**	1,5218
Kwazulu-Natal	0,457814	0,133845	3,42	0,000628***	1,5806
North West	0,407086	0,139903	2,91	0,003626**	1,5024
Gauteng	0,459995	0,13677	3,363	0,000774***	1,5841
Mpumalanga	0,39485	0,137772	2,866	0,004168**	1,4842
Limpopo	0,556036	0,142938	3,89	0,000101***	1,7437
Type of residence (ref=Rural)					
Urban	-0,033638	0,072605	-0,463	0,643157	0,9669
Woman's highest education level (ref=No education)					
Primary education	-0,179293	0,220624	-0,813	0,416434	0,8359
Secondary education	-0,018054	0,224895	-0,08	0,936017	0,9821
Higher education	-0,002704	0,245996	-0,011	0,99123	0,9973
Sex of household head (ref=Male)					
Female	-0,219375	0,06626	-3,311	0,000934***	0,8030
Literacy (ref=Can read)					
Cannot read	0,010613	0,166142	0,064	0,949068	1,0107

Continued…

Wealth index (ref=Poorest)					
Poorer	-0,121217	0,124796	-0,971	0,331416	0,8858
Middle	-0,01027	0,127346	-0,081	0,935727	0,9898
Richer	-0,394082	0,140201	-2,811	0,004953**	0,6743
Richest	-1,126115	0,174102	-6,468	1,05e-10***	0,3243
Ever had a terminated pregnancy (ref=No)					
Yes	0,314463	0,086445	3,638	0,000277***	1,3695
Contraceptive use (ref=Not using)					
Using	0,018269	0,122071	0,15	0,881038	1,0184
Body Mass Index (ref=Underweight (BMI<18))					
Healthy (18≤BMI<25)	-0,1735562	0,169657	-1,023	0,306328	0,8407
Overweight (25≤BMI<30)	-0,247527	0,170754	-1,45	0,147205	0,7807
Obese (BMI≥30)	-0,394102	0,167122	-2,358	0,018388*	0,6743
Current marital status (ref=Single)					
Married	0,381886	0,080966	4,717	2,44e-06***	1,4650
Living with partner	0,452133	0,091347	4,95	7,58e-07***	1,5717
Number of other wives (ref=0)					
One or more	-0,092647	0,092647	-0,545	0,586075	0,9115
Don't know	-0,017783	0,128194	-0,139	0,889675	0,9824
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	0,091619	0,114166	0,803	0,422284	1,0959
Partner's desire for children (ref=Both want same)					
Husband wants more	0,369722	0,077418	4,776	1,82e-06***	1,4473
Husband wants fewer	0,414754	0,128733	3,222	0,001279**	1,5140
Don't know	0,333889	0,066741	5,003	5,77e-07***	1,3964
Partner's education level (ref=No education)					
Primary education	0,144755	0,156019	0,928	0,353537	1,1558
Secondary education	0,085726	0,147189	0,582	0,5603	1,0895
Higher	0,200253	0,167645	1,195	0,232314	1,2217
Don't know	-0,264443	0,420023	-0,63	0,528979	0,7676
Partner's occupation (ref=Unemployed)					
Employed	0,772655	0,230342	3,354	0,000799***	2,1655
Don't know	0,054147	0,080887	0,669	0,503248	1,0556
Woman's occupation (ref=Unemployed)					
Employed	0,192462	0,062245	3,092	0,001994**	1,2122
Don't know	-0,022983	0,141177	-0,163	0,870683	0,9773
Partner's age (ref=Less than 25 years old)					
25 to 34 years old	-0,267872	0,127977	-2,093	0,036368*	0,7650
Greater or equal to 35 years old	-0,422714	0,143649	-2,943	0,003263**	0,6553
The person who usually decides how to spend a woman's earnings (ref=Woman alone)					
Woman and partner	-0,154738	0,065915	-2,348	0,018921*	0,8566
Partner alone	-0,110197	0,116523	-0,946	0,344326	0,8957

Continued…

Woman's earnings compared to partner (ref=More than him)					
Less than him	0,12871	0,074828	1,72	0,085457.	1,1374
About the same	0,261424	0,09791	2,67	0,007599**	1,2988
Partner does not bring in earnings	0,299646	0,121062	2,475	0,013337*	1,3494
Don't know	0,356187	0,15807	2,253	0,024263*	1,4279
Knows STI's (ref=No)					
Yes	0,772655	0,230342	3,354	0,000799***	2,1655
Interaction effects					
Wealth index (ref=Poorest) by contraceptive use (ref=Not using)					
Poorer by using contraceptives	0,274623	0,1663	1,652	0,098473.	1,3160
Middle by using contraceptives	-0,059718	0,1636	-0,364	0,715679	0,9420
Richer by using contraceptives	0,235772	0,1721	1,367	0,171718	1,2659
Richest by using contraceptives	0,460516	0,2056	2,23	0,025785*	1,5849

Significance codes: ‘***’=0≤p-value<0.001, ‘**’=0.001≤p-value<0.01, ‘*’=0.01≤p-value<0.05, and ‘.’=0.05≤p-value<0.1

A woman who knows about STIs is 2.17 (OR=2.1655, p-value=0.000799***) times more likely to experience IPV compared to a woman with no knowledge of STIs.

Interaction Effects

Figure 5.1 shows that all women using contraceptives have a higher prevalence of IPV for all the wealth index classes, compared to those who are not using them.

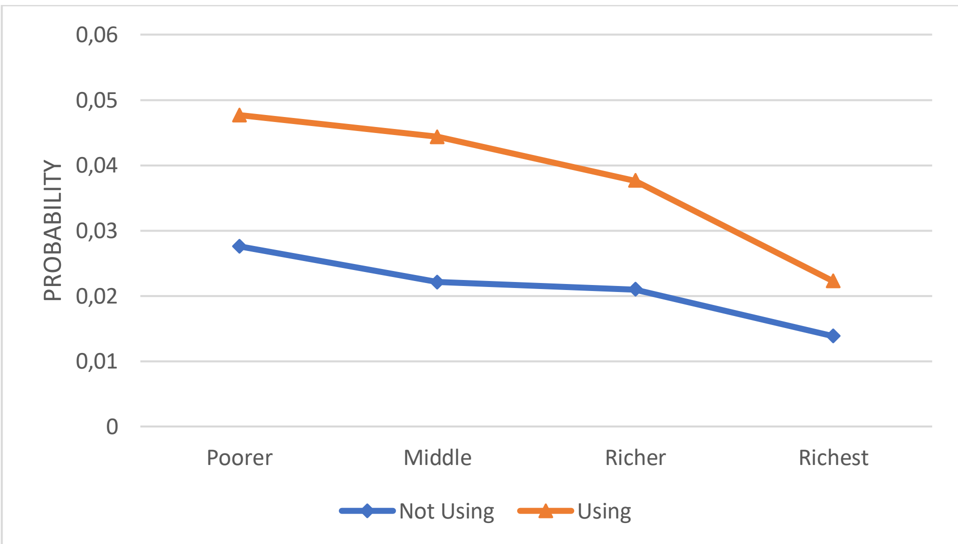

Figure 5.1: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Approximation Smooth Function

Table 5.2: Approximation significance of the smooth term

Smooth terms	Edf	F-value	p-value
S (Respondent's age)	8.702	19.18	<2e-16***

Table 5.2 shows that a woman's age has a significant impact on intimate partner violence. The letter **S** in Table 5.2 represents the smooth term, and the number in parenthesis shows the estimated degree of freedom (**edf**). The test statistics for woman's age (19.18) together with a p-value (<2e-16) show that there is no linear trend associated with IPV. This is confirmed in Figure 5.2, where the trend shows that as a woman's age increases, so does IPV, up to approximately 20 years old; it then remains almost constant until about the age of 38 years. It then starts to decrease from about 39 to 49 years of age.

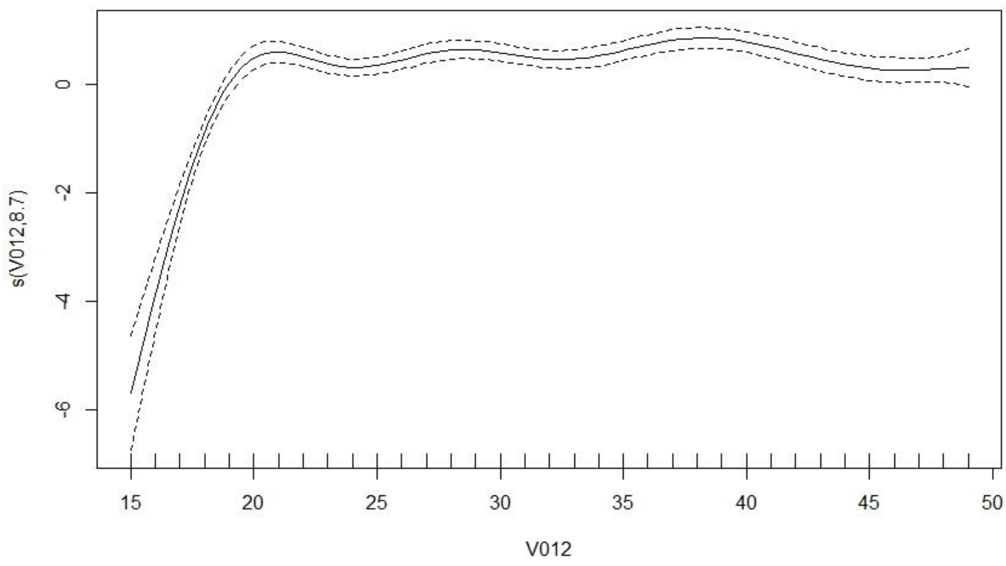

Figure 5.2: Smooth function of the age of the woman (V012)

Tanzania

Table 5.3 shows that a woman whose partner drinks alcohol is 2.60 (OR=2.5950, p-value<2e-16***) times more likely to experience IPV, compared to one whose partner does not drink it. A woman who has witnessed her father beating her mother is 1.71 (OR=1.7100, p-value<2e-16***) times more likely to experience IPV, compared to one who has not witnessed him beating her mother. A woman who does not know if her father beats her mother or not, is 1.52 (OR=1.5145, p-value<7.79e-07***) times more likely to experience IPV, compared to one whose father does not beat her mother. A woman who sees wife-beating as an acceptable act is 1.53 (OR=1.5293, p-value<2e-16***) times more likely to experience IPV, compared to one who sees it as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.74 (OR=0.7448, p-value=0.0197*) times less likely to experience IPV, compared to one who sees it as unacceptable. A woman from a household with five or more members is 0.34 (OR=0.3406, p-value<2e-16***) times less likely to experience IPV, compared to one from a household with five or fewer members. A woman with high exposure to the media is 0.74 (OR=0.7385, p-value=0.0364*) times less likely to experience IPV, compared to one with low exposure to the media.

A woman from Kilimanjaro province is 0.65 (OR=0.6447, p-value=0.0203*) times less likely to experience IPV compared to one from Dodoma province. A woman from Lindi province is 1.61 (OR=1.6072, p-value=0.0093**) times more likely to experience IPV compared to one from Dodoma province. A woman from Mtwara province is 1.60 (OR=1.5971, p-value=0.0113*) times more likely to experience IPV compared to one from Dodoma province. A woman from Tabora province is 1.40 (OR=1.4026, p-value=0.0454*) times more likely to experience IPV compared to one from Dodoma province. A woman from Shinyanga province is 1.52 (OR=1.5215, p-value=0.0144*) times more likely to experience IPV compared to one from Dodoma province. A woman from Mwanza province is 1.50 (OR=1.5031, p-value=0.0208*) times more likely to experience IPV compared to one from Dodoma province. A woman from Mara province is 1.91 (OR=1.9098, p-value=0.0002***) times more likely to experience IPV compared to one from Dodoma province. A woman from Manyanga province is 1.64 (OR=1.6395, p-value=0.0062**) times more likely to experience IPV compared to one from Dodoma province. A woman from Njobe province is 0.66 (OR=0.6633, p-value=0.0250*) times less likely to experience IPV compared to one from Dodoma province. A woman from Geita province is 1.40 (OR=1.4023, p-value=0.0493*) times more likely to experience IPV compared to one from Dodoma province. A woman from Kaskazini Unguja province is 1.88 (OR=1.8757, p-value=0.00099***) times more likely to experience IPV compared to one from Dodoma province. A woman from Mjini Magharibi province is 1.78 (OR=1.7841, p-value=0.0011**) times more likely to experience IPV compared to one from Dodoma province. A woman from Kaskazin Pemba province is 0.52 (OR=0.5248, p-value=0.0016**) times less likely to experience IPV compared to one from Dodoma province.

A woman with primary education is 1.21 (OR=1.2055, p-value=0.0265*) times more likely to experience IPV compared to one who has none. A woman who has a higher education is 0.50 (OR=0.5030, p-value=0.0058**) times less likely to experience IPV compared to one who has no education. A woman who has terminated a pregnancy is 1.13 (OR=1.1260, p-value=0.0375*) times more likely to experience IPV, compared to one who has never terminated a pregnancy. A healthy woman is 1.21 (OR=1.2101, p-value=0.0278*) times more likely to experience IPV compared to an underweight woman. A married woman is 4.87 (OR=4.8652, p-value<2e-16***) times more likely to experience IPV compared to a single woman. A woman living with her partner is 4.42 (OR=4.4229, p-value<2e-16***) times more likely to experience IPV compared to a single woman. A woman who does not know if her partner has other wives/partners or not is 1.79 (OR=1.7887, p-value=0.0157times more likely to experience IPV, compared to one whose partner does not have other wives/partners.

A woman with a partner who wants more children than her is 1.17 (OR=1.1663, p-value=0.0073**) times more likely to experience IPV, compared to one whose partner wants the same number as her. A woman with a partner who wants fewer children than her is 1.35 (OR=1.3480, p-value=0.0028**) times more likely to experience IPV, compared to one whose partner wants the same number as her. A woman who does not know the number of children her partner wants is 1.29 (OR=1.2867, p-value=5.96e-06***) times more likely to experience IPV, compared to one whose partner wants the same number as her.

Table 5.3: The parameter estimates of the IPV for the fixed part of GAMM

Effect	**Estimate**	**Standard error**	**t-value**	**P-value**	**Odds ratio**
Intercept	-2,0863	0,2892	-7,22	5,69e-13***	0,1241
Partner drinks alcohol (ref=No)					
Yes	0,9536	0,0556	17,14	<2e-16***	2,595
Woman's father ever beat her mother (ref=No)					
Yes	0,5365	0,0504	10,64	<2e-16***	2,595
Don't know	0,4151	0,084	4,94	7,79e-07***	1,5145
Access to the media (ref=Low exposure)					
Medium exposure	0,023	0,0509	0,45	0,651	1,0233
High exposure	-0,3032	0,1449	-2,09	0,0364*	0,7385
Wife-beating attitude (ref=Unacceptable)					
Acceptable	0,4248	0,0475	8,94	<2e-16***	1,5293
Don't know	-0,2947	0,1263	-2,33	0,0197*	0,7448
Number of household members (ref=less than 5)					
More than or equal to 5	-1.0769	0.0472	-22.82	<2e-16***	0.3406
Region (ref=Dodoma)					
Arusha	0,2292	0,179	1,28	0,2004	1,2576
Kilimanjaro	-0,4389	0,189	-2,32	0,0203*	0,6447
Tanga	-0,2892	0,1753	-1,65	0,0989	0,7489
Morogoro	0,2606	0,1857	1,4	0,1605	1,2977
Pwari	0,2935	0,1851	1,59	0,1128	1,3411
Dar Es Salaam	0,2892	0,1664	1,74	0,0823	1,3354
Lindi	0,4745	0,1823	2,6	0,0093**	1,6072
Mtwara	0,4682	0,1849	2,53	0,0113*	1,5971
Ruvuma	0,067	0,1813	0,37	0,7119	1,0693
Iringa	-0,0723	0,1907	-0,38	0,7046	0,9303
Mbeya	0,0857	0,186	0,46	0,645	1,0895
Singida	0,0538	0,1786	0,3	0,7631	1,0553
Tabora	0,3383	0,1691	2	0,0454*	1,4026
Rukwa	0,0949	0,1793	0,53	0,5967	1,0995
Kigoma	0,0805	0,1719	0,47	0,6396	1,0838
Shinyanga	0,4197	0,1714	2,45	0,0144*	1,5215
Kagera	0,3513	0,1828	1,92	0,0546	1,4209
Mwanza	0,4075	0,1763	2,31	0,0208*	1,5031
Mara	0,647	0,1722	3,76	0,0002***	1,9098
Manyanga	0,4944	0,1807	2,74	0,0062**	1,6395

Continued…

Njobe	-0,4106	0,1832	-2,24	0,0250*	0,6633
Katavi	-0,0137	0,1755	-0,08	0,938	0,9864
Simiyu	0,2609	0,1684	1,55	0,1212	1,2981
Geita	0,3381	0,172	1,97	0,0493*	1,4023
Kaskazini Unguja	0,629	0,1909	3,3	0,00099***	1,8757
Kusini Unguja	0,6553	0,1943	3,37	0,000749***	1,9257
Mjini Magharibi	0,5789	0,1772	3,27	0,0011**	1,7841
Kaskazini Pemba	-0,6448	0,2047	-3,15	0,0016**	0,5248
Kusini Pemba	-0,268	0,1987	-1,35	0,1774	0,7649
Type of residence (ref= Rural)					
Urban	-0,0635	0,0658	-0,96	0,3348	0,9385
Woman's highest education level (ref=No education)					
Primary education	0,1869	0,0658	2,22	0,0265*	1,2055
Secondary education	-0,1773	0,0842	-1,64	0,1018	0,8375
Higher education	-0,6872	0,1083	-2,76	0,0058**	0,503
Sex of household head (ref=Male)					
Female	0,0716	0,064	1,12	0,2635	1,0742
Literacy (ref=Cannot read)					
Can Read	-0,062	0,0732	-0,85	0,3971	0,9399
Wealth index (ref=Poorest)					
Poorer	-0,081	0,0886	-0,92	0,3604	0,9222
Middle	-0,1153	0,0905	-1,27	0,2025	0,8911
Richer	-0,2036	0,096	-2,12	0,0340*	0,8158
Richest	-0,3128	0,114	2,08	0,0061**	0,7314
Ever had a terminated pregnancy (ref=No)					
Yes	0,1187	0,0571	2,08	0,0375*	1,126
Contraceptive use (ref=Not using)					
Using	0,3207	0,1321	2,43	0,0152*	1,3781
Current marital status (ref=Single)					
Married	1,5821	0,065	24,24	<2e-16***	4,8652
Living with partner	1,4868	0,078	19,45	<2e-16***	4,4229
Body Mass Index(ref=Underweight)					
Healthy	0,1907	0,0866	2,201	0,0278*	1,2101
Overweight	0,1827	0,0963	1,898	0,0578.	1,2005
Obese	0,0849	0,1073	0,791	0,4288	1,0886
Number of other wives/partners (ref=No other wives/partners)					
One or more	0,0614	0,0559	1,099	0,2717	1,0633
Don't know	0,5815	0,2406	2,417	0,0157*	1,7887
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	0,008997	0,091381	0,098	0,9216	1,009
Partner's desire for children (ref=Both want same number)					
Partner wants more	0,1538	0,0574	2,68	0,0073**	1,1663

Continued…

Partner wants fewer	0,2986	0,0999	2,99	0,0028**	1,348
Don't know	0,2521	0,0555	4,54	5,69e-06***	1,2867
Partner's age (ref=Less compared to 25)					
Between 25 and 34	0,0454	0,0891	0,51	0,6103	1,0464
35 and above	-0,1824	0,1066	-1,71	0,0871.	0,8333
Partner's education level (ref=No education)					
Primary education	-0,0731	0,0761	-0,96	0,3369	0,9295
Secondary education	-0,1057	0,0926	-1,14	0,2537	0,8997
Higher	-0,3336	0,1544	-2,16	0,0308*	0,7163
Don't know	-0,2044	0,371	-0,55	0,5818	0,8151
Partner's occupation (ref=Unemployed)					
Employed	-0,3871	0,2233	-1,734	0,083	0,679
Woman's occupation (ref=Unemployed)					
Employed	0,2241	0,0633	3,54	0,000404***	1,2512
Woman's earnings compared to partner (ref=More than him)					
Less than him	0,205	0,0833	2,46	0,0139*	1,2275
About the same	0,2129	0,0928	2,295	0,0218*	1,2373
Partner does not bring in earnings	0,2382	0,2515	0,947	0,3437	1,269
Don't know	0,1965	0,1384	1,42	0,1558	1,2171
The person who usually decides how to spend respondent's earnings (ref=Respondent alone)					
Woman and partner	-0,2313	0,0509	-4,54	5,61e-06***	0,7935
Partner alone	-0,3814	0,0835	-4,57	4,98e-06***	0,6829
Someone else	-0,8631	1,4434	-0,598	0,5499	0,4219
Interaction effects					
Wealth index (Ref=Poorest) by contraceptive use (Ref=Not using)					
Poorer by using contraceptives	0,00418	0,1751	0,024	0,9809	1,0042
Middle by using contraceptives	0,1911	0,1709	1,118	0,2636	1,2106
Richer by using contraceptives	0,1362	0,1621	0,84	0,4009	1,1459
Richest by using contraceptives	-0,1221	0,1607	-0,76	0,4472	0,8851

A woman who has a partner with a higher education is 0.7163 (OR=0.7163, p-value=0.0308*) times less likely to experience IPV, compared to one who has a partner with no education. An employed woman is 1.25 (OR=1.2512, p-value=0.000404***) times more likely to experience IPV compared to one who is unemployed. A woman who earns less than her partner is 1.23 (OR=1.2275, p-value=0.0139*) times more likely to experience IPV, compared to one who earns more than him. A woman who earns about the same as her partner is 1.24 (OR=1.2373, p-value=0.0218*) times more likely to experience IPV, compared to one who earns more than him. A woman who decides with her partner how to spend her earnings is 0.79 (OR=0.7935, p-value=5.61e-06***) times less likely to experience IPV, compared to one who decides how to spend them alone. A woman whose partner alone decides on what to do with her earnings is 0.68 (OR=0.6829, p-value=4.98e-06***) times less likely to experience IPV, compared to one who decides how to spend them alone.

Approximation Smooth Function

Table 5.4: Approximation significance of the smooth term

Smooth terms	edf	F-value	p-value
S (Respondent's age)	6.732	46.94	<2e-16***

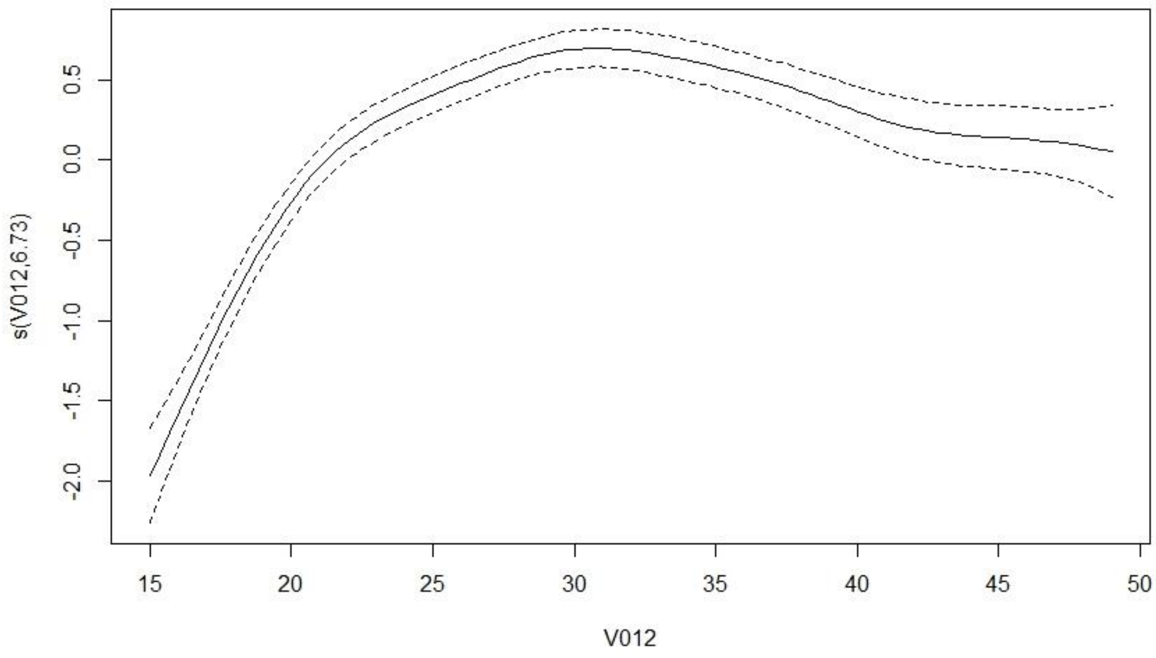

Figure 5.3: Smooth function of the age of a woman (V012)

Table 5.4 shows that a woman's age has a significant impact on intimate partner violence. The letter **S** in Table 5.4 represents the smooth term, and the number in parenthesis shows the estimated degree of freedom (**edf**). The test statistics for woman's age (46.94) together with a p-value (<2e-16) show that there is no linear trend associated with IPV. Figure 5.3 shows that IPV increases as a woman's age increases from 15 years to approximately 32 years of age and starts to decrease from about the age of 33 years until 49 years.

Uganda

Table 5.5 shows that a woman whose partner drinks alcohol is 2.54 (OR=2.5393, p-value<2e-16***) times more likely to experience IPV, compared to one whose partner does not drink it. A woman who has witnessed her father beating her mother is 1.77 (OR=1.7736, p-value<2e-16***) times more likely to experience IPV, compared to one who has never witnessed her father beating her mother. A woman who does not know if her father beats her mother is 1.84 (OR=1.8360, p-value=6.25e-11***) times more likely to experience IPV, compared to one whose father has never beaten her mother. A woman from a household with more than five members is 0.55 (OR=0.5496, p-value<2e-16***) times less likely to experience IPV, compared to one from a household with less than five members. A woman who sees wife-beating as an acceptable act is 1.16 (OR=1.1638, p-value=6.22e-05***) times more likely to experience IPV, compared to one who sees it as unacceptable. A woman who does not know if wife-beating is acceptable or not is 0.63 (OR=0.6278, p-value=0.005756**) times less likely to experience IPV, compared to one who sees it as an unacceptable act. A woman with high exposure to the media is 0.70 (OR=0.6964, p-value=0.0005604**) times less likely to experience IPV, compared to one with low exposure to the media.

A woman from Busoga province is 1.25 (OR=1.2506, p-value=0.042770*) times more likely to experience IPV compared to one from Kampala province. A woman from Teso province is 0.79 (OR=0.7919, p-value=0.048915*) times less likely to experience IPV compared to one from Kampala province. A woman from Karamoja province is 0.56 (OR=0.564452, p-value=5.71e-05***) times less likely to experience IPV compared to one from Kampala province. A woman from Lango province is 0.77 (OR=0.7690, p-value=0.027850*) times less likely to experience IPV compared to one from Kampala province. A woman from Acholi province is 0.72 (OR=0.7178, p-value=0.006486**) times less likely to experience IPV compared to one from Kampala province. A woman from Bunyoro province is 0.68 (OR=0.6776, p-value=0.000954***) times less likely to experience IPV compared to one from Kampala province.

A woman from a household where the head of the house is female is 0.84 (OR=0.8437, p-value=0.000278***) times less likely to experience IPV, compared to one from a household where the head is male. A woman who has terminated a pregnancy is 1.12 (OR=1.1197, p-value=0.010853*) times more likely to experience IPV, compared to one who has never terminated a pregnancy. A woman who is overweight (25≤BMI<30) is 0.82 (OR=0.8237, p-value=0.012199*) times less likely to experience IPV compared to one who is underweight (BMI<18). A married woman is 2.76 (OR=2.7649, p-value<2e-16***) times more likely to experience IPV compared to a single woman. A woman living with her partner is 2.81 (OR=2.8123, p-value<2e-16***) times more likely to experience IPV compared to a single woman.

A woman whose partner has more than one wife is 1.14 (OR=1.1421, p-value=0.002137**) times more likely to experience IPV, compared to one whose partner does not have other wives. A woman who does not know if her partner has other wives or not is 1.51 (OR=1.5089, p-value=5.54e-06***) times more likely to experience IPV, compared to one whose partner does not have other wives/partners.

A woman who does not know the number of children her partner wants is 0.85 (OR=0.8517, p-value=0.001363**) times less likely to experience IPV, compared to one whose partner wants the same number of children as her. A woman who has a partner with secondary education is 0.81 (OR=0.8120, p-value=0.012295**) times less likely to experience IPV, compared to one whose partner has no education. A woman whose partner has a higher education level is 0.73 (OR=0.7318, p-value=0.001321**) times less likely to experience IPV, compared to one whose partner has no education. A woman who does not know her partner's level of education is 0.55 (OR=0.5545, p-value=2.10e-05***) times less likely to experience IPV, compared to one whose partner has no education.

An employed woman is 1.13 (OR=1.1379, p-value=0.022295*) times more likely to experience IPV compared to one who is unemployed. A woman who decides with her partner what to do with her earnings is 0.83 (OR=0.8266, p-value=2.98e-06***) times less likely to experience IPV, compared to one who solely decides how to spend them. A woman whose partner decides how to spend her earnings is 0.70 (OR=0.6984, p-value=3.95e-07***) times less likely to experience IPV, compared to one who decides alone how to spend them. A woman who earns about the same as her partner is 0.69 (OR=0.6855, p-value=1.16e-06***) times less likely to experience IPV, compared to one who earns more than him. A woman whose partner does not bring in earnings is 0.65 (OR=0.6473, p-value=0.021089*) times less likely to experience IPV, compared to one who earns more than him.

Table 5.5: The parameter estimates of IPV for the fixed part of GAMM

Effect	**Estimate**	**Standard error**	**t-value**	**P-value**	**Odds ratio**
Intercept	-0,258	0,4997	-5,163	2,46e-07***	0,7725952
Partner drinks alcohol (ref=No)					
Yes	0,9319	0,038	24,524	<2e-16***	2,5393293
Woman's father ever beat her mother (ref=No)					
Yes	0,573	0,03855	14,865	<2e-16***	1,7735798
Don't know	0,6076	0,09288	6,541	6,25e-11***	1,8360197
Number of household members (ref=more than 5)					
Less than 5	-0,5985	0,03773	-15,861	<2e-16***	0,5496355
Wife beating attitude (ref=Unacceptable)					
Acceptable	0,1517	0,03789	4,005	6,22e-05***	1,163811
I don't know	-0,4655	0,03789	-2,762	0,005756**	0,6278211
Access to the media (ref=Low exposure)					
Medium exposure	-0,03716	0,04047	-0,785	0,432592	0,963522
High exposure	-0,3618	0,01306	-2,77	0,0005604**	0,6964216
Region (ref=Kampala)					
South Buganda	-0,01937	0,104	-0,186	0,852171	0,9808164
North Buganda	-0,0168	0,1093	-0,154	0,877885	0,9833403
Busoga	0,2236	0,1103	2,026	0,042770*	1,2505707
Bukedi	0,03018	0,1166	0,259	0,795697	1,03064
Busigu	-0,1342	0,1208	-1,111	0,266778	0,8744152

Continued…

Teso	-0,2333	0,1185	-1,969	0,048915*	0,791916
Karamoja	-0,5719	0,1421	-4,025	5,71e-05***	0,564452
Lango	-0,2627	0,1195	-2,198	0,027950*	0,7689726
Acholi	-0,3315	0,1218	-2,722	0,006486**	0,7178462
West Nile	-0,06942	0,1172	-0,592	0,553707	0,9329348
Bunyoro	-0,3892	0,1178	-3,304	0,000954***	0,6775987
Tooro	-0,01282	0,1135	-0,113	0,910043	0,9872618
Ankole	0,2093	0,1125	1,861	0,062817.	1,2328148
Kigezi	-0,1163	0,1221	-0,952	0,34094	0,8902081
Type of residence(ref=Rural)					
Urban	-0,1122	0,05547	-2,023	0,043052*	0,8938655
Woman's highest education level (ref=No education)					
Primary education	0,07356	0,06462	1,138	0,255014	1,0763331
Secondary education	0,09883	0,08421	1,174	0,240579	1,1038786
Higher education	0,08128	0,1171	0,694	0,487742	1,0846746
Sex of household head (ref=Male)					
Female	-0,1697	0,04667	-3,636	0,000278***	0,8439
Literacy (ref=Cannot read)					
Can read	-0,01623	0,04681	-0,347	0,728867	0,9839
Wealth index (ref=Poorest)					
Poorer	0,02224	0,0684	0,325	0,745029	1,0225
Middle	-0,09092	0,07466	-1,218	0,223279	0,9131
Richer	-0,1532	0,07973	-1,921	0,054753.	0,8580
Richest	-0,4164	0,0956	-4,356	1,33e-05***	0,6594
Ever had a terminated pregnancy (ref=No)					
Yes	0,1131	0,04441	2,548	0,010853*	1,1197
Contraceptive use (ref=Not using)					
Using	0,3338	0,09199	3,628	0,000286***	1,3963
Body Mass Index (ref=Underweight (BMI<18))					
Healthy (18≤BMI<25)	-0,07095	0,06529	-1,087	0,277192	0,9315
Overweight (25≤BMI<30)	-0,194	0,07739	-2,507	0,012199*	0,8237
Obese (BMI≥30)	-0,1525	0,09212	-1,655	0,097855.	0,8586
Current marital status (ref=Single)					
Married	1,017	0,05666	17,945	<2e-16***	2,7649
Living with partner	1,034	0,05506	18,787	<2e-16***	2,8123
Number of other wives (ref=0)					
One	0,1329	0,04328	3,071	0,002137**	1,1421
Don't know	0,4114	0,09052	4,545	5,54e-06***	1,5089
Cohabitation period (ref=0 to 4 years)					
5 to 9 years	0,06102	0,07522	0,811	0,417251	1,0629
Partner's desire for children (ref=Both want same)					
Husband wants more	0,01193	0,04457	0,268	0,788976	1,0120

Continued…

Husband wants fewer	-0,1247	0,06637	-1,88	0,060177.	0,8828
Don't know	-0,1605	0,0501	-3,203	0,001363**	0,8517
Partner's education level (ref=No education)					
Primary education	-0,1445	0,0765	-1,888	0,058979.	0,8655
Secondary education	-0,2082	0,08314	-2,504	0,012295*	0,8120
Higher	-0,3123	0,1	-3,212	0,001321**	0,7318
Don't know	-0,5896	0,1386	-4,255	2,10e-05***	0,5545
Partner's occupation (ref=Unemployed)					
Employed	0,1322	0,1028	1,285	0,198802	1,1413
Woman's occupation (ref=Unemployed)					
Employed	0,1204	0,05268	2,285	0,022295*	1,1279
Partner's age (ref=Less than 25 years old)					
25 to 34 years old	0,004632	0,07328	0,063	0,949605	1,0046
Greater or equal to 35 years old	0,09827	0,08674	1,133	0,257276	1,1033
The person who usually decide on a woman's earnings (ref=Woman alone)					
Woman and partner	-0,1904	0,04073	-4,674	2,98e-06***	0,8266
Partner alone	-0,3589	0,07074	-5,073	3,95e-07***	0,6984
Someone else	0,08827	0,5098	0,173	0,862538	1,0923
Woman's earnings compared to partner (ref=More than him)					
Less than him	-0,01754	0,06062	-0,289	0,772305	0,9826
About the same	-0,3776	0,07762	-4,864	1,16e-06***	0,6855
Partner does not bring in earnings	-0,435	0,1886	-2,307	0,021089*	0,6473
Don't know	-0,2642	0,1443	-1,83	0,067213.	0,7678
Knows STI's (ref=No)					
Yes	0,2874	0,4959	0,58	0,562199	1,3330
Interaction effects					
Wealth index (ref=poorest) by contraceptive use (ref=not using)					
Poorer by using contraceptives	-0,3365	0,1243	-2,708	0,006785**	0,7143
Middle by using contraceptives	-0,4226	0,1248	-3,386	0,000711***	0,6553
Richer by using contraceptives	-0,3729	0,1245	-2,996	0,002743**	0,6887
Richest by using contraceptives	-0,2243	0,1241	-1,808	0,070664.	0,7991

Significance codes: ‘***’=0≤p-value<0.001, ‘**’=0.001≤p-value<0.01, ‘*’=0.01≤p-value<0.05, and ‘.’=0.05≤p-value<0.1

Interaction Effects

Figure 5.4 we observe that women who are using contraceptives have a higher prevalence of IPV for all the wealth index classes, compared to those who are not using them.

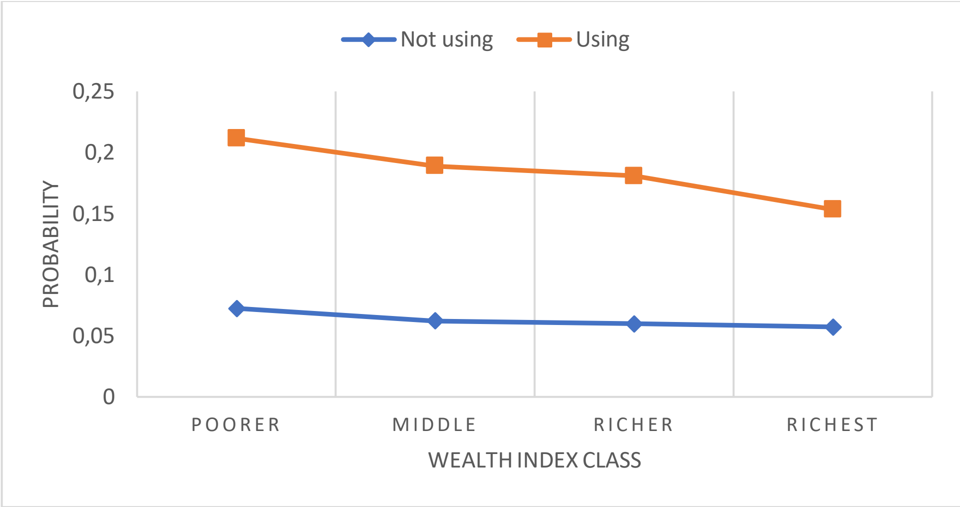

Figure 5.4: Predicted probability of experiencing IPV by wealth index class and contraceptive use

Approximation Smooth Function

Table 5.6 shows that a woman's age has a significant impact on intimate partner violence. The letter **S** in Table 5.6 represents the smooth term, and the number in parenthesis shows the estimated degree of freedom (**edf**). The test statistics for woman's age (33.92) together with a p-value (<2e-16) show that there is no linear trend associated with IPV. Figure 5.5 shows that the relationship between a woman' age and an increase in IPV. As a woman's age increases to approximately 20 years of age, IPV increases sharply. Thereafter, it remains constant until about 35 years of age, and starts to decrease until the age of about 40 years; from then it starts to show an increase again.

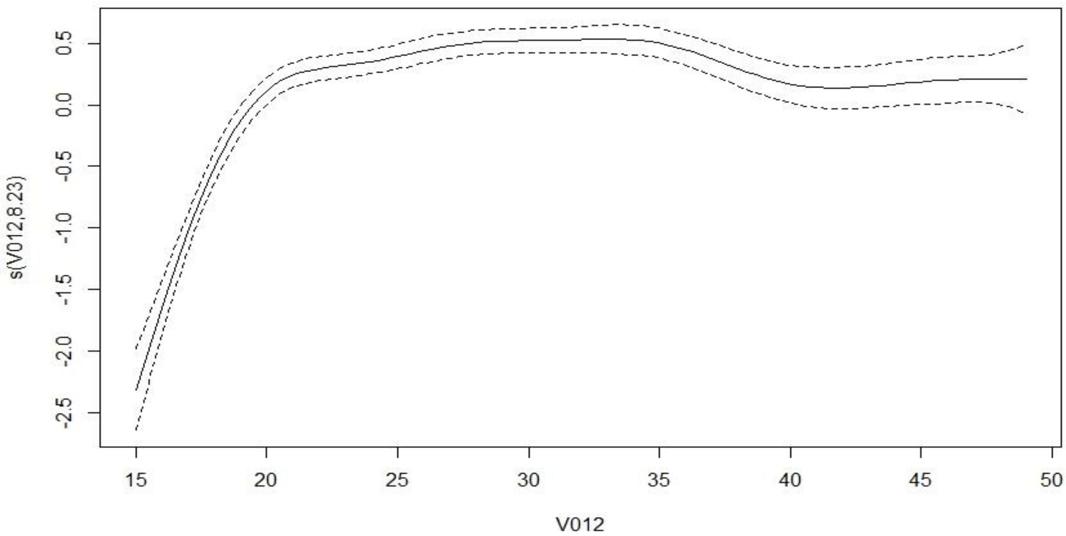

Figure 5.5: Smooth function of the age of a woman (V012)

Table 5.6: Approximation significance of the smooth term

Smooth terms	Edf	F-value	p-value
S (Respondent's age)	8.232	33.92	<2e-16***

5.5 Summary

In this chapter, we have used GAMM to identify the risk factors associated with IPV for women of reproductive age in South Africa, Tanzania, and Uganda. The linearity inherent in GLM and the assumption of normality, are easily relaxed by using a generalized additive mixed model. The flexibility of nonparametric regression for continuous predictor variables, coupled with GLM for predictor variables, offer ways to reveal structure within the data, that may miss GLM assumptions. The results from the generalized additive mixed model give more insight and understanding, especially concerning the distribution of continuous covariates. For all three countries, the results from the parametric part supported IPV being high for: women with partners who drink alcohol, respondents whose father beat their mother, a respondent's media exposure status, the region in which a respondent resides, and their wealth index. This study also confirmed that in all three countries, IPV is high for women who have had a pregnancy terminated, those who are married, different wishes between respondent and her partner regarding desire for children, and respondent's earnings compared to her partner's.

In addition to the above determinants, in Tanzania and Uganda, the results from the parametric part supported that IPV is high in women who use contraceptive methods, those whose partners have other wives/partners, and a partner's education level. In South Africa and Uganda, the results from the parametric part supported that IPV is high in women from a household where the head of the household is male, and was linked to a partner's occupation status, and whether the respondent could or could not refuse sex to her partner. In South Africa and Tanzania, the results from the parametric part supported that IPV is high in women who are either employed or not. In South Africa and Tanzania, the results from the parametric part supported that IPV is affected by the respondent's knowledge of STIs and the respondent's education level, respectively.

This study also confirmed that IPV decreases with an increase in household members in all three countries, and with a decrease in body mass index in South Africa and Uganda. This study also showed that in all three countries, IPV increases with an increase in a woman's age. In addition, the findings from this study showed that IPV is affected by a partner's age.

Chapter 6

Discussion and Conclusion

Whether a woman has experienced intimate violence or not is commonly measured by the type of violence a woman experiences, which is inflicted by her partner. However, collecting data on individual, personal information may be challenging since it might be difficult for women to disclose this, in fear of confidentiality from the party conducting the interview. Additionally, some of the questions may be about past experiences, making it challenging to infer the future. The study's objective was to develop an alternative method to determine which kind of women are more likely to experience intimate partner violence. It also aimed to examine the various statistical methods suitable to identify the risk factors associated with intimate partner violence, for women of reproductive age. To achieve these objectives, a binary response variable was created using the different types of violence (emotional, physical, and psychological) experienced by individual women. Thereafter, based on whether the woman had experienced at least one of these, she was classified as having experienced intimate partner violence or not.

We fitted various statistical models to individual data. Logistic regression and survey logistic regression were first applied to the women's data to identify critical determinants of intimate partner violence, and their results were compared. Comparing the results showed that the sampling weights and sampling stratification significantly affect parameter estimates and standard errors. Therefore, it is better to use survey logistic compared to binary logistic regression, when the data was collected under a multi-stage stratified sampling design, to get a valid statistical inference.

Generalized linear models and generalized linear mixed models estimated how the independent variables are related to the mean value of the response variable. GLMMs were used over logistic regression and survey logistic regression, since they account for variability amongst the units. A GLMM is also more useful, since it takes random effects into account.

By relaxing the assumption of normality and linearity inherent in GLM, we used a generalized additive mixed model (semiparametric); the categorical covariates were modeled parametrically and continuous covariates non-parametrically. A GAMM can reveal some information that may be hidden when only parametric models are used. The findings from all these models are discussed below.

The factors associated with intimate partner violence against women of reproductive age are different in all three countries, but some are common in all of them. For instance, a woman who had a partner who does not drink alcohol was at a lower risk of experiencing intimate partner violence, compared to one who had a partner who drank alcohol. This result is consistent with other findings from previous studies (Ali, et al., 2014; Gage, 2005; Habyarimana, et al., 2018; Obi & Ozumba, 2007). The study's findings also revealed that a woman who had never witnessed her father beating her mother was at low risk of experiencing IPV, compared to one who had witnessed her father beating her mother. In Uganda, the more a woman was exposed to the media, the lesser the risk of experiencing IPV, whereas, in South Africa, the opposite was true.

One of the study's key findings was that as a woman age, she is at a higher risk of experiencing IPV in both South Africa and Uganda. Similar findings were found by Bonomi, et al (2007) and Obi & Ozumba (2007). The region in a woman lives was statistically significant too for both countries. These findings were also found in a study by Habyarimana et al. (2018). It is common in these two countries that a woman from a household with more members is at a lower risk of experiencing IPV, compared to a household with fewer members. In both South Africa and Uganda, if a woman was from a household where the sex of the household head was male, the risk of experiencing IPV was high, compared to a household where the head was female.

In all three countries, a woman from the poorest, poorer, middle, and richer classes was at high risk of experiencing IPV compared to a woman from the richest class. This is supported by a study by Bamiwuye & Odimegwu (2014) that found that domestic violence was higher among women from rich families in Mozambique and Zambia. This study also found that violence in Nigeria and Cameroon was higher in the middle classes, and higher in low-income families in Kenya and Zimbabwe. A woman who had never terminated a pregnancy was at lower risk of experiencing IPV, than one who ds had a terminated pregnancy. A woman who did not use contraceptives was at lower risk of experiencing IPV, compared to one who did use them. In a study by Koeing et al. (2003), this variable was used and was found to be statistically significant in their model. In South Africa, a woman with a higher body mass index was at lower risk of experiencing IPV compared to one with a lower body mass index. In contrast, in Uganda and Tanzania, the BMI of a woman was statistically insignificant.

A married woman or one staying with her partner was at higher risk of experiencing IPV compared to a single woman. Similar results were found in a study by Usta et al. (2007). The study also found that the number of other wives her partner might have, was statistically insignificant in South Africa. In Uganda, a woman whose partner had other wives/partners was at higher risk of experiencing IPV, compared to a woman whose partner had no other wives/partners. Similar findings were noted by Koeing et al. (2003). A woman who had been staying with her partner for 5-9 years was at lower risk of experiencing IPV, compared to one who had stayed with him for a period of fewer than five years. This confirmed the findings of Koeing et al. (2003). In South Africa, if a woman's partner wanted the same number of children, she was at lower risk of experiencing IPV, compared to one whose partner wanted more or fewer children, or in such cases where the woman did not know her partner's desire for children. A woman from Uganda whose partner wanted more children than her, was at a higher risk of experiencing IPV, compared to one whose partner wanted the same number as her. Likewise in Uganda, a woman who wanted the same number of children as her partner was at higher risk of experiencing

IPV, compared to one whose partner wanted fewer or did not know her partner's desire for children. A partner's level of education was found to be statistically insignificant in South Africa. In Uganda, however, a woman who had a partner with no education was at a higher risk of experiencing IPV, compared to one whose partner had a primary, secondary or higher education; the same was the case for a woman who did not know her partner's level of education. Similar findings were confirmed by Usta et al. (2007), and Habyarimana et al. (2018).

The study's findings also revealed that if a woman's partner was employed, she was at a lower risk of experiencing IPV. Similar results were found in a study by Koeing, et al. (2003). In Uganda, a woman's employment status and her partner's age were statistically insignificant. In South Africa, an employed woman was at higher risk of experiencing IPV compared to an unemployed woman. The study's findings also suggest that in South Africa, the older the woman's partner, the lower was the risk of experiencing IPV. In the study mentioned above by Koeing et al. (2003), similar results were found. As a woman's earnings got higher compared to those of her partner, the risk of IPV got lower. The above finding is similar to the ones by Obi & Ozumba (2007). This study's findings also show that women with knowledge of sexually transmitted infections were at a higher risk of experiencing IPV. However, it was found that the same knowledge was statistically insignificant in Uganda. Women from South Africa who could not refuse sex from their partner had a lower risk of experiencing IPV, but in Uganda, the opposite was true. An increase in a woman's age and not using contraceptive methods increased the risk of experiencing IPV. The study's findings also demonstrate that women from the different wealth index levels who were not using contraceptive methods, were at higher risk of experiencing IPV, compared those from the richest wealth index class, also not using these methods.

The study also revealed some common determinants between South Africa and Uganda: a woman's partner alcohol consumption status, if a woman's father ever beat her mother, the region in which a woman lives, the number of household members, and the current marital status of a woman. In addition to the above factors, these two countries share other factors in common: the sex of the household head, the wealth index, whether a woman had terminated a pregnancy, and her occupation.

This current study highlights novel findings, such as knowledge of sexually transmitted infections and contraceptive methods used by women, as significant IPV factors. Perhaps governments need to educate couples contemplating marriage and married couples, through going on a short course that addresses these issues. Likewise, religious organizations can attempt to assist couples at a grassroots level.

The findings also suggest that knowledge of sexually transmitted infections may have a significant impact on intimate partner violence. A woman's exposure to the media could help in reducing the high rate of IPV. This study’s findings further suggest that polygamy should be discouraged, and policymakers should encourage men and women to learn more about contraceptive methods, sexually transmitted infections, and the concept of rape. A recommendation of this study, based on its findings, is that women and men should be taught intimate partner violence at an early age, so as to avoid the high increase in violent cases as a woman grows older. The current work also revealed that motivating women to empower themselves and be independent, might reduce the rate of IPV. Women may be encouraged to pursue their studies and open a business, which could help them earn a living and be independent. The policymakers could use different platforms to engage with the targeted group of individuals. Some of these platforms and possible ways to address IPV, could be through social media, radio talk shows where women can talk about their experiences anonymously, television documentaries with willing participants outlining the different types of violence and some of their health consequences.

The current study used the DHS cross-sectional data sets, and this type of data may not address specific issues, such as causality. For future research, a longitudinal study may be more apt to determine causality.

References

Abrahams, N., Jewkes, R., Laubscher, R. & Hoffman, M., 2006. Intimate partner violence: Prevalence and risk factors for men in Cape Town, South Africa. *Violence Vict,* Volume 21, pp. 247-264.

Adjah, E. S. & Agbemafle, I., 2016. Determinants of domestic violence against women in Ghana. *BMC public health,* Volume 16(1), p. 368.

Ajah, L. et al., 2014. Comparison of domestic violence against women in urban versus rural areas of southeast Nigeria. *International Journal in Women's Health,* Volume 6, pp. 865-872.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control,* 19(6), pp. 716-723.

Ali, A., Yassin, K. & Omer, R., 2014. Domestic violence against women in Eastern Sudan. *BMC public health,* 14(1), pp. 1114-1136.

An, A. B., 2002. Performing logistic regression on survey data with the new survey logistic procedure. *In Proceedings of the twenty-seventh annual SAS users group international conference,* pp. 258-270.

Audi, C. A. et al., 2008. Violence against pregnant women: prevlence and associated factors. *Revista de Saude publica,* Volume 42(5), pp. 877-885.

Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J., 2011. Multiple imputation by chained equations: what is it and how does it work?. *International Journal of Methods in Psychiatric Research,* 20(1), pp. 40-49.

Bamiwuye, S. O. & Odimegwu, C., 2014. Spousal violence in sub-Saharan Africa: does household poverty-wealth matter?. *Reproductive health,* 11(1), p. 45.

Bolker, B. M. et al., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24(3), pp. 127-135.

Bonomi, A. E. et al., 2007. Intimate Partner Violence in Older Women. *The Gerontologist,* 47(1), pp. 34-41.

Bowman, C. G., 2003. Domestic Violence: Does the African Context Demand a Different Approach?. *International Journal of Law and Psychiatry,* Volume 26, pp. 473-491.

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

Bridgette, 2013. *Discussions regarding The DHS program data and results.* [Online] Available at: https://userforum.dhsprogram.com/index.php?t=msg&goto=220&S=Google

Brink, O., Vesterby, A. & Jensen, J., 1998. Pattern of injuries due to interpersonal violence. *Injury,* 29(9), pp. 705-709.

Chapman, H. & Gillespie, S. M., 2018. The Revist Conflict tactics Scales (CTS2): A review of the properties, reliability, and validity of the CTS2 as a measure of partner abuse in community and clinical samples. *Aggression and Violent Behavior,* Volume 44, pp. 27-35.

Choi, K. H., Hoff, C. & Gregorich, S. E., 2008. The efficacy of female condom skills training in HIV risk reduction among women: a randomized controlled trial. *Am J Public Health,* 98(10), pp. 1841-1848.

Coker , A. & Richter, D., 1998. Violence against women in Sierra Leone: frequency and correlates of intimate partner violence and forced sexual intercourse.. *African Journal of Reproductive Health,* Volume 2, pp. 61-72.

Cucchiara, A., 2012. Applied Logistic Regression. *Technometrics.*

Devries, K. et al., 2010. Intimate partner violence during pregnancy: analysis of prevalence data from 19 countries. *Reprod Health Matters,* Volume 18(36), pp. 158-170.

DHSProgram, 2016. [Online] Available at: www.dhsprogram.com [Accessed 2021].

Dlamini, W. J., 2016. *Statisticals models to understand factors associated with under-five child mortality in Tanzania,* Pietermaritzburg: University of Kwazulu-Natal.

Ellsberg, M. et al., 1999. Wife abuse among women of childbearing age in Nicaragua. *American Journal of Public Health,* Volume 89:24, pp. 1-4.

Finnbogadottir, H., Dykes, A.-K. & Wann-Hansson, C., 2014. Prevalence of domestic violence during pregnancy and related risk factors: a cross-sectional study in southern Sweden. *BMC Women's Health,* Volume 14:63.

Gage, A. J., 2005. Women's experience of intimate partner violence in Haiti. *Social Science & Medicine,* 61(2), pp. 343-364.

Garcia-Moreno, C. et al., 2006. WHO multi-country study on women's health and domestic violence against women study team: prevalence of intimate partner violence: findings from the WHO multi country study on women's health and domestic violence. *Lancet*, Volume 368(9543), pp. 1260-1269.

Garcia-Moreno, C. & Pallitto, C., 2013. *Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence*, Geneva: WHO.

Gazmararian, J. et al., 1996. Prevalence of violence against pregnant women. *Journal of American Medical Association,* 275(19), pp. 15-20.

Graham, J. W., 2009. Missing data analysis: making it work in the real world.. *Annual Review Psychology,* Volume 60, pp. 549-576.

Greenland, S., Arminda, L. S. & Waleska, T. C., 1994. alternative models for ordinal logistic regression. *Statistics in medicine,* 13(16), pp. 1665-1677.

Green, P. J. & Silverman, B. W., 1993. Nonparametric regression and generalized linear models. *London: CRC press.*

Green, P. J. & Silverman, B. W., 1993. Nonparametric regression and generalized linear models. *London: CRC press.*

Green, P. J. & Silverman, B. W., 1993. Nonparametric regression and generalized linear models. *London: CRC Press.*

Habyarimana, F., 2016. Measuring poverty and child malnutrition with their determinants from household survey data. *University of Kwazulu-Natal.*

Habyarimana, F., 2016. *Measuring poverty and child malnutrition with their determinants from household survey data,* Pietermaritzburg: University of Kwazulu-Natal.

Habyarimana, F., Zewotir, T. & Ramroop, S., 2014. A potential odds model with complex sampling design to identify key determinants of malnutrition of childrenunder five years in Rwanda. *Mediterranean Journal of Social Sciences,* 5(23), pp. 1642-1648.

Habyarimana, F., Zewotir, T. & Ramroop, S., 2014. A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under five years in rwanda. *Mediterranean Journal of Social Sciences,* 5(23), pp. 1642-1648.

Habyarimana, F., Zewotir, T. & Ramroop, S., 2018. Determinants of Domestic Violence in Women of Reproductive Age in Rwanda. *Journal of Economics and Behavioral Studies,* 10 (1)(2220-6140), pp. 101-111.

Haitovsky, Y., 1968. Missing Data in Regression Analysis. *Journal of the Statistical Society. Series B (Methodological),* 30(1), pp. 67-82.

Hardle, W., 1999. Applied nonparametric regression. *Cambridge university Press.*

Hastie , T. J. & Tibshirani, R. J., 1990. Generalized additive models. *New York: CRC press.*

Hastie, T. J. & Tibshirani, R. J., 1990. Generalized additive models. *CRC Press.*

Heeringa, S. G., West, B. T. & Berglund, P. A., 2010. Applied survey data analysis.

Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X., 2013. *Applied logistic regression,* s.l.: Wiley.

Hosmer, D. W. & Lemeshow, S., 2004. Applied logistic regression. *John Wiley & Sons.*

Ibrahim, J. G., Chen , M.-H., Lipsitz, S. R. & Herring, A. H., 2011. Missing-Data Methods for Generalized Linear Models. *Journal of the American Statistical Association,* 100(469), pp. 332-346.

Jasinki , J., 2004. Pregnancy and domestic violence: a review of the literature. *Trauma Violence Abuse,* Volume 5(1), pp. 47-64.

Jewkes, R., Levin, J. & Penn-Kekana, L., 2002. Risk factors for domestic violence: findings from a South African cross-sectiopnal study. *Socialo Science & Medicine,* Volume 55, pp. 1603-1617.

Jewkes, R. et al., 2001. Prevalence of emotional, physical, and sexual abuse of women in three South African provinces. *South Aftrican Medical Journal,* 91(42), pp. 1-8.

Joanne, Z. M., 2018. *Africana Collection: List of Sub-Saharan African Countries.* [Online] Available at: https://www.loc.gov/rr/amed/guide/afr-countrylist.html

Kachman, S. D., 2000. An introduction to generalized linear mixed models, in `Proceedings of a symposium at the organizational meeting for a NCR coordinating committee on Implementation Strategies for National Beef Cattle Evaluation. *Athens',* pp. 59-73.

Kim, K. & Cho, Y., 1992. Epidemiological survey of spousal abuse in Korea, Intimate Violence: Interdisciplinary Perspectives.. *Washington: Hemisphere Publishing corporatiion.*

Kincaid, C., 2005. Guidlines for selecting the covariance structure in mixed model analysis.. *In Proceedings of the Thirtieth Annual SAS Users Group International Conference, number 198-30..*

Koeing, M. A. et al., 2003. Domestic violence in rural Uganda: evidence from a community-based study. *Bulletin of the World Health Orginization,* Volume 81 (1).

Koenig, M. A. et al., 2003. Domestic violence in rural Uganda: evidence from a community-based study. *Bulletin of the World Health Organization,* Volume 81(1), pp. 53-60.

Kohn, R., Ansley, C. F. & Tharm, D., 1991. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of applied statist. Ass.,* 86(416), pp. 1042-1050.

Krug, E. G., Mercy, J. A., Dahlberg, L. L. & Zwi, A. B., 2002. *The world report on violence and health,* Geneva: World Health Organization.

Lee, K. J. & Carlin, J. B., 2010. Multivariate Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. *American Journal of Epidemiology,* 171(5).

Lemeshow, S. & Hosmer, D., 2000. Applied logistic regression. *Wiley Series in Probability and Statistics.*

Lindsey, J., 1997. Applying generalized linear models. *Springer Verlag.*

Lin, X. & Zhang, D., 1999. Inference in generalized additive mixed models. *Journal of the royal statistical society,* 61(2), pp. 381-400.

Littell, R. C., Pendergast, J. & Natarajan, R., 2000. Tutorial in biostatistics: modelling covariance structure in the analysis of repeated measures data. *Statistics in medicine,* 19(1793), p. 1819.

Little, R. J. & Rubin, D. B., 2002. *Statistical Analysis With Missing Data.* 2nd ed ed. New York: John Wiley & Sons Inc.

Maman, S., Campbell, J., Sweat, M. & Gielen, A., 2000. The intersections of HIV and violence: directions for future research and interventions. *Social Science and Medicine*, Volume 50:4, pp. 59-78.

Manning, C., 2007. Generalized linear mixed models (illustrated with R on bresman et al. datives data). *Unpublished handout.*

Martin, S. et al., 1999. Sexual behaviors and reproductive health outcomes: associations with wife abuse in India. *Journal of the American Medical Association,* 282(19), pp. 67-72.

McCullagh, P. & Nelder, J. A., 1989. Generalized linear model. *Chapman & Hall/CRC.*

McCulloch, C. E. & Neuhaus, J. M., 2011. Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics,* 67(1), pp. 270-279.

McCulloch, C. E., Searle, S. R. & John, M., 2008. Generalized Linear and Mixed Models.

Mculloch, C. E., Searles, S. R. & John, M., 2008. Generalized Linear and Mixed Models. *A John Wiley & Sons,Inc.*

Moeti, A., 2010. *Factors affecting the health status of the people of Lesotho,* Pietermaritzburg: University of Kwazulu-Natal.

MoHCDGEC, M. o. H. C. D. G. E. a. C. et al., 2016. *Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16,* Dar es Salaam, Tanzania, and Rockville, Maryland, USA: MoHCDGEC, MoH, NBS, OCGS, and ICF: Ministry of Health.

Molenberghs, G. & Verbeke, G., 2005. Models for discrete longitudinal data. *New York: Springer-Verlag.*

Molenberghs, G. & Verbeke, G., 2006. Models for discrete longitudinal data. *Springer Science & Business Media.*

Morrell, R., Jewkes, R. & Lindegger, G., 2012. Hegemonic Masculinity/Masculinities in South Africa: Culture, power, and gender politics. *SAGE J,* Volume 15, pp. 11-30.

Muluneh, M. D., Stulz, V., Francis, L. & Agho, K., 2020. Gender Based Violence against Women in Sub-Saharan Africa: A Systematic Review and Meta-Analysis of Cross-Sectional Studies. *International Journal of Environmental Research and Public Health,* Volume 17, pp. 903-921.

NBS, N. B. o. S. & OCGS, O. o. t. C. G. S., 2019. *Microdata Library.* [Online] Available at: https://microdata.worldbank.org/index.php/catalog/2739 [Accessed 2021].

NDoH, N. D. o. H., StatsSA, S. S. A., SAMRC, S. A. M. R. C. & ICF, January 2019. *South Africa Demographic and Health Survey 2016*, Pretoria, South Africa, and Rockville, Maryland: National Department of Health.

Obi, S. N. & Ozumba, B. C., 2007. Factors associated with domestic violence in south-east Nigeria. *Journal of Obstetrics and Gynaecology,* Volume 27:1, pp. 75-78.

Palermo, T., Bleck, J. & Peterman, A., 2014. Tip of the iceberg:Responding and gender-based violence in developing countries. *A Journal in Epidemiology,* Volume 179, pp. 602-612.

Pallitto, C. C. et al., 2006. WHO Multi-country Study on Women's Health and Domestic Violence against Women Study Team. Prevalence of intimate partner violence: Findings from the WHO multi-country study on women's health and domestic violence.. *Lancet,* Volume 368, pp. 1260-1269.

Park, I. & Lee, H., 2001. The design effect: Do we know all about it. *In Proceedings of the Annual Meeting of the American Statistical Association,* August.pp. 5-9.

Raghunathan, T. E., Lepkowski, J. M. & Van Hoewyk, J., 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol,* 27(1), pp. 85-95.

Raghunathan, T. W., Lepkowksi, J. M., Van Hoewyk, J. & Solenbeger, P., 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology,* Volume 27, pp. 85-95.

Rao, V., 1997. Wife-beating in rural South India: a qualitative and econometric analysis. *Social Science and Medicine,* Volume 44, pp. 1169-1180.

Robinson, G. K., 1991. That blup is a good thing: the estimation of random effects. *Statistical Science,* 6(1), pp. 15-32.

Royston, P., 2004. Multiple imputation of missing values. *Stata J,* 4(3), pp. 227-241.

Rubin, D., 1987. Multiple Imputation for Nonresponse in Surveys.

Rubin, D. B., 1976. Inference and Missing Data. *Biometrika,* Volume 63, pp. 581-592.

SA), S. S. A. (., 2012. *2011 Statistical Release,* Pretoria: s.n.

Schafer, J. I. & Graham, J. W., 2002. Missing data: our view of state of the art. *Psychological Methods,* Volume 7, pp. 147-177.

Schafer, J. L., 1997. Analysis of Incomplete Multivariate Data.

Schall, R., 1991. Estimation in generalized linear models with random effects. *Biometrika,* 78(4), pp. 719-727.

Scheffer, J., 2002. Dealing with Missing Data. *Res.Lett.Inf.Math.Sci,* Volume 3, pp. 153-160.

Schei , B. & Bakketeig, L., 1989. Gynaecological impact of sexual and physical abuse by spouse. A study of a random sample of Norwegian women. *British Journal of Obstetrics and Gynaecology,* Volume 96:13, pp. 79-83.

Schneider, M. et al., 2018. Patterns of intimate partner violence among perinatal women with depression symptoms in Khayelitsha, South Africa: A longitudinal analysis. *Global Mental Health,* Volume 5, p. 11.

Seitzman, R. L., Mahajan, V. B. & Mangione, C., 2008. Estrogen receptor alpha and matrix metalloproteinase 2 polymorphisms and age-related maculopathy in older women. *Am J Epidemiol,* 167(10), pp. 1217-1225.

Shackman, G., 2001. Sample size and design effect. *Albany Chapter of the American Statistical Association.*

Simundi´c, A.-M., 2008. Measures of diagnostic accuracy: basic definitions. *Med Biol Sci,* pp. 61-65.

StatsSA, S. S. A., 2019. *Microdata Library.* [Online] Available at: https://microdata.worldbank.org/index.php/catalog/3408 [Accessed 2021].

Swahnberg, I. & Wijma, B., 2003. The NorVold Abuse Questionnaire (NorAQ): validation of new measures of emotional, physical, and sexual abuse, and abuse in the health care system among women.. *Eur J Public Health,* Volume 13(4), pp. 361-366.

Tierny, L. & Kadane, J. B., 1986. Accute approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393), pp. 82-86.

UBOS, B. o. S., 2019. *Microdata Library.* [Online] Available at: https://microdata.worldbank.org/index.php/catalog/2979

Uganda Bureau of Statistics (UBOS) and ICF, January 2018. *Uganda Demographic and Health Survey 2016,* Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF: UBOS and ICF.

United Nations, 1993. *Declaration on the Elimination of Violence against Women,* s.l.: s.n.

Usta, J., Farver, J. A. M. & Pashayan, N., 2007. Domestic violence: The Lebanese experience. *Journal of the royal institute of public health,* Volume 121, pp. 208-219.

Valdez-Santiago, R. & Sanin-Aguirre, L., 1996. Domestic violence during pregnancy and its relationship with birth weight. *Salud Publica Mexicana,* Volume 38:3, pp. 52-62.

Van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, Volume 16, pp. 219-242.

van Buuren, S., Boshuizen, H. C. & Knook, D. L., 1999. A multiple-imputation of missing blood pressure covariates in survival analysis. *Stat Med*, 18(6), pp. 681-694.

van Buuren, S., Boshuizen, H. C. & Knook, D. L., 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine,* Volume 18, pp. 681-694.

van der Straten, A. et al., 1998. Sexual coercion,physical violence, and HIV infection among women in steady relationships in Kigali,Rwanda. *AIDS and Behavior,* Volume 2, pp. 61-73.

Verbeke, G. & Molenberghs, G., 2000. Linear mixed models for longitudinal data. *New York: Springer-Verlag.*

Vittinghoff, E., Glidden, D. V., Shiboski, S. C. & McCulloch, C. E., 2011. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. *Springer Science & Business Media.*

VPA, 2002. *VIOLENCE PREVENTION ALLIANCE.* [Online] Available at: https://www.who.int/violenceprevention/approach/definition/en/#:~:text=Interpersonal%20violence%20refers%20to%20violence,partner%20violence%20and%20community%20violence.&text=Collective%20violence%20refers%20to%20violence,social%2C%20political%20and%20eco

Waagepetersen, R., 2007. Computation of the likelihood function for GLMMs.

Watts, C., Keough, E., Ndlovu, M. & Kwaramba, R., 1998. Withholding of sex and forced sex: dimensions of violence against Zimbabwean women. *Reproductive Health Matters,* Volume 6, pp. 57-65.

Watts, C., Keough, E., Ndlovu, M. & Kwaramba, R., 1998. Withholding of sex and forced sex: dimensions of violence against Zimbabwean women. *Reproductive Health Matters,* Volume 6, pp. 57-65.

WHO, 1996. *Violence: a public health priority,* Geneva: World Health Organization.

WHO, 2002. *World report on violence and health,* Geneva: World Health Organization.

WHO, 2010. *Injuries and violence: the facts,* Geneva: World Health Organization.

WHO, 2013. *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence,* Geneva: World Health Organization.

WHO, 2020. *Health topics.* [Online] Available at: https://www.who.int/health-topics/violence-against-women#tab=tab_1

Worldometer.info, 2021. *Worldometer.* [Online] Available at: https://www.worldometers.info/population/countries-in-africa-by-population/ [Accessed 22 March 2021].

Yang, Y., 2011. Multiple Imputation Using SAS Software. *Journal of Statistical Software,* 45(6).

Zeger, S. L. & Karim, M. R., 1991. Generalized linear models with random effects; a gibs sampling approach. *Journal of the American statistical association,* 86(413), pp. 79-86.

Zhang, D., Lin, X., Raz, J. & Sowers, M., 1998. Semi-parametric stochastic mixed models for longitudinal data. *Journal of american statistical association,* 93(442), pp. 710-719.

Appendices

Appendix A

Different Codes Used

The variables used to fit the models are described below with full names.

D113: Partner alcohol drinking status, D121: Woman's father ever beat her mother, ATM: Access to media, V012: Age of respondent, V024: Region, V025: Type of residence, V106: Highest education level (Woman), V136: Number of household members, V151: Sex of household head, V155: Literacy, V190: Wealth index , V228: Ever had a terminated pregnancy, V312: Use of contraceptive methods, V445: Body mass index (Woman), V501: Current marital status, V505: Number of other partners, V513: Cohabitation duration, V621: Partner's desire for children, V701: Partner's education level, V705: Partner's occupation, V717: Woman's occupation, V730: Partner's age, V746: Woman earnings compared to her partner, V751: Ever heard of sexually transmitted infections (STIs) and V739: Person who usually decide on what to do with woman's earnings

A.1 Missing Values (Multiple Imputation by Chained Equations)

```
proc mi data=data name nimpute=1 out=data name seed=19944 minimum=0;
class d113 d121 v705 v746 v621 v701 ATM IPV v119 v850a v312;
fcs     nbiter=5     logistic(d113/details)     logistic(d121/details)
logistic(v119/details) logistic(v621/details)
logistic(v701/details) logistic(v705/details) logistic(v746/details)
logistic(v850a/details) regpmm(v445/details)
regpmm(v505/details) regpmm(v730/details);
var D113 D121 V705 V746 ATM IPV v001 v012 v013 v021 v023 v024 V025
V106 V119 V136 V151 V155 V190 V228 V312
V445 V501 V505 V513 V621 V701 V717 V730 v750 v850a;
run;
```

A.2 Logistic Regression SAS Code

```
ods graphics on;
proc logistic data=data name plots=effect plots=roc;
class  d113(ref='1')  d121(ref='1')  atm(ref='1')  v024(ref='2')  v025
v106(ref='0')    v119    v151    v155(ref='1')    v190(ref='5')    v228
v312(ref='1') v501(ref='0') v513(ref='1') v621(ref='1') v701(ref='0')
v705 v717(ref='0') v746(ref='2') v750 v850a(ref='1')/param=glm;
model ipv(event='1') =d113 d121 atm v012 v024 v025 v106 v119 v136 v151
v155 v190 v228 v312 v445 v501 v505 v513 v621 v701v705 v717 v730 v746
v750  v012*v312  v190*v312/link=logit  selection=stepwise  lackfit  rsq
expb;run;
ods html close;
      ods graphics off;
```

A.3 Survey Logistic Regression SAS Code

```
      ods graphics on;
proc surveylogistic data=data name;
ods output oddsratios=domainors;
stratum v023;
cluster v001;
class d113(ref='1') d121(ref='1') atm(ref='1') v024(ref='2') v025
v106(ref='0') v119 v151 v155(ref='1') v190(ref='5') v228
v312(ref='1') v501(ref='0') v513(ref='1') v621(ref='1') v701(ref='0')
v705 v717(ref='0') v746(ref='2') v750 v850a(ref='1')/param=glm;
model ipv(event='1') =d113 d121 atm v012 v024 v136 v151 v190 v228 v312
v445 v501 v513 v621 v705 v717 v730 v746 v750 v850a
v012*v312 v190*v312 /link=logit expb;run;
ods graphics off;
```

A.4 Generalized Linear Mixed Model SAS Code

```
proc glimmix data=data name method=laplace;
class d113(ref='1') d121(ref='1') atm(ref='1') v024(ref='2') v025
v106(ref='0') v151 v155(ref='1') v190(ref='5') v228 v312(ref='1')
v501(ref='0') v513(ref='1') v621(ref='1') v701(ref='0') v705
v717(ref='0') v746(ref='2') v750 v850a(ref='1');
model ipv(event='1') =d113 d121 atm v012 v024 v025 v106 v136 v151 v155
v190 v228 v312 v445 v501 v505 v513 v621 v701 v705 v717 v730 v746 v750
v850a v012*v312 v190*v312/dist=binary link=logit oddsratio s;
random intercept/ subject=v001;run;
```

A.5 Generalized Additive Mixed Model R Code

```
Name of data=gamm(IPV~s(V012)+factor(V136)+factor(V445)+s(V505)+factor(V730)+

factor(D113)+factor(D121)+factor(ATM)+factor(V024)+factor(V025)+

factor(V106)+factor(V151)+factor(V155)+factor(V190)+factor(V228)+

factor(V312)+factor(V501)+factor(V513)+factor(V621)+factor(V701)+

factor(V705)+factor(V717)+factor(V746)+factor(V750)+factor(V850A),

random=list(V024=~1),family=binomial(link = logit),data=Name of imported data)

summary(Name of data$gam)
```

A.6 Multicollinearity

```
proc reg data=data_name;
model _Y=d113 d121 v012 v024 v025 v106 v136 v151 v155 v190 v228 v312
v445 v501 v505 v513 v621 v701 v705 v717 v730 v739 v746 v750/vif tol
collin; run;
```