

ESSAYS ON HEALTH AND HEALTH CARE UTILIZATION IN BOTSWANA

David Mmopelwa

School of Economics

THESIS
Submitted to the University of Nottingham for the Degree of Doctor of Philosophy

December 2020

ABSTRACT

The role of health in both economic growth and development has been established in the literature. Consequently, public health policies aim to ensure an equitable provision of quality curative health services and promotion of preventive care. However, some national survey-based health indicators in Botswana show that both child and adult health deteriorated over time, which warrants investigation that is now timely given that both the current health policy and the integrated health service plan are nearing their end. Therefore, this three-essay thesis attempts to provide insights into deteriorating health.

Following an increase in the proportion of low birth weight (i.e., less than 2.5kg) children, during a period in which the country was among the worst performers in the WHO African region and the upper-middle-income category, the first essay assesses the role of prenatal care utilization on infant`s health. A joint treatment effect model for care utilization and infant outcome is estimated, accounting for the endogeneity of care utilization. As not all infants have birth weight reported, the essay also estimates the sample selection model to control for potential sample selection bias. Results show that adequate prenatal care utilization is positively associated with infant health and that failing to account for endogeneity reduces its effect. On average, birth weight improves by 0.67, 0.73 and 0.64 kg in full, urban and rural samples, respectively. These findings imply the need to audit care quality since utilization rates are high, but outcomes have been deteriorating.

The second essay investigates the effect of household size and child`s birth order on child nutrition as measured by the three anthropometric indicators: height for age, weight for age and weight for height. The essay is motivated by one of the theoretical predictions of the resource dilution hypothesis relating to the impact of family size on child capital. The hypothesis postulates that large family sizes and high child`s birth order are more likely to have a negative effect on child capital. However, there are arguments that the assumption of a fixed and narrow flow of resources from parents underpinning the theory may not always hold. We estimate the random effects model to explore the within and between household effect. We find that household size is negatively associated with child nutrition and high birth order children (i.e. later born) are more likely to fare worse than their lower birth order counterparts. A variation decomposition analysis shows a higher variation in child nutrition across than within households. Of the total variation, a large share is accounted for by the unexplained component.

For all the three indicators the between household explained shares are in the range 5-8%, while for within household explained shares are less than 1%.

The third essay investigates whether sexual behaviour (indicators of activity and use of a condom) is responsive to both objective risk (HIV prevalence rate) and perceived risk (concern on the likelihood of being infected by HIV due to ART). Using three of the four waves of the Botswana AIDS Impact Survey (BAIS) collected by Statistics Botswana, we account for the endogeneity of both risk indicators through the method of instrumental variables. HIV prevalence rate is instrumented by the distance from respondent`s area of residence to the district with the highest HIV rate, while the perceived risk is instrumented with duration (in months) between antiretroviral therapy (ART) rollout and the survey date. Overall, results suggest there is more responsiveness to subjective rather than objective risk indicators. Compared to those who indicate to have not changed their level of concern about being infected by HIV due to ART, the less (more) concerned are less likely to abstain and use a condom (more likely to use condoms). The HIV prevalence rate has no effect on both the number of sexual partners and use of condoms. From a policy perspective, the non-behavioural response to HIV prevalence could be a cause for concern if there are unobserved fatal behavioural patterns. Thus, because programmes such as ART prolongs the lives of the infected and affects perception about the likelihood of infection risk, there could be some negative externalities as revealed by the results.

ACKNOWLEDGEMENTS

The production of this thesis was under the guidance and supervision of Professor Oliver Morrissey and Professor Trudy Owens, to whom I am profoundly indebted. I am forever grateful for their patience, support and advice throughout the process and offer warm appreciation to them. I would also like to thank my sponsor and employer, the Botswana Institute for Development Policy Analysis (BIDPA), for making sure that my welfare was well-taken care of throughout the study period. Special thanks to the national statistics office, Statistics Botswana, for their generosity in availing survey data used in exploring the issues in this thesis and to Dr Serai Daniel Rakgoasi of the University of Botswana for some additional data for the third essay. To Dr Nkomo and Ms Pono in the Ministry of Health and Wellness, thanks for the HIV/AIDS programmes` information provided. I would also like to extend special thanks to the members of staff of the School of Economics, University of Nottingham, who provided a facilitative learning environment. Lastly, I am grateful to my family and friends, who supported me all the way.

Table of Contents

CHAPTER 1: INTRODUCTION 1

1.1. Botswana`s health sector and some selected health indicators 4

1.1.1 Fertility Indicator (Total Fertility Rate) 6

1.1.2 Morbidity Indicators 7

1.1.3 Mortality Indicators 9

1.2. Data 11

1.3. Structure, organization and contribution 11

CHAPTER 2: PRENATAL CARE UTILIZATION AND INFANT HEALTH IN BOTSWANA 15

2.1 Introduction 16

2.2 Literature Review 21

2.3 Data and Summary Statistics 28

2.4 Model Specification 33

2.4.1 Theoretical Model 33

2.4.2 Empirical Model 35

2.5 Results and Discussions 40

2.5.1 Testing for Regressor Endogeneity 44

2.5.2 Relevance and Validity of Instruments 44

2.6 Robustness Checks 45

2.6.1 Quality of Care 46

2.6.2 Sample Selection 46

2.7 Conclusion 47

APPENDIX FOR CHAPTER 2 49

CHAPTER 3: HOUSEHOLD SIZE, BIRTH ORDER AND CHILD HEALTH IN BOTSWANA 53

3.1 Introduction 54

3.2 Review of related literature 58

3.2.1 Household Size and Child Health in Developed Countries 58

3.2.2 Household Size and Child Health in Developing Countries 59

3.2.3 Birth Order and Child Health in Developed Countries 60

3.2.4 Birth Order and Child Health in Developing Countries 61

3.3 Data and Summary Statistics 64

3.4 Model Specification 71

3.4.1 Theoretical Model 71

3.4.2 Empirical Model 72

3.5 Results and Discussion 77

3.6 Robustness Checks 85

3.7 Conclusion 86

APPENDIX FOR CHAPTER 3 ..... **88**
CHAPTER 4: HIV RISK AND SEXUAL BEHAVIOUR IN BOTSWANA **98**
4.1 Introduction **99**
4.2 Related Literature **102**
4.2.1 Concept and Measurement of HIV Risk 102
4.2.2 Evidence on HIV Risk and Sexual behaviour 103
4.3 Data and Descriptive Statistics **109**
4.4 Model Specification **115**
4.4.1 Theoretical 115
4.4.2 Empirical Strategy 116
4.5 Results and Discussions **122**
4.5.1 Objective risk and sexual behaviour 122
4.5.2 Perceived risk and sexual behaviour 129
4.6 Robustness checks **135**
4.6.1 Is there evidence of sample selection? 135
4.6.2 Subsample Effect 135
4.7 Discussion and Relation to Literature **138**
4.8 Conclusion **139**
APPENDIX FOR CHAPTER 4 ..... **142**
CHAPTER 5: CONCLUSIONS **168**
5.1 Summary of findings **168**
5.2 Implications **169**
5.3 Further Research **171**
BIBLIOGRAPHY **173**

LIST OF TABLES

Table 1.1. IMR and UMR (per 1,000 live births):1971-2017 11
Table 2.1: Low Birth Weight and Care Coverage:2000-2009 17
Table 2.2: Variable Definitions 31
Table 2.3: Summary Statistics 33
Table 2.4: First Stage Regression Results 41
Table 2.5: OLS and Second Stage Regression Results 42
Table 3.1: Definitions of variables 66
Table 3.2: Summary Statistics 70
Table 3.3: Nutritional Indicators Summary Statistics by Birth order 70
Table 3.4: Summary Statistics for Child with Mothers in Household 71
Table 3.5: Demographic Indicators: 1971-2011 75
Table 3.6: Regression Results for Child Nutritional Indicators 79
Table 3.7: Probit Marginal Effects for Child Malnutrition 80
Table 3.8: Regression results for Child Nutritional Indicators (omitting household size) 82
Table 3.9: Regression results for Children residing with their biological mothers 83
Table 3.10: Probit Marginal effects for children residing with their biological mothers 84
Table 4.1: Description of Variables 112
Table 4.2: Descriptive Statistics 113
Table 4.3: Mean differences by Sexual activity report 115
Table 4.4: Cross tabulation of knowledge and education variables 122
Table 4.5: Marginal Effects (at means) for indicators of sexual partners (Ordered Probit) 123
Table 4.6: Probit Marginal Effects for condom use 125
Table 4.7: First stage results for objective risk (HIV Prevalence) 127
Table 4.8: Second Stage Results for Condom Use 128
Table 4.9: Marginal Effects (at means) from Ordered Probit (Perceived Risk) 130
Table 4.10: Probit Marginal Effects for Perceived Risk on the use of a condom 131
Table 4.11: First Stage results for the change in the level of concern about HIV infection 133
Table 4.12: IV (2SLS) Regression results for condom use 134

LIST OF FIGURES

Figure 1.1:Trends in PMTCT Programme Enrolment:2002-2010 6
Figure 1.2: Trends in Total Fertility Rates: 1971-2017 7
Figure 1.3: HIV/AIDS prevalence rate among pregnant women aged 15-49:1992-2011 7
Figure 1.4: Tuberculosis Notification Rate (per 100,000) in Botswana:1975-2018 8
Figure 1.5: Malaria Cases in Botswana:1987-2018 9
Figure 1.6: Maternal Mortality Ratio (per 100,000 live births) in Botswana:2005-2018 10
Figure 3.1: Trends in Malnutrition Rates: 1996-2009 54
Figure 3.2: Household Tree diagram 69

CHAPTER 1: INTRODUCTION

Botswana`s government, through the country`s health policy, is committed to improving health through equitable provision of quality curative and preventive health services (Republic of Botswana, 2011). During the millennium development goals era, the country joined others by committing to (i) reduce child mortality, (ii) improve maternal health and (iii) combat HIV/AIDS, malaria and other diseases (Republic of Botswana and United Nations, 2015). Attaining these objectives was to be facilitated by the free provision of public health care and the social protection system, which had existed since the country's independence (RHVP, 2011). Specifically, the government introduced the Vulnerable Group Feeding Programme through which all children aged 6-60 months receive a monthly food ration when they attend a health care facility which monitors children's growth (BIDPA, 2010). As was envisaged through the country`s Vision 2016, an antiretroviral therapy programme was introduced to reduce HIV/AIDS spread and related mortality; and there were efforts to promote less risky behaviour through national campaigns (NACA, 2015). Not only have these objectives been maintained in the post-2015 development agenda (WHO, 2015a), but in Botswana`s Vision 2036 (Presidential Task Force, 2016).

However, the above intentions notwithstanding, some national survey-based health indicators suggest that both child and adult health deteriorated over time. Some might use these statistics to question the effectiveness of the care provision from a public policy perspective. For instance, the proportion of children born below a normal birth weight (2.5kg) increased from 8% to 13% during the period 2000-2008, placing the country among the worst performers in the World Health Organization (WHO) African region and upper-middle-income countries (WHO, 2011a). Similarly, from 2000 the proportion of under-five children with a low height for age (stunting), low weight for age (underweight) and inadequate weight for height (wasting) increased (World Bank, 2015). Moreover, the rates recorded in 2009/10 suggests a reversal of the achievement made during the 1990s. That achievement (a decline in malnutrition) however, was observed to have mainly been driven by changes in moderate malnutrition, while severe malnutrition remained almost constant (Seleka et al., 2008).

The Botswana AIDS Impact surveys show a steady increase in HIV prevalence rates of 17.1%, 17.6% and 18.5% in 2004, 2008 and 2013 respectively (NACA and CSO, 2005; CSO and NACA, 2009; Statistics Botswana and NACA, 2016). While increasing prevalence may have arguably been due to the reduced mortality as a result of the antiretroviral therapy programme

(Bonita et al., 2006; MFDP, 2017), survey results reveal some behavioural patterns (such as unsafe sexual intercourse while on treatment) that warrants further investigation. The increasing HIV rate prevailed with the increased use of contraceptives over time (Statistics Botswana, 2018) raising the question if the policy of reducing fertility resulted in neglecting measures to reduce HIV/AIDS infection?

Given an established link between child and adult health as well as health and the economy (see Strauss and Thomas, 2008 for a review), a pattern of deteriorating health performance warrants investigation, which is now timely given that both the current health policy and the integrated health service plan for Botswana are nearing their end. The reason for the focus in Botswana is that it was excluded from existing cross-country studies given absence of comparable survey data in Botswana. For instance, unlike in other countries, Botswana separately conducts the family health survey, AIDS Impact surveys and the demographic surveys whereas other countries have demographic and health surveys. Resultantly, it is common for those studies that pool sub–Saharan African country surveys to exclude Botswana due to the data non-comparability. Examples include Bingenheimer (2010), Oster (2012), Tennekoon and Rosenman (2014), Fox (2014), Howell et al. (2016), Lucas and Wilson (2018).

The three essays in this thesis provide insights on three specific issues:

1. Does prenatal care utilization have any effect on infant birth weight in Botswana?
2. What are the roles of birth order and household size on child nutritional indicators?
3. What is the impact of objective and perceived HIV risk on sexual behaviour?

The aim of answering these questions is to provide information that could be useful in guiding future public policy to address issues that threaten the policy objective of better health for all.

The relationship between health status and health care utilization adopted here is influenced by Grossman (1972), who considers health care not to be demanded for its own sake, but for the production of health (i.e. demand for health care is derived from the demand for health). Thus, when people seek health care, the aim is to either prevent or recover from ill-health. While health has mostly been viewed as the outcome of formal health care systems (Culyer, 1971), Grossman recognizes the role of individuals and households in health production (Wagstaff, 1986). The model treats health both as consumption and an investment commodity. From the

consumption viewpoint, health directly affects utility (e.g. for mothers with healthy children) while as an investment commodity it affects the amount of time available for market and non-market activities (like in the case of an ill individual not being able to go to work). This framework lays the basis for the three essays explored in this thesis. Thus, treating health as a commodity to be produced allows us to explore individual investment decisions such as utilizing prenatal care, distributing household resources and using/not using protection in sexual intercourse, investigated in the first, second and third essay respectively. In emphasizing the role of non-health care factors in health production, Ludbrook and Cohen (2003) state that during the 18th and 19th century improved health outcomes were attributable to factors outside the formal health sector such as improved sanitation. Furthermore, according to Wojtczak (2005:87) "it is estimated that 10-25% of improvement in health status can be attributed to medical intervention". This implies that a larger share is due to factors outside formal care.

A question might be asked, why research on health and health care utilization? It is important to note that "health is a complete state of physical, mental and social wellbeing, not merely the absence of disease or infirmity" (WHO, 2006:1). This definition implies that health has several indicators as we show for Botswana in the next subsection. It has been recognized that health affects individuals` availability for work and their productivity (Mushkin, 1962), which subsequently affects economic growth and development (Bloom and Canning, 2001). As mentioned at the outset, evidence suggests an association between infant/childhood health and future adult health and human welfare (Alexander et al., 2014; Gajate-Garrido, 2013; Smith, 2009; Victora et al., 2008; Black et al., 2007), as well as between adult health and the economy (Strauss and Thomas, 1998; 2008; Finlay, 2007). Hence public policy has centred around improving health status.

However, although health might depreciate exogenously through aging (see Kramer, 1987 in case of reproductive health), as producers of health, individual behaviour may have elements that are not always observed, necessitating a need for empirical investigation into factors that affect health and health care utilization. This is further complicated by the nature of health care, the characteristics of which can be assessed on three factors: consumer rationality, uncertainty and externalities (Culyer, 1971). For instance, in the case of communicable diseases such as HIV/AIDS there is the possibility that an individual may not be aware of their infection status, or even when aware, may engage in a behaviour which some might label "irrational', yet maximizes their expected utility (Phillipson and Posner, 1993). Moreover, assumptions, such

as people care about others, as a rationale for public health policies (Hauck, 2018) have been questioned by empirical evidence (Gong, 2015; Baird et al., 2014). There is also evidence that the impact of some public health measures aimed at improving health may have unintended consequences. For example, those infected by HIV may increase the number of partners as a result of improved health due to antiretroviral therapy (Lakdawalla et al., 2006). These illustrate the complex nature of issues related to health and care utilization. These are the issues investigated empirically with respect to Botswana in this thesis.

1.1. Botswana`s health sector and some selected health indicators

The delivery of health care services in Botswana is through public, private (profit and non-profit), and traditional medicine practice (Republic of Botswana, 2011). The sector is largely public, led by the Ministry of Health and Wellness through health care facilities such as hospitals, clinics, health posts[1], mobile stops, as well as community-based preventative and promotive services. The Ministry is also responsible for the formulation of policies, regulatory tools and standards for health services. Delivery of health care is in five levels: the individual, primary health care clinic, primary hospital, district hospital and referral hospital. Including the individual in this recognises that they play a role in determining their health through their behaviour (Republic of Botswana, 2011). As of 2016, the country had three national Referral Hospitals, 18 District Hospitals, 17 Primary Hospitals and 318 primary health care clinics (Statistics Botswana, 2017).

As health outcomes are also affected by non-formal health sector factors, a variety of indicators (fertility, morbidity and mortality) at the population level are reported to illustrate features of the health sector in Botswana. Part of the reason for the selected indicators is that they represent some of the factors recognized as potential challenges in the implementation of the country`s integrated health service plan (Republic of Botswana, 2010a). Some of the targets of the selected indicators outlined in the revised national population policy are as follows: (MFDP and UNFPA, 2010:26-27):

1. Achieve full coverage of pregnant women attending antenatal care with at least four follow up visits.

1 Health posts are “facilities with no beds, staffed by a nurse with visits by a midwife, mental health nurse, eye nurse and doctor.” A mobile stop, on the other hand, “has no permanent structure, visited by a nurse, health education assistant and lay counsellor.”(MOHW, 2017:online)

2. Monitor the pace of fertility decline and introduce appropriate measures to maintain total fertility rate at three children per woman.
3. Reduce HIV national prevalence from 17.1% to 10% in 2016 and further to 2% by 2025.
4. Reduce TB incidence from 602 per 100, 000 persons in 2005 to 202 per 100, 000 persons by 2015 and eliminate malaria in the country by 2015.
5. Reduce maternal mortality to less than 50 deaths per 100, 000 live births.
6. Reduce infant mortality from 46.5 in 2007 to less than 23 per 1000 live births in 2020 and under-five mortality from 68 in 2007 to less than 29 in 2020.

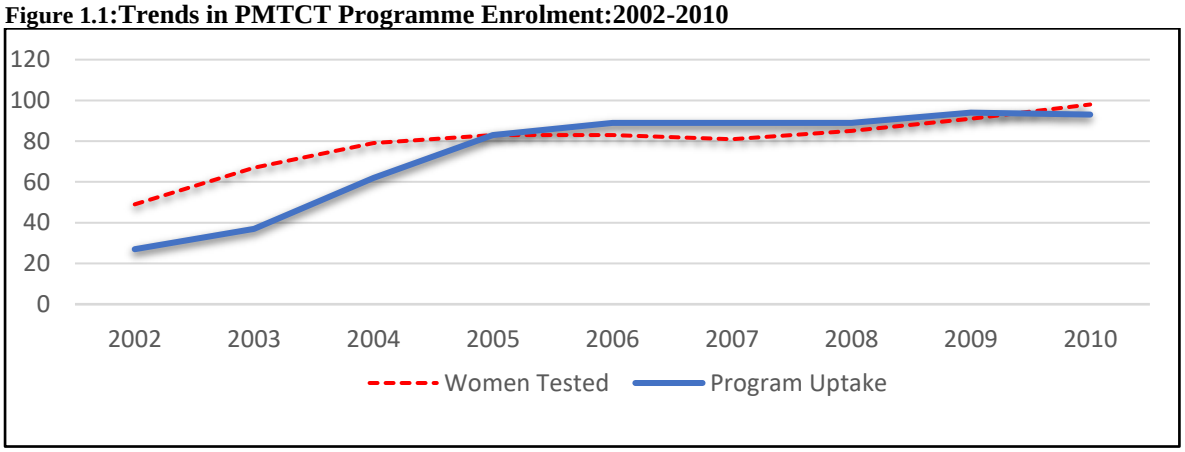
Historical efforts to provide care to pregnant women in Botswana date back to the early 1980s during which it was provided at a locality (village) level. Specifically, this was through the Bamalete Lutheran Hospital in Ramotswa village, located in the South-East administrative district of the country (Fako et al., 2004). An extension of coverage resulted in the launch of the Botswana Obstetric Record (BOR) in 1984, as a primary tool for monitoring pregnancies. This collected information on the mother`s history of utilization of prenatal care, labour and delivery. Following the 1987 safe motherhood conference recommendation, Botswana adopted the safe motherhood initiative in 1990 (Mogobe et al., 2007). The initiative put much emphasis on the skills required for the provision of obstetric care, with the conference specifically recommending to halve maternal mortality by the year 2000. According to Fako et al. (2004), BOR was reviewed in 1997 with the intention to incorporate the ability of health professionals in identifying pregnancy risk factors, most notably HIV/AIDS.

Consequently, in 1999 the government introduced and incorporated into the prenatal care services the Prevention of Mother to Child Transmission (PMTCT) programme whose components were: (i) prevention of pregnancies among young girls, (ii) prevention of unwanted pregnancies among HIV positive mothers, (iii) Antiretroviral (ARV) prophylaxis for prevention of HIV infection from mother to child, and (iv) support for mothers and family (BIDPA, 2010). Before the programme guidelines were revised in 2011, an optional HIV testing was available for pregnant women and the subsequent assistance offered based on the test results (MOHW, nd). Thus, depending on the CD4 count, a woman could be enrolled on either the Highly Active Antiretroviral Therapy (HAART) or AZT (Zidovudine) (BIDPA, 2010). Those with $\leq$ 250 cells/μL were eligible for HAART, and those with more than 250 cells/μL were eligible for Zidovudine (Dryden-Peterson et al., 2011). In their study involving

428 infants and 439 HIV-infected mothers who delivered live infants at the two hospitals (Scottish Livingstone and Princess Marina) in Botswana, Dryden-Peterson et al. (2011) found that HAART reduced the mother to child transmission rates more than the AZT. The authors tracked infants` health from birth to 6 months of age and found that 2.5% of infants got infected in the Zidovudine group compared to the 0.4% in the HAART group.

Currently, all HIV positive pregnant women are eligible for ART irrespective of their CD4 count (MOHW, nd). Monitoring is continued even after the birth of the child. Official statistics reveal that over time, both the trend in testing and programme uptake increased (Figure 1.1). The proportion of women testing rose from 49% to 98% between 2002 and 2010. Similarly, programme uptake increased from 27% to 93% during the same period.

Figure 1.1:Trends in PMTCT Programme Enrolment:2002-2010

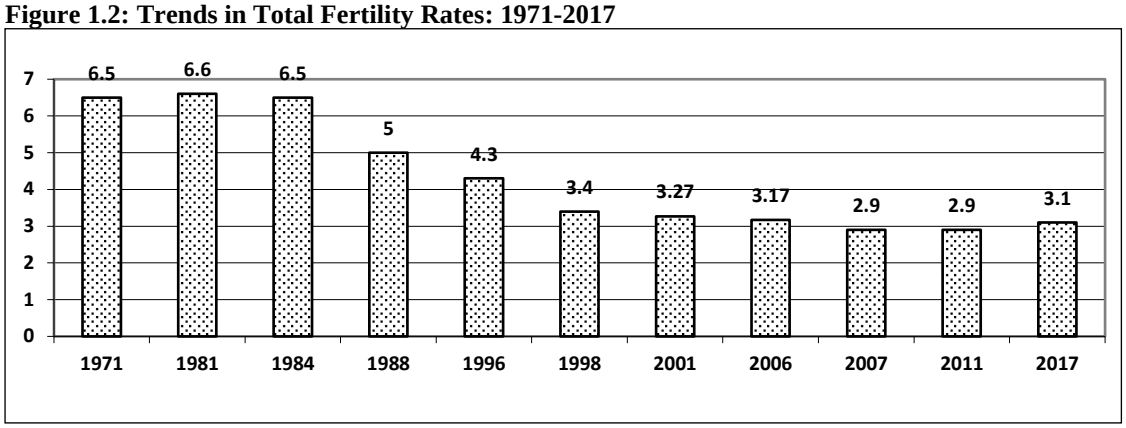


Source: BIDPA (2010: 21)

1.1.1 Fertility Indicator (Total Fertility Rate)

Total Fertility Rate (TFR) is calculated as the mean total number of births for women of reproductive age (15-49 years) (Moultrie et al., 2013). Figure 1.2 shows that Total Fertility Rates declined from 6.5 births in 1971 to 3.1 births by 2017. While declining TFR could be attributed to several factors, the driving force has been credited to the Botswana National Family Planning Programme, increased age at first birth, increased female education and women`s participation in the labour force (Mills et al., 2010). Although statistics suggest that this indicator is on track in relation to the revised population policy target, future trends might be affected by HIV prevalence. As one study has established, for women on antiretroviral therapy, perceived stigmatization from both family members and health care workers and a belief that it is irresponsible to have a child when HIV positive are negatively associated with planning to have a child (Schaan et al., 2014).

Figure 1.2: Trends in Total Fertility Rates: 1971-2017



Source: CSO (1989; 2009b); Majelantle (2014); Statistics Botswana (2018)

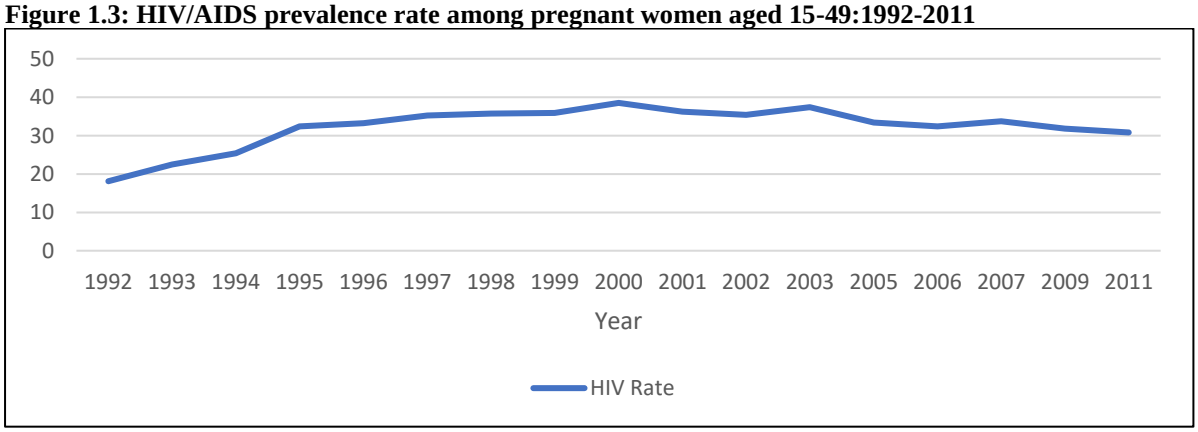
1.1.2 Morbidity Indicators

This subsection discusses the magnitude of diseases at population level focusing on HIV/AIDS, Tuberculosis and Malaria. Addressing these diseases remains a policy target both at national and international level (MFDP and UNFPA, 2010; WHO, 2015a).

HIV/AIDS prevalence

The country`s first recorded case of HIV was in 1985, after which the government responded through various means including information, education and communications programmes (UNDP, 2000). As part of monitoring the disease, the first surveillance was conducted in 1992, although this focused on pregnant women who utilized antenatal care. Figure 1.3 shows the trend in the prevalence rate, which consistently increased until reaching a peak of 38.5% in 2000. Although it declined slightly after that, it remained above 35% until 2003. From 2005 to 2011, the prevalence rate was in the range of 30-34%.

Figure 1.3: HIV/AIDS prevalence rate among pregnant women aged 15-49:1992-2011



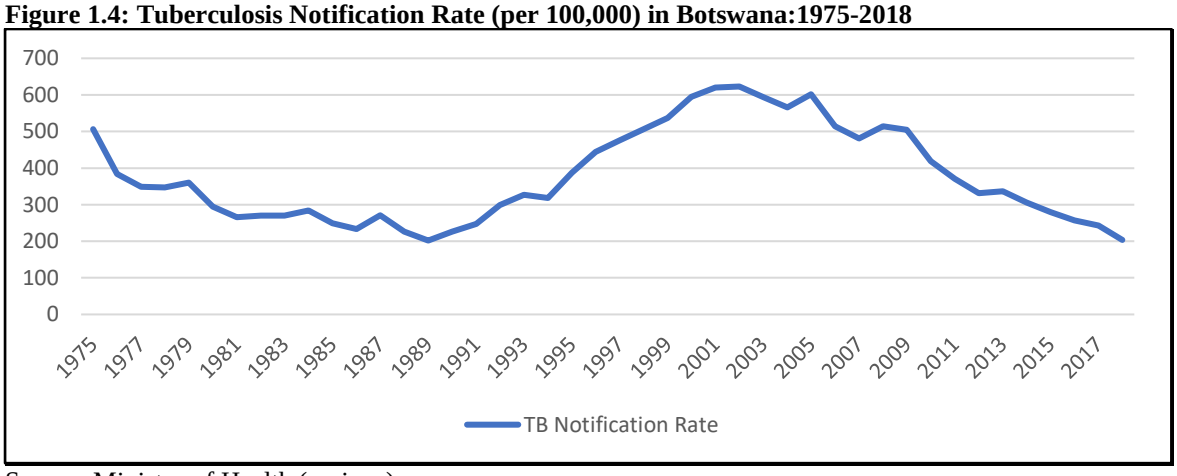
Source: NACA (2003); MOH (2011a)
Notes: Data for 2004, 2008 and 2010 not available

The national AIDS impact surveys indicate that HIV prevalence increased from 17.1% in 2004 to 18.6% in 2013 (NACA and CSO, 2005; Statistics Botswana, 2016), implying that the policy target to reduce HIV national prevalence from 17.1% to 10% in 2016 and further to 2% by 2025 is yet to be realised. Results further show that prevalence is generally higher among females, the less educated and those residing in rural areas.

Tuberculosis

Related to HIV are the Tuberculosis notification rates per 100,000 people presented in Figure 1.4. As evident, rates declined in the 1970s until 1990-2002 during which the opposite was witnessed due to HIV and multi-drug resistance (MoH, 2011b). While generally, notification rates declined, an observation was made that some strategic plan targets were not met. For instance, during the period 2006-2008 treatment success rates were less than 60% against the target of 71% (MoH, 2010) and stood at 79%, 81% and 82% during 2009-2011 against the 85% and 87% Botswana's and global target respectively (MoH, 2012).

Figure 1.4: Tuberculosis Notification Rate (per 100,000) in Botswana:1975-2018



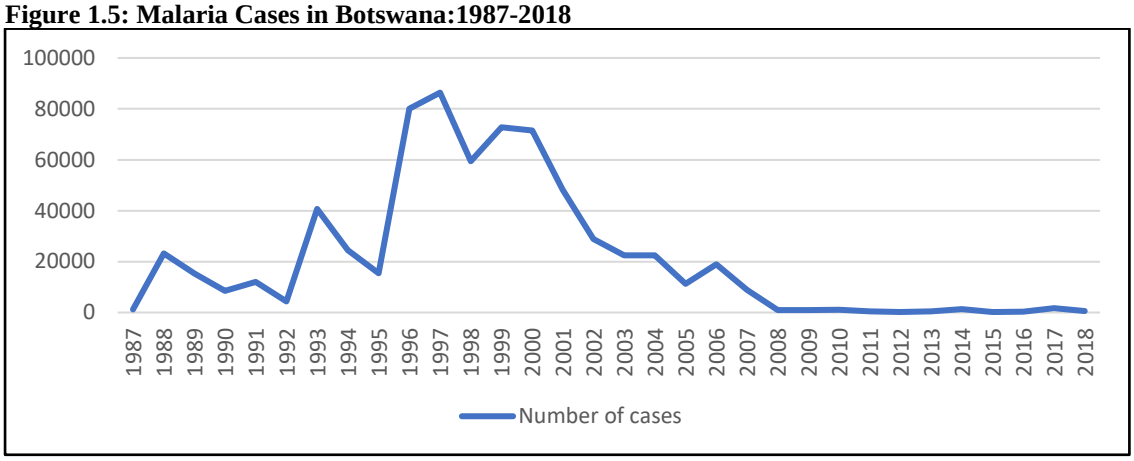
Source: Ministry of Health (various)

Malaria

Figure 1.5 presents the number of malaria cases for Botswana and shows that from the late 1980s cases generally declined, a trend attributed to the drought that the country experienced (UCSF, 2016). However, following an increase (decrease) in the number of cases in 1993 (1994-5), a sharp increase was witnessed in 1996/97. UCSF (2016) notes that this increase resulted in 141 deaths. From about 71,000 cases recorded in 2000, a consistent improvement was recorded until the last year of the series, partly reflecting the outcome of a policy shift from control to elimination of malaria. Specifically, the target was to have the disease eliminated by the year 2018 (UCSF, 2016). However, the objective to eliminate malaria by

2015 (Republic of Botswana and UNFPA, 2010) and by 2018 was not met. Nevertheless, Figure 1.5 suggests a promising pattern towards that objective. The number of cases for the period from 2008 to 2018 ranges between 188 (in 2009) and 1713 (in 2017).

Figure 1.5: Malaria Cases in Botswana:1987-2018



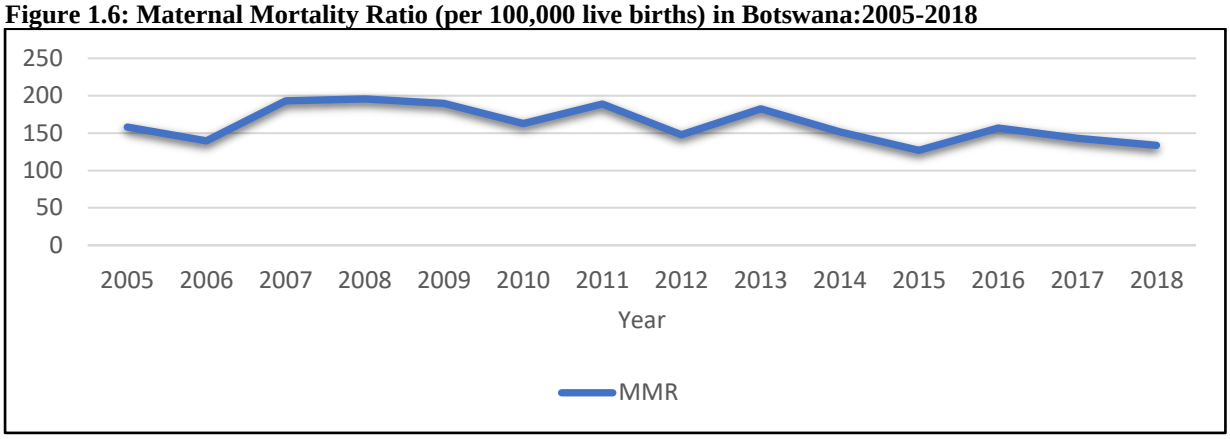
Source: WHO (2015b;2019); Mosweunyane (2007)

1.1.3 Mortality Indicators

Maternal Mortality

In addition to ensuring an improvement in infant health, prenatal care services aim to prevent maternal deaths (WHO, 2016). Maternal death is "the death of a woman while pregnant or within 42 days of termination of pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes" (WHO, 2004:98). One indicator used to monitor maternal deaths in Botswana is Maternal Mortality Ratio (MMR), which is the ratio of maternal deaths per 100, 000 live births (Moultrie et al., 2013). Figure 1.6 presents the trend in Botswana`s MMR for the period from 2005 to 2018. While the ratio recorded in 2005 represented an improvement from the 1990`s of 326 (Republic of Botswana and United Nations, 2015), Figure 1.6 shows that prior to 2015, except for the years 2006 and 2013 the ratio was consistently above 150 deaths per 100,000 live births. One implication is derived from the trend: the country`s target of 150 deaths per 100,000 live births by 2011 and global target of 82 per 100,000 live births in 2015 (Republic of Botswana and United Nations, 2015) were not met in Botswana.

Figure 1.6: Maternal Mortality Ratio (per 100,000 live births) in Botswana:2005-2018



Source: Statistics Botswana (various)

Infant Mortality Rates and Under-Five Mortality Rates

Infant mortality rate (IMR) measures the rate of death for children within the first year of life, while under-five mortality rate (UMR) is that for those under-five years (Bonita et al., 2006:26). Table 1.1 shows that the two indicators declined during the period from 1971 to 1996, followed by an increase until 2001 for infant mortality rate and 2007 for under-five mortality rates. The increasing trends for both indicators may partly be attributed to the prevalence of HIV/AIDS (MFDP and UNFPA, 2010). However, a significant fall was witnessed during the period from 2007 to 2011, where IMR decreased from 57 to 17 and UMR decreased from 76 to 28 deaths per 1000 live births; a pattern attributable to a change in estimation technique (Majelantle, 2014). Both measures can be estimated both directly and indirectly. The indirect technique estimates mortality rates using aggregate information on the number of children ever born, those still alive and those dead from women categorised by age group (Moultrie et al., 2013). It assumes that child mortality risks for women who reported their childbearing experience and those who did not are the same and that fertility remained constant during the previous census or surveys (United Nations, 1990).

However, if fertility declined, as was the case with Botswana (Figure 1.2), the technique results in an over-estimation in child mortality. It was observed that with the 2011 population and housing census data, the use of the indirect technique produced biased estimates (Majelantle, 2014). The direct estimation technique disaggregates the child`s information (i.e. names, dates of birth and death for each child, etc.) for a particular age cohort. From such information, mortality rates for specified age cohort`s births are estimated by dividing the numbers of births that died by the number of births. The two indicators have since risen in 2017, with rates above

the targets for 2020 (23 per 1000 live births for IMR and 29 per 1000 live births for UMR) (MFDP and UNFPA, 2010).

Table 1.1. IMR and UMR (per 1,000 live births):1971-2017

Year	IMR	UMR
1971	97	152
1981	71	105
1984	70	-
1988	37	53.4
1991	48	63
1996	37	45
1998	51	67
2000	57	75
2001	56	74
2006	48	76
2007	57	76
2011	17	28
2017	38	48

Source: (BIDPA, 2010), CSO (1989; 2009b); Majelantle (2014); Statistics Botswana (2018)
Notes: IMR: Infant Mortality Rate, UMR: Under-five Mortality Rate

1.2. Data

This thesis uses data from national surveys conducted by Statistics Botswana. The first essay uses the Botswana Family Health Survey (BFHS) for 2007/08, which includes under-five child and women characteristics. Attempt to reduce the impact of recall bias evident of household surveys (Ansah and Powell-Jackson, 2013) and the nature of prenatal care measure considered for this essay resulted in the sample of 1395 under-five children. The second essay uses the Botswana Core Welfare Indicator Survey (BCWIS) for 2009/10, which was an improvement on the previous income and expenditure surveys. It includes non-income welfare measures, with health status captured by child height and weight measurements for a sample of 3873 children. The third essay uses the last three waves of the four Botswana AIDS Impact Surveys (the first survey lacks relevant data) for 2004, 2008 and 2013 for a pooled sample of 31948 individuals. All these surveys are conducted following the same two-stage sampling methodology. The first stage involves the selection of enumeration areas, while the second involves the selection of households using the probability to proportional to size of the area, where the number of households is used as a measure of size. Moreover, they are all representative at the national level.

1.3. Structure, organization and contribution

The rest of the thesis is organised in the following manner. Chapter two (first essay) investigates the role of prenatal care in infant health. It estimates a treatment effect equation to

account for the endogeneity in care utilization. It also checks for sample selection bias since not all children have reported birth weight (83% have). The third chapter (second essay) considers birth order and household size effect on child nutritional indicators. It estimates the random effect model to explore the within and between household effect, and a household fixed effect for robustness. Chapter four (third essay) assesses responsiveness of sexual behaviour to both objective risk (HIV prevalence rate) and perceived risk (concern about the likelihood of being infected by HIV/AIDS as a result of the antiretroviral therapy). Endogeneity of both risks is accounted for through instrumental variables. Conclusion with policy implications and recommendations for further research are presented in chapter five.

The thesis contributes to the existing inconclusive literature on prenatal care and infant health, household size and birth order`s effect on child nutrition, as well as sexual behaviour and HIV/AIDS. The effect of prenatal care on infant health, explored in the first essay, is difficult to identify given endogeneity emanating from preterm delivery, pregnancy resolution, treatment of the distribution of health outcomes as well as the use of care itself (Rosenzweig and Schultz, 1982; 1983; Grossman and Joyce, 1990; Warner, 1995; Fiscella, 1995; Frick and Lantz, 1996; Warner, 1998; Conway and Deb, 2005; Evans and Lien, 2005; Jewell, 2007; Abrevaya and Dahl, 2008; Mwabu, 2008; Habibov and Fan, 2011; Gajate-Garrido, 2013; Habibov et al., 2017). Consequently, evidence for the relevance of prenatal care is limited; according to Kramer (1987:716), "no firm conclusions can be drawn, about the possible benefits of frequent antenatal visits on either intrauterine growth or gestation duration". Fiscella (1995:468) concludes that "prenatal care has not demonstrated to improve birth outcomes conclusively". Moreover, Corman et al. (2019:1) note that "an increasingly methodologically rigorous literature suggests that the effect of prenatal care timing and quantity on birth outcomes, particularly low birth weight are modest at the population level". Our results suggest that not accounting for endogeneity, which has not been addressed in previous empirical work in Botswana, could underestimate the effect of prenatal care utilization on infant health.

Regarding the second essay, theory predicts high childbirth order and larger family size to have a negative impact on child nutrition (Becker and Lewis, 1974; Becker and Tomes, 1976). This prediction is based on resource dilution hypothesis, according to which low birth order children should be healthier as they do not experience competition for resources, while a larger family size implies depletion of resources (De Keyser and Rossem, 2017; Horton, 1988; Blake, 1981). However, it has been argued and shown that the assumption underpinning this prediction is

that resources remain fixed, and their flow is narrow in a nuclear family set up neglecting the effect of buffering, either at household or national level (Bras et al., 2010; Lu and Treiman, 2008; Desai, 1995; Blake, 1981). This also relates to anthropological literature which argues that a definition of a household as recommended by the United Nations misses both the social and economic interaction of members residing elsewhere (Bongaarts, 2001), which characterises an African household (see Randall and Coast, 2015).

Furthermore, existing evidence in Botswana has not controlled for (or explored) the possible unobserved heterogeneity within households that might explain deteriorating child nutrition. Therefore, in contributing to these debates, we adopt a household definition allowing for resources to be from outside the nuclear family and explore both the within and between household variation in child nutrition. Our results show that both household size and high birth order are negatively associated with child nutrition. There is higher variation in child nutrition between than within households. As we decompose the variation into explained and unexplained, we show that a large proportion of this variation is unexplained by the considered factors. This implies the impact of household size on child nutrition could be due to intrahousehold factors not observed due to data limitation.

Lastly, the relationship between sexual behaviour and HIV/AIDS risk is endogenous, leading to inconclusive empirical evidence (Guillon and Thuilliez, 2015; Oster, 2012; Tenkorang et al., 2011). The third essay examines the impact of objective and perceived risk (concern about the likelihood of being infected with HIV/AIDS due to the existence of ART) on sexual behaviour. In the US unintended consequences of receiving treatment for HIV have been evidenced (an increase in the number of partners) (Lakdawalla et al., 2006) although Auld (2006) finds that an increasing prevalence leads to a larger decrease in risky behaviour among low-risk individuals than among risky ones. Studies for Africa provide mixed results for the effect of learning of an HIV test result on risky behaviour (Gong, 2015; Thornton, 2008). In Botswana, although Keetile & Letamo (2015) find that believing that ARV cures AIDS is positively associated with the increased likelihood of inconsistent condom use, Letamo et al. (2017) find that the same belief reduces inconsistent use. Moreover, previous empirical work that investigated influential factors for sexual behaviour only considers respondents who report to being sexually active (Keetile and Kgosidintsi, 2018; Balekang and Dintwa, 2016; Keetile, 2014; Malema, 2012), potentially suffering from endogeneity (Wooldridge, 2010). Our results

show that individuals respond more to perceived risk than objective risk, and that an approach to focus only on those sexually active may lead to biased results.

CHAPTER 2:PRENATAL CARE UTILIZATION AND INFANT HEALTH IN BOTSWANA

Abstract

Despite the freely and publicly provided health care in Botswana, the proportion of low birth weight infants increased from 8% to 13% during 2000-2007/08 period. The latter rate placed the country among the worst performers in the World Health Organization African region and upper-middle-income countries and may question the effectiveness of care. Using the 2007/08 Botswana Family Health Survey data collected by Statistics Botswana, the paper jointly estimates the adequate prenatal care utilization (input demand) and infant birth weight (outcome) functions through the treatment effect model. As birth weight information is not available for all infants, we also estimate a Heckman sample selection model to account for potential bias. Estimating models for rural and urban samples separately, we find that lower level of mother`s education reduces the likelihood to both adequately utilize prenatal care and report infant birth weight. The likelihood for prenatal care utilization increases with the probability of a care facility being sufficiently close and availability of care facilities. Adequate prenatal care utilization is positively associated with birth weight, and failing to account for endogeneity reduces its effect. On average, birth weight increases by 0.67, 0.73 and 0.64 kg in full, urban and rural samples, respectively.

2.1 Introduction

The Botswana government`s developmental efforts have, among others, been demonstrated through its public budget allocation to the health sector. During the period 1980-2013 for instance, public health expenditure share to total public spending increased from 4.6% to 10.7% (MFDP, various), real health spending grew by 9 per cent per annum, and the share to GDP almost doubled. Such commitments are in part, based on the recognition of health as an important dimension of welfare (World Bank, 2001), and a major input for economic growth and development (Bloom and Canning, 2001). One crucial component in the production of both infant and mother's health is prenatal care, which in Botswana, has been provided since the 1980s (Fako et al., 2004). Prenatal care benefits both the infant and expectant mother through among others, identification of those who are at risk of poor infant health outcome as well as the promotion of proper nutrition and cessation of risky behaviour during pregnancy (WHO, 2016). In this chapter we investigate the role of prenatal care on birth weight in Botswana, through a health production function which identifies households and individuals as critical players, through "their health affecting consumption patterns, utilization of health care and their environment" (Wagstaff, 1986:3). We ask a question, does the utilization of prenatal care have any effect on infant birth weight in Botswana? Answering this question for Botswana is essential because of the following.

Despite the publicly provided prenatal care in Botswana, which (should and) may arguably have contributed to the observed high utilization rates (see Table 2.1), the proportion of low birth weight (i.e. below 2.5 kg) infants increased from 8% in 2000 to 13% in 2007/08 (CSO, 2009a), placing the country among the worst performers in the World Health Organization African region and upper-middle-income countries during the said period (WHO, 2011a). Since prenatal care utilization is recognized as an essential input in reducing the likelihood of poor-health risks (WHO, 2016), a pattern of increasing rates of low birth weight infants, given high care utilization rates might lead some to question its effectiveness. However, this is not to suggest that care has not been effective in Botswana, but we argue that such a pattern warrants investigation, carried out in this chapter.

While Table 2.1 may be said to signal a possible paradox for Botswana compared with other countries (such as Angola, Cameroon, DRC, Kenya, Lesotho, Nigeria, Swaziland[now Eswatini], Togo, Zambia, and Zimbabwe), high rates of prenatal care utilization could relate to mothers` health. Thus, either low-risk mothers utilize care more because they recognize its

important role or higher risk mothers utilize it more because of their poor health (Rosenzweig and Schultz 1982; Joyce, 1994). None of these possibilities is directly observed in the data and is acknowledged as a source of bias affecting the effectiveness of care (Frick and Lantz, 1996). There is, however, a consensus that prenatal care utilization should positively correlate with both mothers` and infants` health (WHO, 2016).

Table 2.1: Low Birth Weight and Care Coverage:2000-2009

Country/Region	Low Birth Weight (%)	Prenatal Care Coverage (%)
Angola	12	68
Botswana	13	94
Cameroon	11	82
The Central African Republic	13	69
Democratic Republic of Congo	10	87
Eritrea	14	70
Ethiopia	20	28
Gabon	14	94
The Gambia	20	98
Ghana	13	90
Kenya	8	92
Lesotho	13	92
Liberia	14	79
Madagascar	16	86
Malawi	14	92
Mauritius	14	-
Mozambique	15	92
Namibia	16	95
Nigeria	12	58
Senegal	19	87
Sierra Leone	14	87
South Africa	-	92
Swaziland	9	85
Togo	12	84
Uganda	14	94
Tanzania	10	96
Zambia	11	94
Zimbabwe	11	93
WHO Region		
African	13	74
Americas	8	95
South-East Asia	24	76
European	7	97
Eastern Mediterranean	21	68
Western Pacific	5	91

Source: Awiti (2014:2); WHO (2011a).
*Note*s: Within each WHO region, countries are sorted by the latest available data since 2000 for coverage of at least one visit. Regional averages are also for coverage of at least one visit.

Birth weight is one important indicator facilitating the examination of effective interventions towards better health outcomes. There is suggestive evidence that poor infant`s health is associated with poor future adult health and human capital development (Victora et al., 2008; Gajate-Garrido, 2013; Alexander et al., 2014). Among others, economic benefits of addressing low birth weight include reduced (i) infant mortality, (ii) costs of neonatal medical attention, (iii) costs of subsequent illness and related medical care for infants and children, and (iv) discounted lifetime productivity gain, and reduction in intergenerational impacts (Alderman and Behrman, 2006). Resultantly, public health policy has centred around improving health status.

Although there is generally an agreement that low birth weight could result from either retarded growth or preterm birth (short gestation period) or both (Kramer, 1987; Conway and Deb, 2005; Gajate-Garrido, 2013), there exists debate and inconsistencies on the role of prenatal care utilization on birth weight in the literature. Such is due to bias from care utilization, preterm delivery, pregnancy resolution as well as treatment of the distribution of health outcomes and measure of care itself (Rosenzweig and Schultz 1982, 1983; Grossman and Joyce, 1990; Warner, 1995; Fiscella, 1995; Frick and Lantz, 1996; Warner, 1998; Conway and Deb, 2005; Evans and Lien, 2005; Jewell, 2007; Abrevaya and Dahl, 2008; Mwabu, 2008; Reichman et al., 2009; Habibov and Fan, 2011; Gajate-Garrido, 2013; Habibov et al., 2017). Consequently, evidence for the relevance of prenatal care is limited; according to Kramer (1987:716), "no firm conclusions can be drawn, about the possible benefits of frequent antenatal visits on either intrauterine growth or gestation duration". Fiscella (1995:468) concludes that "prenatal care has not demonstrated to improve birth outcomes conclusively". Similarly, in a review of studies for the period 1983-2018, it emerges that the effect of prenatal care on health outcomes is modest (Corman et al., 2019). They observe that "an increasingly methodologically rigorous literature suggests that the effect of prenatal care timing and quantity on birth outcomes, particularly low birth weight are modest at the population level" (p.1).

While it is acknowledged that random trials for the impact of prenatal care on health outcomes are limited by ethical considerations (Grossman and Joyce, 1990; Liu, 1998), evidence from those conducted among low-risk women in developed countries could be used by some to question the effectiveness of prenatal care. In London, Sikorski et al. (1996) find that while women randomly assigned to fewer (compared to standardized) visits have less content (fewer

scans and caregivers), their health outcomes (caesarean section rate, morbidity indicators, infant`s birth weight) are not statistically different from those who received standard care. Moreover, although no significant difference is evident in the reported smoking and alcohol consumption behaviour during pregnancy, women with less care expressed more worry about babies` welfare even postnatally. However, a follow-up study by Clement et al. (1999), though with a reduced sample (40% of the original) due to attrition, suggests that the post-natal psychological effects of reduced visits are temporary. In the US, McDuffie et al. (1996) also find no significant difference in maternal behaviour, morbidity indicators and child`s birth weight between those with less and those with standardized care.

A similar pattern for the impact of reduced prenatal care visits on birth weight is observed in some developing countries such as Argentina, Cuba, Saudi Arabia and Thailand (Villar et al., 2001) and Zimbabwe (Majoko et al., 2007; Munjanja et al., 1996). Together with the review by Carroli et al. (2001) this evidence led to the development of the new model (WHO, 2002)[2]. However, a pooled analysis of the above studies by Dowswell et al. (2010) suggests a higher likelihood of perinatal mortality among the reduced visits in developing countries. According to Vogel et al. (2013), the latter results need careful interpretation given the possibilities of heterogeneity (variations in the definitions of reduced and standard visits and outcomes being considered across the studies). Nonetheless, they find similar results after adjusting for risk factors. Based on these and other factors, a new model of care utilization (of 8 contacts), with arguments that the previous one was not convincingly effective, was recommended by the World Health Organization (WHO, 2016). Notwithstanding this, some still demonstrate that reduced visit model could still yield the same results if coupled with better monitoring (Tobah et al., 2019). This demonstrates the debate in the literature, which is likely to continue even in the subsequent evidence.

Notwithstanding the above, existing evidence in Botswana (Letamo and Majelantle, 2001 and Okurut et al., 2013) consider care as an exogenous variable, a shortcoming which we address. Thus, the chapter contributes to the debate on the role of prenatal care on birth weight and highlights potential areas for policy intervention. We consider the self-reported level of use of

[2] Evaluations on the impact of the model in some SSA countries reveal that this model led to an improved quality and continuity of care in Ghana (Birungi et al., 2006), and improvement in early access to care in Malawi (Mchenga et al., 2019). However, no change in quality is evident in Malawi (Mchenga et al., 2019) and South Africa (Chege, 2005) and underutilization of care also witnessed in Malawi.

prenatal care, since its impact does not only depend on the occurrence but intensity as well (Gajate-Garrido, 2013). However, some data limitations are worth noting. Firstly, we are not able to control for the activities performed during care utilization. Secondly, contrary to the practice with demographic and health surveys in other countries, information on the level of care utilization was, during a survey, collected as a categorical variable (i.e. 1-3 & 4 or more visits). Hence, we only have a dummy for adequate care utilization as per the Botswana`s revised population policy objective (MFDP and UNFPA, 2010) and the previous World Health Organization`s recommendation (WHO, 2002; 2016). Although this represents loss of data to some extent, it could be argued to reduce the measurement errors associated with recall bias as demonstrated with other indicators elsewhere with other health indicators (Ansah and Powell-Jackson, 2013; Lawson and Appleton, 2007).

Thirdly, we use the 2007/08 BFHS, which is quite old. While there is the 2009/10 Botswana Core Welfare Indicator Survey and the 2015/16 Botswana Multi Topic Household survey, these surveys only asked whether women used prenatal care or not, which was considered by one previous study (Okurut et al., 2013), though it did not account for endogeneity of care use. As indicated, while we also use binary indicator for level of use due to data constraint, we adopt a different measure which captures intensity of use. This is because with the former measure, care use is more than 90%, and it conceals the fact that there is considerable difference on the level of prenatal care utilization as evident in only the 2007/08 BFHS. It is also important to note that the survey does not have information on the child`s father, who may impact birth weight directly through biological factors or indirectly by affecting care utilization or contributing to household resources (Lindo ,2011).

The main finding is that adequate prenatal care utilization is positively associated with birth weight and failure to account for its endogeneity could underestimate its effect on infant health. The finding is important for Botswana`s health policy, whose duration nears the end, as well as for the revised national population policy. Thus, efforts should go beyond access to care, but promote adequate use of care among those that do not fully use it as it can improve infant health outcomes. Such efforts should be coupled with the need to reflect on care quality. It is also important considering one previous study, which did not account for this endogeneity and found that 4-5 visits increased the likelihood of low birth weight in Botswana (Letamo and Majelantle, 2001). The chapter proceeds as follows. Section 2.2 provides an overview of related literature, while section 2. 3 discuses data issues. Modelling and estimation issues are presented

in section 2.4 and results in section 2.5. Section 2.6 discusses checks for robustness while 2.7 concludes.

2.2 Literature Review

The chapter relates broadly to the literature on factors influencing birth weight, and especially to the role of prenatal care utilization. The correlates of birth weight include socio-economic, demographic, biological, and behavioural factors (Kramer, 1987), evidence for which exists in both developed and developing countries, although with varying conclusions. For instance, using the 1968-97 waves of the Panel Study of Income Dynamics data, Lindo (2011) finds that husband`s job displacement (a measure of income shock) negatively affect child`s birth weight. On the other hand, Dehejia and Lleras-Muney (2004) find that the incidence of children born with low weight is reduced on infants conceived during the periods of high unemployment rate in the US. Contrary to the theoretical prediction that hard economic conditions may reduce consumption of health-related services like prenatal care, their results show that care use improves for both black and white mothers. Moreover, they also demonstrate that fertility selection decisions play a role: the incidence of birth reduces among less-educated black mothers improving child health, while it increases among the less educated white mothers.

Other insights are provided in the form of stressful economic news (Carlson, 2015), which affect birth weight through mothers` health (stress) or their coping risky behaviour. A positive association between mother`s education and infant birth weight has been found in Lithuania (Dickute et al., 2004), India (Joshi et al., 2005) and Pakistan (Khan et al., 2014). In Lesotho, Mathule et al. (2005) find that women more exposed to the hungry season (more pronounced food insecurity) during pregnancy are more likely to have low mean birth weight infants. Behavioural factors include smoking (see Silles, 2012 for a review), whose negative effect on birth weight is witnessed in among others, Alberta (Tough et al., 1999), Switzerland (Chiolero et al., 2005) and Nagpur (Deshmukh et al., 1998). In the following, we provide an overview of studies that consider the effect of prenatal care on birth weight, in both developed and developing countries.

Evidence in developed countries

A vital recognition by Rosenzweig and Schultz (1982, 1983) was that prior evidence was threatened by potential sample selection, inadequate control for endogenous health-related behaviour, and failure to account for heterogeneity in unobserved characteristics that may

affect infant health. For their analysis of the national US natality follow-back surveys for 1967 to 1969, Rosenzweig and Schultz (1982, 1983) account for endogeneity through a 2SLS estimator. Their results indicate that delay in seeking prenatal care reduces birth weight, unhealthy women started prenatal care earlier and that their babies had lower birth weight, confirming evidence of adverse selection. This pattern of the effect remains even with a consideration of an additional 1980 wave (Rosenzweig and Schultz, 1988). Notably, their approach became influential to subsequent research work on the effect of care on birth outcomes as we show in this section.

Another form of potential bias emanating from the decision to have a child was incorporated into the analysis by Grossman and Joyce (1990). These authors argue that previous empirical evidence had neglected the possible correlation between unobservables affecting a decision to give birth (as opposed to abortion) and those affecting prenatal care use decisions, an issue which they maintain is inadequately addressed by instrumental variables method. In their analysis, they provide three estimates: probit model for birth resolution, Heckman 2 step model to correct for selectivity, 2SLS estimator controlling for both birth selectivity and endogeneity in prenatal care using the 1984 data for pregnant women in New York. However, they find evidence of selectivity bias among black women only: a positive correlation between unobserved factors that increase the likelihood to give birth and unobservables that reduce delay in care initiation, which increases infant birth weight. They show that potential mean birth weight for women who resolved to abort is lower than for those who did not.

In subsequent work, Joyce (1994) investigates the effect of levels of prenatal care on birth weight. Using data for black, white and Hispanic women aged at least 20 years in 1984, the author modifies the measure of prenatal care by combining both care initiation and visits. The argument is that this measure avoids possible misclassification associated with either visits or delay when considered separately, such as the possibility of a woman initiating care early and not completing the recommended visits (pp.768). The effect is investigated both among all women residents and those who chose a certain level of care to demonstrate the effect of switching the levels of care. Specifically, the expected mean birth weight of women with adequate care is compared to their expected mean birth weight with either intermediate or inadequate care. Results show that among all women, the effect of moving from inadequate to an intermediate level of care improves birth weight more than moving from intermediate to

adequate level. Similarly, mothers with adequate care would have had lower birth weight infants if they had an inadequate level of care.

Using data for the 1980-1990 black women`s births, Warner (1995) investigates the impact of prenatal care delay and number of visits adjusted for duration of pregnancy on birth weight. Through a 2SLS method, his results show that contrary to expectation, although women whose access to care was facilitated by Medicaid initiate care earlier, Medicaid was negatively associated with visits. According to the author, this is attributed to the observation that obstetricians were less willing to assist Medicaid beneficiaries due to the low rates of reimbursement. While delay is negatively associated with birth weight, visits have no effect. In his modelling approach, he considers the interaction of these two measures of prenatal care to assess their degree of substitutability or complementarity. The interaction, although insignificant, carries a negative sign implying that these two could be substitutes in the child health production. Using the National Maternal and Infant Health Survey data for 1988, Warner (1998) further considers the interaction of the number of care visits with visits delay and finds that, contrary to his previous work, delay does not have an effect on the fetal growth rate for both black and white women, concluding that the effect of prenatal care on birth weight is through visits rather than delay. Moreover, in the latter paper, the interaction effect implies that delay and visits could be neutral, or slightly complementary if anything, contrary to the previous findings.

Liu (1998) also accounts for pregnancy resolution following Grossman and Joyce (1990). Using the 1984 data for the Commonwealth of Virginia, analysis reveals that selection bias emanating from the use of prenatal care reduces the effectiveness of prenatal care on birth weight, while pregnancy resolution bias overestimates it. Estimates show that a one month`s delay in seeking prenatal care leads to an average birth weight reduction by 14 grams, 160 grams and 76 grams through OLS, 2SLS and 2SLS with selection respectively. However, contrary to findings by Grossman and Joyce (1990), Liu`s estimates yield a negative coefficient for the bias correction term (Mill`s ratio), interpreted to imply that the cost of abortion could be more prohibitive of mother`s termination of unwanted pregnancies in Virginia than in New York. A similar approach of controlling for sample selection in pregnancy resolution and endogeneity of prenatal care use is adopted by Rous et al. (2003) who find that increased use of prenatal care improves child`s birth weight in Texas. They control for the bias through the discrete factor method, which they argue performs better since its estimators do not necessarily

assume a normal correlation between errors terms. Results show that unaccounted endogeneity of prenatal care leads to an estimate that is about 45 per cent lower than when it is accounted for.

Evans and Lien (2005) investigate the effect of prenatal care on several indicators of birth outcomes (birth weight, gestation, maternal smoking and maternal weight gain) in Pennsylvania. They account for the endogeneity of care visits through the 2SLS estimator, with bus strike as an instrumental variable. Having established that the use of public transport was higher in Allegheny county, Pittsburgh city and among the blacks, they estimate the effect for these subsamples. Their results show that the effect of the strike in reducing the number of prenatal care visits is much pronounced among the black women and those who were at late stages of pregnancy during the strike. However, except for the reduction in the rate of maternal smoking, no significant effect is evident in other indicators of birth outcomes.

Evidence in developing countries

As indicated, the role of prenatal care on birth weight has been considered in developing countries. Guilkey et al. (1989) allow gestational duration to have an intermediate effect on infant birth weight and find that the frequent use of care is associated with few incidences of low birth weight in the Philippines. Through their multivariate analytical approach, they find that geographical area, type of provider and facility affect the effectiveness of prenatal care on birth weight. Visits to public practitioners indirectly improve birth weight through behavioural changes while visits to private practitioners have effect only in urban areas and no effect in rural areas. They find that increased prenatal care visits (to public and traditional practitioners) improve birth weight in urban areas, while private urban care visits negatively affect it. Using data from Cebu Longitudinal Health and Nutrition survey, Gajate-Garrido (2013) considers the effect of an adequate level of prenatal care use on birth weight, being small for gestational age and the incidence of premature birth in urban areas. Endogeneity in care use is accounted through the 2SLS method, with standardized accumulated below-average rainfall shock during mothers' pregnancy as an instrument. Results show that an adequate level of prenatal care use increases birth weight by 666 grams and lowers the likelihood of being small for gestational age and of premature birth.

The effect of prenatal care on birth weight has been investigated in South America (Jewell and Triunfo, 2006; Jewell, 2007). Jewell and Triunfo (2006) use data from the Latin American

Center for Perinatology and Human Development in Uruguay and account for care endogeneity through the 2SLS with marital status as the instrument. Their estimates show that early initiation of prenatal care is positively associated with birth weight, which increases by 57 grams compared to the 24 grams obtained through OLS. However, we concur with Gajate-Garrido (2013) that marital status could be a choice variable; hence its exogeneity may be questionable. Jewell (2007) uses demographic survey data for four countries; Peru, Columbia, Bolivia and Brazil and considers care measures of visits and initiation. The author also uses 2SLS and finds that delaying care initiation by a month reduces birth weight by 62.5 grams while increasing visits increases birth weight by 50.7 grams. Common in both studies is that OLS understates the effect of prenatal care on birth weight.

Similar results through the same empirical strategy are also found in Turkey, where the number of prenatal care visits increases birth weight by 39 grams for a sample of ever-married women aged 15 to 49 (Celik and Younis, 2007). Lin (2004) arrives at the same conclusion in Taiwan controlling for both pregnancy resolution and endogeneity of care use by married women. Selectivity to birth is controlled through the Heckman two-step procedure while care endogeneity is instrumented with the predicted value of its demand. Deb and Sosa-Rubi (2005) investigate the effect of both contents of prenatal care (a proxy for quality) and its time of initiation on birth weight in Mexico. In their analysis, they account for the endogeneity of prenatal care through the instrumental variables method (they instrument with indicators of care access, the density of health personnel and facilities, and eligibility of Progresa programme at municipality level). The authors find that while the content of prenatal care has a positive impact on birth weight, the timing of care has no effect.

In Azerbaijan, Habibov and Fan (2011) investigate the impact of quality care (an index based on what mothers reported to have been carried out during visits) and visits on birth weight. Using the 2006 Azerbaijan demographic and health survey, 2SLS controls for prenatal care endogeneity and Heckman approach control for potential sample selection due to non-availability of birth weight information. Their results show that care quality and visits increase birth weight by 41 grams and 26 grams respectively. However, there is no evidence of sample selection bias. In Tajikistan, Habibov et al. (2017) use the 2012 demographic and health survey to investigate the impact of prenatal care visits, quality and timing on birth weight. They instrument for care endogeneity with the number of previously born children and control for potential sample selection bias. Results show that visits, timing, (unlike in Mexico) and care

quality improve birth weight by 89, 304 and 130 grams, respectively. However, their instrument could be correlated with birth weight as established in the US (Rosenzweig and Schultz, 1983).

The effect of health inputs on birth weight has also been investigated in Kenya (Awiti, 2014; Mwabu, 2008). Awiti analyses the effect of adequate use of care and controls for both sample selection and endogeneity of care use (through the two-stage residual inclusion method) using the 2008 Kenyan demographic and health survey. The author finds that adequate use of care is positively associated with birth weight. Mwabu investigates the effect of vaccination against tetanus on birth weight, accounting for selectivity due to missing birth weight information through one step Heckman procedure. The analysis is for the 1996 data collected by the National Bureau of Statistics. Although results show a positive association between vaccination against tetanus and birth weight, the author highlights and demonstrates that this association may have been due to the possible omission of factors related to vaccination. This follows theoretical prediction that vaccination induces additional efforts for child`s health production and that the direct biological association between vaccination and birth weight has not been established (Dow et al., 1999). Dow et al. (1999) find evidence for the complementarity hypothesis using the demographic and health survey data for Malawi, Tanzania, Zambia, and Zimbabwe through family fixed effect estimation and instrumenting for a vaccination with the previous child`s vaccination status. Moreover, according to studies cited in WHO (2017:63) effectiveness of immunization during pregnancy declines with decreasing number of doses. For instance, in New Guinea, effectiveness stood at 94%, 65% and no effect for 3, 2, and 1 dose respectively. It appears, therefore, that, consideration of a binary variable for vaccination against tetanus as in the case of Mwabu (2008), although understandable due to data constraint, may not capture the hypothesis adequately.

Another significant contribution in the literature is the recognition that effectiveness of prenatal care in birth outcomes might be affected by the distribution in such outcomes (Conway and Deb, 2005; Deb and Sosa-Rubi, 2005; Abrevaya and Dahl, 2008). Conway and Deb caution about the approach to treat outcomes the same, yet some pregnancies are “normal” while some are “complicated”. In their analysis, they consider the time of initiating care, arguing that visits could be an outcome of complicated pregnancies. While this is notable, evidence has established that what could be referred to as potentially complicated pregnancies makes mothers initiate care earlier (Rosenzweig and Schultz, 1983; 1988) and failure to account for

visits could strongly assume that such women adequately utilized care (Joyce, 1994). Two main issues emerge from their results: the effect of initiating prenatal care earlier is evident among the normal pregnancies, and the majority of factors affecting birth weight through 2SLS estimator have an effect only on normal pregnancies. They, therefore, conclude that an approach to combine these types of pregnancies leads to prenatal care appearing ineffective.

Similarly, in the Mexican study Deb and Sosa-Rubi (2005) compare the impact of care for a sample of women with a normal birth weight range (2.5-4KG) to a sample of those with low birth weight children. They find that quality of care has a larger effect than the one evident in the entire sample. However, both quality and timing have significant effects in the low birth weight sample, although care quality effect is less than that in normal birth weight children. It is worth noting, however, that the authors caution that their analysis is threatened by potential sample selection bias. Abrevaya and Dahl (2008) also find that prenatal care has larger effects on birth weight of lower quantiles of distribution in the two states of Washington and Arizona.

In Botswana, to our knowledge, except for two studies (Okurut et al., 2013; Letamo and Majelantle, 2001), research on child health mostly focused on either child mortality or anthropometric indicators for children aged five years and below. Okurut et al. (2013) explore the determinants of the child`s birth weight with data for 1996, which is more than two decades ago. Since then, as indicated, the proportion of infants with low birth weight increased. Furthermore, a policy shift to have HIV/AIDS testing component into the prenatal care programme (BIDPA, 2010) might have affected the intensity of care use and suggests that the content and supply of care might have changed. As evident in Figure 1.1, enrolment into the prevention from mother to child transmission programme increased over time. Through the Heckman model, they consider vaccination against tetanus to be endogenous. However, according to Mwabu (2008), the positive association between tetanus vaccination and birth weight is “due to the omission of birth weight improving behaviours and investment that are induced by tetanus vaccination from the birth weight production function” (p. 217). Thus, the effect evident in the paper by Okurut et al. (2013) potentially suffers from this problem. Moreover, while they control for potential sample selection bias, they consider mothers` prenatal care utilization as a binary exogenous variable, yet women would endogenously differ in their levels of care utilization.

Letamo and Majelantle (2001) use a logistic regression approach for low birth weight and prematurity and consider care utilization as an exogenous variable as well. They find that women with 4-5 visits, which ideally should have been deemed adequate (WHO, 2016) are more likely to have low birth weight and premature children compared to those with at least six visits. Failure to account for care endogeneity may render care to be ineffective. Their results further show that women with experience of prior pregnancy termination are more likely to have low birth weight infants and no effect on prematurity. It could, however, be argued that this experience may condition the intensity of care use and failure to account for this is not without bias (Grossman and Joyce, 1990).

In summary and as the foregoing shows, mixed results on the effect of prenatal care are evident in the literature, to which we contribute. We note three aspects emerging from the review. First, scholars who control for pregnancy resolution decision find that the effect of bias is in different directions (Grossman and Joyce, 1990; Liu, 1998; Rous et al., 2003). Second, while evidence regarding the role of prenatal care on birth weight exists in both the developed and developing countries, the intensity (and perhaps approach) to research on this issue is affected by variability, complexity and difficulty in measuring the adequacy of its use and content (Alexander and Korenbrot, 1995). Consequently, as some maintain, the mechanism through which prenatal care may affect a child`s birth weight remains unclear (Lewit, 1977). It is evident that care initiation/delay is not consistently affecting birth weight (Warner, 1995; 1998; Deb and Sosa-Rubi, 2005). Thirdly, among those who control for potential sample selection bias due to non-availability of birth weight information, results are inconsistent as well (Habibov and Fan, 2011; Okurut et al., 2013). Therefore, a question on the role of prenatal care on infant health remains relevant particularly for Botswana given deteriorating health under a largely free care system, which among others might affect outcomes through visits and care facility deliveries (see Dzakpasu et al., 2014 for a review).

2.3 Data and Summary Statistics

We use cross-section data from the 2007/08 Botswana Family Health Survey (BFHS), which was conducted by the national statistics office, Statistics Botswana. This is the fourth (after 1984, 1988 and 1996) survey in the series. The survey collected information for under-five children, households and individual females aged 12-49 years. Additionally, for the first time, information facilitating the measure of adult nutrition was collected. The under-five child data file had a total of 2825 observations, with 106 observations having missing information (of

these, 44 were reported not to have been available for interview, 7 refused and 55 were for other reasons not specified). We excluded those children (N=717) whose respondents were not their mothers (and could not provide information for the mother) since we do not consider the characteristics of (non-mother) carers to be appropriate for the infant`s health production function. This reduced the sample to 2002 children. The sample was further reduced to 1986 after excluding 16 children whose mothers were aged more than 49 years, for whom data were unavailable (infants have to be merged with 12-49 aged female data). These children were identified by the availability of mother`s age variable in the under-five child data file.

The individual female data file had a total of 7319 cases. However, according to the survey report, "only 6916 women aged 12-49 years were successfully interviewed" (CSO, 2009a: 12) suggesting that 7319 represented all women eligible for the interview. The rest had the interview result codes of 2,3,4,5, and 6 for present but not available for the interview, postponed, refused, partially completed and other, respectively. From the 7319 cases, we excluded all except those with *completed* and *partially completed* interview outcomes, resulting in 6935 females. This file was then merged with the under-five children data file because it contains mothers` variables whose effects on infant birth weight are investigated. After merging the two data files, it emerged that 36 children had some missing information for their mothers (i.e. not matched from the master file). These could be those whose mothers had interview result codes of 2, 3, 4 and 6, whom we excluded because of lack of information. Such children were also excluded from the sample, resulting in a total sample of 1950 children matched to mothers.

The utilization of prenatal care information was sourced from the child and female questionnaires. On the child questionnaire, the reference point was the five years preceding the date of the interview. However, through the female questionnaire, when investigating the level of care utilization by mothers, reference was made to the live births that occurred within two years before the interview date. Based on the measure of prenatal care adopted in this chapter as well as the need to reduce the recall bias associated with a household survey (Ansah and Powell-Jackson, 2013), a consideration of the latter resulted in the final sample of 1395 children.

An examination into the availability of infant birth weight information (which was extracted from the growth monitoring cards and is the outcome of interest) reveals that 236 children had missing information. The question, as stated by Heckman (1979:154), remains "why are the data missing?" Of the 236 with missing birth weight information, 182 children had missing information because the interviewers did not see their cards while 54 children had missing information though their cards were seen. The reasons for the cards not seen include being either lost or burnt. Moreover, it could possibly have been due to such children being born at home and therefore not weighed at/after birth. Birth weight information availability may be linked to contact with health professionals (Mwabu, 2008). We explore this possibility by considering "assistance at delivery", from which the infant`s place of birth can be inferred. Of the total 54 children, 22 were reported to have been assisted by relatives/friends, 16 assisted by nurse/midwife, 6 assisted by traditional birth attendants, 7 assisted by self, while 3 were said to have been assisted by doctors. From these, it may be concluded that those who were not assisted by professional health personnel are 35. These are the children whom we may conclude were probably born at home because health professionals would not visit homes for birth assistance. The possible reason for the missing of birth weight information for those who were assisted by doctors and nurses could be that they lost their first cards.

We also linked some health district-level data (average retail food prices, distance to the nearest health facility as well per capita health care facilities) with the survey data. Monthly data on food prices are collected by Statistics Botswana from 46 areas, including cities and towns, urban villages and rural areas. Food prices are considered as they affect purchasing power with implications on mothers' health and health care utilization (Thompson, 2009). This is likely to be the case for Botswana, where the majority of households are net food buyers. Their consideration is also worthy given that food purchases were identified as the most important household source of food, even among the subsistence agricultural households in the country (Mmopelwa and Seleka, 2012; Seleka and Lekobane, 2017). We used the average prices for cereals (the most consumed in Botswana).

Statistics Botswana also produced data on primary health care accessibility for the period 2006/2007. The data shows the proportions of the population who are within three categories of distance to the nearest health care facilities; 0-5, 5-8 and 8-15 kilometres in health districts (Statistics Botswana, 2012; 2017). The data reveals that at the national level, 84% of the population were within the 5km radius, 11 per cent within 5 and 8km while the rest were within

8 and 15 kilometre radius. Moreover, the proportion at urban areas stood at 96% and 4% for 0-5 km and 5-8 kilometres respectively, whereas 72%, 17% and 11% per cent were in the 0-5, 5-8 and 8-15 kilometres range respectively in rural areas. We used this information to create a dummy variable for the probability of care facility being sufficiently close, also factoring in rural-urban differences. For instance, we have a dummy of 1 if mothers resided in a district whose proportion of those in 0-8km is higher than 95% (observed at a national level), otherwise 0. However, since it has been observed that in urban districts the proportion within 8km is 100, we consider 0-5km for the urban subsample. Thus, an urban district was assigned a value of 1 if the proportion of those residing within 5 km radius is higher than that of all urban districts combined. Table 2.2 presents variable definitions.

Table 2.2: Variable Definitions

Variable	Definition
A. Outcome variable	
Birth Weight (KG)	Infant`s weight at birth (in kilograms)
B. Instrumented Variable	
APC	Adequate Prenatal Care utilization: Indicator for 4 or more visits (1=Yes)
C. Infant Characteristics	
Male	A dummy for sex of the infant (1=Male)
Birth weight reported	A dummy for the availability of infant`s birth weight (1=Yes, 0 otherwise)
QoB1	An indicator for the infant born in January-March (1=Yes, 0 otherwise)
QoB2	An indicator for the infant born in April-June (1=Yes, 0 otherwise)
QoB3	An indicator for the infant born in July-September (1=Yes, 0 otherwise)
QoB4	An indicator for the infant born in October-December (1=Yes, 0 otherwise)
D. Mothers Characteristics	
Age	Mothers` age in completed years
Pre-Secondary	Dummy for mothers with pre-secondary education (1=Yes, 0 otherwise)
Married	Dummy for married mothers (1=Yes, 0 otherwise)
Christianity	Dummy for Christianity as mother`s main religion (1=Yes, 0 otherwise)
Height	Height in centimetres
Rural	A dummy for mother residing in rural areas (1=Yes, 0 otherwise)
E. Variables at the district level	
Rice price	Price for Rice per kg
Maize price	Price for maize meal per 5kg
Sorghum price	Price for sorghum meal per 5kg
Proximity	Dummy for higher proportion within a certain distance radius than at national level
Care Facility	Number of fixed health care facilities in health districts

Table 2.3 presents summary statistics, comparing (through two-sample t-test) urban and rural areas, given the possible differences in characteristics between the two. An urban area is defined as "any settlement with a population of at least 5000 of whom 75% of the workforce is engaged in non-agricultural activities" (Republic of Botswana, 2014:3). However, since the district is larger, it is possible to find both urban and rural areas under the same district. The

average birth weight is about 3 kilograms. Of the 95% of infants who were taken for prenatal care during pregnancy, 73% had adequate utilization, with a higher average in urban areas. About 49 per cent of infants are males, and 83 per cent of infants had their birth weight reported. Possible reasons for this are explored in the preceding sub-section. The average shares for the quarter of birth dummies are between 23 and 28 per cent, with third quarter dominating.

The average mother`s age is 28 years. About 70 per cent of infants are born of mothers with a secondary level of education, followed by 22 per cent of those with a primary education level. The least share is accounted for by those whose mothers had non-formal education. In our estimation, we re-define the education variable to have two categories pre-secondary and post-primary. This is because, for child health, it has been found that primary education level is sufficient (Charmarbagwala et al., 2004). Only 17 per cent are from married mothers; the majority of infants were born to cohabiting and never married (single) mothers. About one per cent have either widowed, separated or divorced mothers. A pattern of relatively lower averages for *married* mothers evident in the table has been observed in previous studies; a possible explanation is that cohabitation (which increased over time) serves as a temporary phase (i.e. a prelude) before marriage (Mokomane, 2005). Thus, cohabitation is seen not as a substitute for marriage, but rather as delaying its timing. Indeed, there have been initiatives promoting marriage and creating awareness on the side effects of cohabitation (Morwaagole, 2013). The average for those residing in rural areas is 46 per cent. Average retail prices are higher in urban than in rural areas.

Table 2.3: Summary Statistics

	Full (1)	Urban (2)	Rural (3)	Differences (Urban-Rural) (4)
A. Outcome Variable				
Birth Weight	3.085	3.101	3.064	0.038
B. Instrumented Variable				
Adequate Prenatal Care	0.725	0.766	0.677	0.089***
C. Infant Characteristics				
Male	0.497	0.507	0.486	0.021
Birth weight reported	0.831	0.874	0.781	0.092**
QoB1	0.232	0.246	0.214	0.032
QoB2	0.241	0.244	0.239	0.005
QoB3	0.269	0.268	0.270	-0.003
QoB4	0.258	0.242	0.276	-0.034
D. Mother`s Characteristics				
Age	28.014	28.340	27.633	0.706**
Pre-Secondary	0.295	0.215	0.388	-0.172***
Married	0.170	0.209	0.124	0.084***
Christianity	0.806	0.827	0.781	0.046**
Height	161.432	162.021	160.744	1.278***
Rural	0.462			
E. Variables at District level				
Rice price	8.785	8.570	9.036	-0.466***
Maize price	21.906	21.468	22.416	-0.949***
Sorghum price	22.861	22.732	23.011	-0.279***
Proximity	0.832	0.874	0.784	0.089***
Care Facility	5.788	6.020	5.518	0.502

Notes: The first column provides means for the full sample, while columns 2 and 3 provide means by area of residence. The fourth column presents the difference between the urban and rural means for the considered variables. ***significant at the 1 per cent level, **significant at the 5 per cent level.

2.4 Model Specification

2.4.1 Theoretical Model

Behrman and Deolalikar (1988:637) state that "the proximate determinants of individual`s health and nutrition usually are decisions made by the individual or by the household in which he or she lives". Moreover, since many individuals live in household consumption units, models have been constructed such that households maximize utility subject to given constraints, the framework standard utility function is specified as follows (Rosenzweig and Schultz, 1982):

$$U = U(X, Y, H) \tag{2.1}$$

In (2.1), *X* represents a neutral commodity (e.g. transport) that does not directly affect infant health; *Y* represents a health-related commodity (such as food) that influences infant health,[3] and *H* is infant`s health status. Infant health is further affected by household characteristics and other inputs (*Z*) acquired by the family, such as health care, as well as by family health endowments (μ), known to them (such as genetic or environmental factors) but over which they have no control. Therefore, an infant health production function is represented as follows:

$$H = F(Y, Z, \mu),\ F_y, F_z, F_\mu \neq 0 \qquad (2.2)$$

The family maximizes utility represented by equation (2.1), given the health production (2.2), subject to the following budget constraint:

$$I = PxX + PyY + PzZ \qquad (2.3)$$

Where *I*, P_x, P_y and P_z are income and prices of neutral good, health-related good and health investment goods respectively. As food is the major component of household expenditure, especially in rural areas, in general, P_xX is likely to have only a small effect. Health is not directly acquired and is regarded as "utility-augmenting" good. The health-related good (*Y*), has both a direct effect in augmenting utility and indirect effect on utility, through *H*, as captured by equation (2.2), while *Z* has only indirect effect and *X* has only direct effect on utility. The model can be represented as demand equations for the three commodities in terms of prices and income:

$$X = Dx(Px, Py, Pz, I, \mu) \qquad (2.4)$$
$$Y = Dy(Px, Py, Pz, I, \mu) \qquad (2.5)$$
$$Z = Dz(Px, Py, Pz, I, \mu) \qquad (2.6)$$

Substituting the effect of changes in prices on infant health from equations (2.4-2.6) the change in infant health can be expressed as:

$$dH = FydY + FzdZ + F\mu d\mu \qquad (2.7)$$

3 This good may still be consumed even if it happens that it has no effect on health at all, or if it severely affects health.

The next section explains the empirical specifications we use to test the model.

2.4.2 Empirical Model

Prior to presenting the empirical model(s), we discuss potential issues surrounding the estimation strategy. The first issue is *endogeneity*. As noted already, efforts to unravel the effect of prenatal care on infant birth weight may be affected by bias brought by the association of birth weight with prenatal care utilization, which is an outcome of the selection process partly determined by non-observable factors. The selection processes that may introduce bias when investigating the role of prenatal care on infant health may be classified into four types; favourable, adverse, estrangement and confidence selection (Frick and Lantz, 1996). The favourable selection process is said to occur when women at low risk of poor birth outcomes are also higher users of prenatal care, in which case there is potential to bias the effect of prenatal care upward, leading to its overestimation: "epidemiological literature has acknowledged that favourable selection may be a more serious source of confounding" (Grossman and Joyce, 1990: 984-985).

Adverse selection occurs when high-risk women are higher users of prenatal care, potentially biasing the effect of prenatal care downwards. In this case, poor birth outcomes would be because of the factors that cause women to utilize care more, instead of the care itself. Joyce (1994) notes that studies that corrected for adverse selection showed larger effects of prenatal care than those that did not. With estrangement selection, some high-risk women (such as those who suffer from abuse) are among the lowest users of prenatal care, including those with no care at all. Lastly, confidence selection process implies that women of higher risk use care more, and it is similar to the adverse selection process (Frick and Lantz, 1996).

The second issue is that of *sample selection bias*, which is said to occur because of using a subset of a random sample, either because of the survey design or non-response on survey questions (Wooldridge, 2010:791; Heckman, 1979). As indicated in section 2.3, some infants had missing birth weight information, which is not an outcome of a random process. According to Awiti (2014:4), "sample selection bias will occur if the unobservable factors affecting the decision to report the birth weight of the child are correlated with the unobservable factors affecting the birth weight itself".

The above issues are addressed using an instrumental variable (IV) strategy and a control function approach (Angrist and Pischke, 2009; Wooldridge, 2010; Heckman, 1979). Provided the instrument is good, "IV methods solve the problem of missing or unknown control variables, much as a randomized trial obviates extensive controls in a regression" (Angrist and Pischke, 2009:115). We, therefore, use the two-stage regression to control for endogeneity in prenatal care utilization and Heckman selection model to control for potential sample selection bias. Instrumental variable strategy requires a variable or variables not affecting outcome directly (instruments), but that explains the endogenous regressor and requires assumptions for the purposes of causal interpretation. First, instrument(s) should have a clear effect on the endogenous variable (prenatal care use), implying that they should have a statistically significant effect or that they should be relevant. Second, their effect on birth weight should be through the first stage. The second assumption, known as exclusion restriction, has two components; instrument(s) should be such that they are not correlated with the error term in the birth weight equation, implying that they should be exogenous and that they have no effect other through the first stage (Angrist and Pischke, 2009).

Following Rosenzweig and Schultz (1982, 1983) and other studies influenced by their approach, the equation of interest (birth weight production function) may be described as follows:

$$HBW_i = \alpha_0 + \alpha_1 APC_i + \alpha_2 X_i + \varepsilon_1 \quad (2.8)$$

Where HBW_i is the log of birthweight (in Kilograms) for infant i, APC_i is an indicator for the adequate level of prenatal care utilization by mother during pregnancy for infant i, X_i is a set of control variables for the mother and the infant and ε_1 is the stochastic error term. The first stage equation, which captures the effect of instruments (*Inst*) on adequately utilizing prenatal care, is then specified as follows:

$$APC_i = \beta_0 + \beta_1 Inst_i + \beta_2 X_i + \varepsilon_2 \quad (2.9)$$

Endogeneity of care utilization is captured through the relationship between the two errors ε_1 and ε_2, represented by the correlation coefficient, rho (ρ). In this case, we test the null hypothesis that rho=0, against the alternative that it is not.

Since our endogenous regressor (APC) is a binary variable, we do not directly apply the two-stage least squares (2SLS) estimator. This is because "the conditional expectation function (CEF) associated with the first stage is probably nonlinear; hence OLS first stage will be an approximation to the underlying nonlinear CEF" (Angrist and Pischke, 2009: 190). In such a case Angrist and Pischke caution against "forbidden regressions", which are situations in which either the fitted values from the non-linear (for instance, probit) first stage are substituted for actual values in the second stage equation or used as instruments. As they put it, "naively plugging in the first stage fitted values in non-linear models is a bad idea" (pp.192). While other problems (such as the non-normality and heteroscedasticity of the error term as well possibilities of the fitted values lying outside the 0-1 range), associated with applying a linear model on binary variables may be overcome, the fundamental problem remains in the assumption that the marginal effect of explanatory variables is on the likelihood of the event occurring constant throughout (Gujarati, 2004:595). Such assumption may not hold in our case given endogeneity of care utilization. Assuming linearity led to insignificant and smaller effect of family income on education in Norway (Løken et al., 2012). However, when the linearity assumption/restriction was released, results were along with the theoretical prediction.

However, this binary variable may be viewed as a treatment indicator, resulting in the estimation of the treatment effect model (Cameron and Trivedi, 2010; Wooldridge, 2010). The model we consider specifically "estimates average treatment effect and the other parameters of linear regression model augmented with an endogenous binary treatment variable" (StataCorp, 2017:36), thus allowing for correlation between the unobservables that may affect treatment and unobservables that only affect potential outcomes. We maintain the stable unit treatment value assumption (SUTVA), that is, adequate utilization of prenatal care by one mother should not affect the infant birth weight for the other mother. Another important assumption in the estimation of the average treatment effect is that of ignorability or unconfoundedness (Morgan and Winship, 2007). The assumption states that conditional on observable covariates, assignment to treatment is independent of potential outcomes. However, because differences in mothers` health awareness are unobserved, we hypothesize that assignment to treatment (adequate utilization of prenatal care) and outcome (infant birth weight) may not only be determined by observable variables; hence we may not assume conditional independence between the two. Formally, the model, which uses both equations (2.8) and (2.9), where assignment to treatment is through an unobservable latent variable, is re-specified as follows:

$$HBW_i = \alpha_0 + \alpha_1 APC_i + \alpha_2 X_i + \varepsilon_1$$

$$APC_i = \begin{cases} 1 & if \quad APC^*_i > 0 \\ 0 & otherwise \end{cases} \tag{2.10}$$

The birth weight information was obtained from the card. However, as not all children had birth weight reported, an approach to consider only those with the information could introduce sample selectivity bias (Heckman, 1979). This would arise if the unobservables determining the inclusion of infants into the sample are correlated with those affecting infant birth weight. Following Vella (1998:129), our sample selection model is specified as follows:

$$HBW_i^* = x_i'\beta + \varepsilon_i;\ i = 1, \ldots, N \tag{2.11}$$

$$BWR_i^* = z_i'\gamma + v_i;\ i = 1, \ldots, N \tag{2.12}$$

$$BWR_i = 1\ if\ BWR_i^* > 0;\ BWR_i = 0\ otherwise \tag{2.13}$$

$$HBW_i = HBW_i^* * BWR_i \tag{2.14}$$

Where, HBW_i^*, the infant health measured by birth weight, is sample-induced endogenous, which is observed when $BWR_i = 1$. Equation (2.12) is a probit selection equation where BWR_i is a dummy variable with a value of 1 if the infant`s birth weight information is available and 0 otherwise. x_i and z_i are vectors of independent variables affecting infant birth weight and selection into sample respectively, while β and γ are parameters to be estimated. The error terms, ε_1 and v_i in equations (2.11) and (2.12) are normally distributed, N(0,Σ), where the covariance is

$$\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v^2 \end{pmatrix} \tag{2.15}$$

These two errors may be correlated (Hartman, 1991; Cameron and Trivedi, 2010) which, as indicated, assumes unobserved sample-induced heterogeneity, also captured through ρ. However, following Mwabu (2008), it may be assumed that the two errors are not independent of the instruments as non-linear interactions between unobservables and care utilisation may be omitted from the equation of interest. But instruments and *vi* are assumed to be independent (Mwabu, 2008). We estimate this through the Heckman sample selection model (Heckman, 1979). Estimation of the Heckman model is done using two approaches: (i) the two-step

approach and (ii) one step (full information maximum likelihood) approach. The first step of the two-step approach involves generating a correction factor (inverse of Mills ratio) from the probit estimates of the birth weight report model (equation 2.12). In the second step, an OLS model is estimated with the inverse of the Mills ratio as an additional regressor (Vella, 1998). While this approach is said to be consistent, it is less efficient than the full information maximum likelihood approach (Wooldridge, 2010; Hartman, 1991), which we adopt.

According to Mwabu (2008), when the above issues are considered and given that they cannot be determined a priori, birth weight equation (2.8) may be re-specified as follows:

$$HBW_i = \alpha_0 + \alpha_1 APC_i + \alpha_2 X_i + \gamma_i E_i + \Psi_i (E_i \times APC_i) + \lambda_i \beta_i + \varepsilon_1 \quad (2.16)$$

where the correction terms *E*, (*E*×*APC*) and β represent residuals of the endogenous treatment variable, possible non-linear interaction of the residual with the endogenous treatment, and inverse Mills ratio respectively. The rest of the variables are as previously defined and α, γ, Ψ, and λ are parameters to be estimated. The residuals account for the unobserved factors that might be correlated with utilization of prenatal care (Mwabu, 2008; Heckman and Robb, 1985). Interacting the residual with health care input accounts for the possible non-linear interaction of unobservable factors with other factors affecting infant birth`s weight.

Together with the approach to variables in previous studies (Rosenzweig and Schultz, 1982; 1983; Habibov et al., 2017) as well as those cited in Awiti (2014) and our assumptions, we instrument for adequate prenatal care utilization with the probability of care facility being sufficiently close and availability of care facilities in health districts. We use these instruments even for identifying the birth weight reporting. Our choice of proximity to a health care facility is based on the following. First, along with the theory of time allocation (Becker, 1965), longer distances might present higher opportunity costs for non-market activities. There is evidence that distance to health care facility affects care utilization. For instance, shorter distances led to hospital births in the Netherlands (Daysal et al., 2015) and higher health care demand in China (Qian et al., 2009). It has also affected adequate prenatal care use in Kenya (Awiti, 2014) and the use of preventive care among children in America (Curry and Reagan, 2003). However, the potential weakness with our distance variable is that it is at the district level, hence a "second-best proxy". It is for this reason that we include district fixed effect and cluster the standard errors, as there can be differences within a district.

While there could be worries that mothers may choose to reside nearer to the health care facilities because of some unobserved factors (Rosenzweig and Schultz, 1982), we believe our measure for distance addresses that. Firstly, as per the land allocation system, one does not necessarily choose where to be allocated. An application for a plot of residence is submitted, and the final decision lies with the land board department; hence the effect of being close to the health care facility does not lie with the mother. Secondly, we consider distance at the health district level and argue that even if mothers were to choose to be nearer care facilities, this would not be a simultaneous decision for all of them (Awiti, 2014). We also separate rural and urban samples since migration is likely to be from rural to urban areas (CSO, 2009b).

In the context of access to care, the policy for health care facility distribution can be argued to be exogenously determined by the suppliers (private and public). Although some women may temporarily migrate to urban areas when pregnant, they are more likely to still utilize public health care facilities, which provide services freely. To our knowledge, no study has been conducted in Botswana to examine the pregnancy-induced migration. However, studies conducted elsewhere find no statistically significant effect of migration in response to the expanded care (Schwartz and Sommers, 2014). Moreover, according Letamo and Majelantle (2001), women in Botswana utilize prenatal care and give birth in their usual place of residence.

2.5 Results and Discussions

Three models are estimated; ordinary least squares, linear regression with endogenous treatment accounting for endogeneity of prenatal care utilization and Heckman sample selection model, which accounts for potential sample selection bias and heterogeneity. Table 2.4 presents results for adequate prenatal utilization (treatment variable for the treatment model) estimated using the maximum likelihood estimator. The table shows that having pre-secondary education compared to post-primary education (the reference category) reduces the likelihood to utilize prenatal care adequately. This finding implies that mother`s education is an essential factor in influencing demand for care as has been observed in the literature (Simkhada et al., 2007; Mwabu, 2008; Awiti, 2014). Moreover, the higher probability of care facility being close, and availability of care facilities are positively associated with the likelihood of adequate prenatal care utilization. Rural residence reduces the likelihood to utilize prenatal care adequately.

Table 2.4: First Stage Regression Results

Variable	Full (1)	Urban (2)	Rural (3)
Age	0.021	-0.018	0.049
	(0.054)	(0.058)	(0.088)
Age squared	-0.000	0.001	-0.000
	(0.001)	(0.001)	(0.002)
Married	-0.030	0.071	-0.268
	(0.111)	(0.094)	(0.279)
Male	0.064	0.046	0.090
	(0.076)	(0.110)	(0.099)
Height	0.005	0.002	0.010
	(0.007)	(0.009)	(0.011)
Pre-Secondary	-0.241**	-0.216	-0.254
	(0.116)	(0.155)	(0.211)
Other religion	-0.299	-0.286	-0.277
	(0.224)	(0.233)	(0.273)
No religion	0.081	0.115	0.038
	(0.144)	(0.204)	(0.283)
Proximity Likelihood	0.455***	0.226**	0.578***
	(0.085)	(0.110)	(0.087)
Care Facility	0.041***	0.046*	0.059***
	(0.011)	(0.026)	(0.019)
Rural	-0.268***		
	(0.081)		
Joint Significance (Chi2)	36.35***	42.17***	49.44***

Notes: Each column provides estimates from separate regressions. The first column provides results for the full sample, while the remaining columns provide for area of residence. The dependent variable is a dummy for adequate care utilization. All standard errors, in the parenthesis, are clustered at fixed effects. Omitted category for religion is Christianity. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table 2.5 reports the ordinary least squares and second stage regression results for the treatment effect model. Adequate prenatal care utilization is positively associated with birth weight, and failing to account for endogeneity reduces its effect. From the treatment effect model results, moving from inadequate to adequate care utilization improves birth weight by 22, 24 and 21 per cent in full, urban and rural samples respectively. These translate to the respective improvements in birth weight by 0.67, 0.73 and 0.64 kg compared to 0.12, 0.10 and 0.15 kg when endogeneity is unaccounted for. Age of the mother is associated with birth weight in full and urban samples only. While mother`s age carries a positive sign, age squared carries a negative sign suggesting that the likelihood of either impaired intrauterine growth or gestational duration increases for aging mothers. By implication, reproductive health stock depreciates with mother`s age.

Table 2.5: OLS and Second Stage Regression Results

	OLS			Treatment Effect Model		
	Full	Urban	Rural	Full	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)
APC	0.038***	0.032*	0.049**	0.216***	0.237***	0.209***
	(0.012)	(0.018)	(0.019)	(0.048)	(0.039)	(0.076)
Age	0.017**	0.015	0.019*	0.017**	0.016**	0.015
	(0.007)	(0.009)	(0.009)	(0.007)	(0.008)	(0.011)
Age squared	-0.000**	-0.000	-0.000*	-0.000**	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.042**	0.027	0.072**	0.043**	0.023	0.083***
	(0.018)	(0.026)	(0.021)	(0.018)	(0.023)	(0.027)
Male	0.028***	0.023*	0.029**	0.022**	0.018	0.024
	(0.009)	(0.013)	(0.013)	(0.010)	(0.015)	(0.016)
Height	0.003***	0.004***	0.002**	0.003***	0.003***	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Pre-secondary	0.007	0.000	0.016	0.023*	0.013	0.032*
	(0.011)	(0.021)	(0.010)	(0.012)	(0.021)	(0.018)
Rice price	0.038***	-0.008	0.055***	0.045***	0.018	0.058***
	(0.009)	(0.014)	(0.008)	(0.008)	(0.012)	(0.008)
Maize Price	-0.007*	0.020***	-0.026***	-0.017***	0.007	-0.034***
	(0.004)	(0.006)	(0.004)	(0.004)	(0.007)	(0.007)
Sorghum price	-0.003	0.002	-0.006	-0.006	-0.007	-0.001
	(0.003)	(0.007)	(0.004)	(0.004)	(0.007)	(0.004)
QoB1	-0.019	0.000	-0.038*	-0.020	-0.002	-0.034*
	(0.013)	(0.022)	(0.017)	(0.013)	(0.023)	(0.017)
QoB2	-0.033	-0.029	-0.037	-0.033	-0.025	-0.038*
	(0.021)	(0.033)	(0.024)	(0.020)	(0.031)	(0.023)
QoB4	-0.004	0.022	-0.037*	-0.005	0.023	-0.037*
	(0.014)	(0.022)	(0.021)	(0.013)	(0.022)	(0.021)
Other religion	-0.003	0.060	-0.055	0.012	0.080*	-0.044
	(0.027)	(0.036)	(0.055)	(0.019)	(0.042)	(0.043)
No religion	0.014	0.021	0.016	0.011	0.009	0.017
	(0.011)	(0.018)	(0.015)	(0.012)	(0.020)	(0.022)
Rural	-0.007			0.009		
	(0.014)			(0.014)		
Proximity	0.003	0.005	0.006			
	(0.009)	(0.011)	(0.014)			
Care Facility	0.000	0.000	0.002			
	(0.000)	(0.000)	(0.001)			
lambda				-0.108	-0.122	-0.099
				(0.031)	(0.027)	(0.051)
Observations	1159	656	503	1159	656	503

Notes: The first three columns provide OLS regression results for full, urban, and rural samples. Columns 4-6 provide results from the treatment effect model. The dependent variable is log of birth weight. All standard errors, in parenthesis, are clustered at fixed effects. Omitted categories are QoB3 for the quarter of birth and Christianity for religion. APC: Adequate utilization of prenatal care. QoB: Quarter of birth. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

The incidence of being married is positively associated with birth weight in full and rural samples. Such an effect might signal the role played by husbands, which is through their characteristics that may add on to the mother`s socio-economic environment. In Kenya, for instance, husband`s education was found to have an effect on the likelihood of vaccination against tetanus (Mwabu, 2008). Warner (1995) also found that the relation of the child`s parent benefited the infant. Due to data limitations, we are not able to investigate the effect of husband`s characteristics on either infant birth weight or utilization of prenatal care. While BFHS interviewed males in the 12-49 age category, there is no variable indicating that they were the fathers of these children aged below five years or husbands to the women considered by the survey. Furthermore, the 12-49 aged female questionnaire made no reference to the characteristics of husbands. Being a male infant is positively associated with birth weight, although in the full sample only. Kramer (1987) reviews studies whose conclusions were that infant`s sex did not affect either gestational age or prematurity. Nonetheless, they found that male infants were found to have larger weight and lower risk of intrauterine growth rate. Mother's height is also positively associated with birth weight across all the samples.

For full and rural samples, having pre-secondary education is positively associated with birth weight. It is not easy to interpret this effect, given that pre-secondary education also reduces the likelihood of adequate prenatal care utilization. While pre-secondary education may be enough for child nutrition (Charmarbagwala et al., 2004), Celik and Younis (2007) find that education has no direct effect on birth weight. In fact, Rosenzweig and Schultz (1982:59) maintain that "it is doubtful that schooling can affect the production of health without being associated with some alteration in an input". They argue that the effect of education is through the changing perceptions of the relationship between health and its inputs. Warner (1998) also finds no effect of mothers' education for black mothers and small effect for white mothers. From that, the author also concludes that the pattern was also suggestive of Rosenzweig and Schultz`s view.

With regards to average retail food prices, since we do not observe whether households produced the considered food, results are likely to remain ambiguous. Theoretically, changes in food prices might affect household food security and mothers' nutrition. However, their effects would vary depending on several factors, including household`s food price change sensitivity, coping strategies as well as food source diversity, which might in part be affected by benefiting from food transfers. Nonetheless, the price of rice is positively correlated with

birth weight across all the samples. This could, in part, reflect the fact that rice is not locally grown in the country and has no import restrictions, unlike maize and sorghum. Maize meal price is negatively associated with birth weight in full and rural samples. In Kenya, Grace et al. (2014) also found that maize prices alone did not have an effect on birth weight but became significant when interacted with vegetation index. Overall, it would appear that perhaps food prices are capturing some of the unobserved variation in district/community characteristics.

Table 2.5 further shows that against the third quarter of birth dummy, all other quarters reduce (at 10% level) infant`s birth weight in the rural sample. This might suggest some temperature and precipitation effects, which we do not observe in our data. In their study, Grace et al. (2015) provide literature, although with inconclusive results, with some effects of weather on birth weight. From such review (i) the effect of extreme temperature has been found to be more than that of seasonal temperature variation, (ii) low birth weight was observed to be more frequent in autumn-winter periods and (iii) high precipitation could lead to low birth weight (Grace et al., 2015).

2.5.1 Testing for Regressor Endogeneity

Treatment effect model produces a parameter that captures the relationship between unobservables determining the assignment to treatment and unobservables affecting the potential outcome, which as indicated in section 2.4, is represented by rho (ρ). Other parameters are sigma (σ) and lambda (ρσ). According to Cameron and Trivedi (2010:194), /athrho (=0.5*ln[1+ρ]/[1-ρ]) and /lnsigma reported in the bottom of Table 2.5 are transformed outcomes, that ensures $\hat{\sigma} > 0$ and $|\hat{\rho}| < 1$. Regressor endogeneity is tested against the null hypothesis that $\rho=0$. Across all the samples, the null hypothesis of no correlation between these unobservables is rejected through the Wald test, at 1 per cent level for full and urban samples, and at 10 per cent level for the rural sample. We, therefore, conclude that adequate prenatal care utilization is indeed an endogenous variable. The negative signs for the rhos in Table 2.5 (-0.560, -0.594, -0.569), imply that unobservables that raise birth weight occur with unobservables that lower adequate prenatal care utilization.

2.5.2 Relevance and Validity of Instruments

Literature acknowledges the difficulty in identifying good and strong instruments (Bound et al., 1995; Crown et al., 2011). According to Crown et al. (2011), there is a likelihood for strong

and valid instruments to be also correlated with unobservables. This, therefore, presents a trade-off, resulting in the use of weak instruments, which at the same introduces larger bias. Instruments are termed good if they are exogenous (not correlated with the error term of the structural equation) and relevant (being statistically significant for the endogenous regressor). However, the possibility of weak but relevant, and exogenous, but irrelevant instruments have been observed (Stock et al., 2002). According to Stock et al. (2002), instruments are not relevant if their joint significance on an endogenous explanatory variable is statistically not different from zero. Table 2.4 shows the overall effect (joint significance) of our excluded instruments and suggest that they are valid and relevant. We also find no evidence of a direct association between instruments and birth weight, with the highest correlation coefficient of 0.43 between the probability of health care facility being sufficiently close and birth weight in the rural sample.

2.6 Robustness Checks

Establishing the validity and strength of instruments is difficult, especially when the endogenous variable is binary (as applies for APC). Moreover, the treatment effect model has been questioned for the assumption of joint normality of errors and non-allowance of the calculation of causal effects under observed heterogeneity (Cerulli, 2014). Our results have so far demonstrated the relevance of the instruments on the basis of their statistically significant effect on adequate prenatal care utilization. As we do not directly apply 2SLS from which we could derive the F statistic for joint significance in the first stage, we appeal to two alternative techniques; control function approach and the Two-Step IV model (Wooldridge, 2010; 2015). The control function approach we employ here requires the use of instrumental variables and is estimated in two steps. The first step involves the estimation of the probit model for adequate prenatal care (APC) utilization and obtaining the generalised residual as follows (Wooldridge, 2015).

$$\widehat{gr}_i = APC_i\lambda(Inst_i\gamma) - (1 - APC_i)\lambda(-Inst_i\gamma) \qquad i=1, \ldots, N. \qquad (2.17)$$

where, as in the previous, APC_i is adequate prenatal care utilization, and $Inst_i$ is a set of instrumental variables. The second step involves OLS estimation of birth weight with the generalised residuals as an additional regressor. Statistical significance of the residuals suggests that the considered endogenous variable is indeed endogenous. Along equation (2.16), residuals could be interacted with this endogenous variable. Table A2.1 shows that for full and

urban samples, the generalised residuals are significant, while in the rural sample, it is significant when considered with the residual-care utilization interaction.

Although the control function approach is considered efficient even with weak instruments (Wooldridge, 2007), it may equally be argued that it could be sensitive to the specification of the first step probit model (Wooldridge, 2015). For that reason, we also consider the Two-step IV model. With this model, we first estimate a probit model by maximum likelihood, obtain fitted probabilities, and estimate the second step by IV using fitted probabilities and exogenous variables. This method is robust to misspecification of the first (probit) stage model, i.e., it has the property of robustness: as predicted probabilities are used as instruments, there is no strict requirement for a correct specification of the probit model (Wooldridge, 2010). This is different from using fitted values of the endogenous regressor in place of the actual values, which is inappropriate (Angrist and Pischke, 2009). As evident in Table A2.2, birth weight increases by about 13, 34 and 11 per cent in full, urban and rural areas respectively (see results in Appendix).

2.6.1 Quality of Care

It is not possible to explore the quality of prenatal care due to data limitations. We do not observe, for instance, the detailed contents of care as well as the time during which it was initiated. The complementary hypothesis suggests that mothers may invest more in care when they believe it to be beneficial (Mwabu, 2008). By implication, some may decide to alter their subsequent recommended visits. One possible effect is being vaccinated against tetanus, which has been found to positively correlate with birth weight in Botswana (Okurut et al., 2013). Hypothesising that vaccination may in part reflect quality, we constructed a dummy variable of 1 if care was adequately sought simultaneously with receipt of injection against tetanus and re-estimated models. For the full sample, this increases birth weight by about 2 percentage points (coefficient is 0.232 as opposed to 0.216), above the effect when injection against tetanus is not considered for the treatment effect model (see Table A2.3). While the channel of the effect may not be known with certainty, the conclusion on care effectiveness still holds.

2.6.2 Sample Selection

As not all infants had birth weight information recorded in their growth monitoring cards, estimating for only infants with birth weight information may introduce selection bias (Heckman, 1979). However, applying the Heckman sample selection model (equations 2.11-2.13) revealed no presence of bias in rural and urban areas separated (see Table A2.4). While

the absence of birth weight information is generally linked to the absence of health professionals at the birth (i.e. home births), these results suggest that this may not always be the case; for example, mothers could lose the growth monitoring cards from which birth weight information is extracted during the surveys. For Botswana, this is relevant given that the surveys are conducted months after birth so as to collect information on the three anthropometric indicators (wasting, underweight and stunting) of child nutrition. In the full sample, of the 54 infants whose birth weight information was not available although the interviewers saw their cards, only 35 were possibly born at home since assistance at delivery was from non-health professionals.

2.7 Conclusion

This chapter investigates the role of prenatal care utilization on infant birth weight in Botswana, following an increase in the proportion of low birth weight infants, which was among the highest in the WHO African region and upper-middle-income countries. Using two-stage regression, we account for potential endogeneity in prenatal care utilization, whose effect on infant health has been debatable. We also account for the possible sample selection bias since birth weight information is not available for all infants. We find that the age of the mother, mother`s incidence of being married, male infant and mother's height, are positively associated with infant birth weight. The effect of age declines as age increases, implying a decrease in the reproductive health stock for aging mothers. Maize meal price is negatively associated with birth weight, while the price for rice consistently carries a positive sign. However, given that we do not observe whether mothers (or their households) produced food, the effect of food prices would remain ambiguous. Thus, perhaps food prices are capturing some unobserved variation in districts or community characteristics.

The main result is that adequate prenatal care utilization is positively associated with birth weight and failure to account for its endogeneity reduces its effect. We derive two implications from these results. First, mothers, especially those who make the least use of prenatal care, should be encouraged to adequately utilize prenatal care to improve infant health. Secondly, it would appear that since adequate care utilization improves the birth weight, perhaps a possible policy issue for the observed deteriorating infant health relates to the quality of prenatal care. Hence, promotional efforts for adequate care utilization should be coupled with the need to reflect on the quality of care. Previous reviews of administrative data on maternal mortality suggest that there were issues of substandard care, including delayed referrals, intervention and

inappropriate management (MOH, nd; MOHW, nd). This pattern threatens the objective to "ensure access to quality antenatal care, and quality care during and after delivery to mothers and their babies" outlined in the country` essential health service package (Republic of Botswana, 2010b:13).

APPENDIX FOR CHAPTER 2

Table A2.1: Control function approach (Probit Reduced Form) results

	Full		Urban		Rural	
	(1)	(2)	(3)	(4)	(5)	(6)
APC	0.168**	0.183**	0.304*	0.307*	0.130**	0.146***
	(0.063)	(0.066)	(0.146)	(0.150)	(0.059)	(0.052)
Age	0.017**	0.017**	0.018*	0.018*	0.018*	0.017*
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Age squared	-0.000**	-0.000**	-0.000*	-0.000*	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.043**	0.042**	0.025	0.026	0.077***	0.074***
	(0.018)	(0.017)	(0.029)	(0.029)	(0.021)	(0.020)
Male	0.025***	0.025***	0.017	0.017	0.029**	0.028**
	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)	(0.013)
Height	0.003***	0.003***	0.003***	0.003***	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Pre-secondary	0.021	0.019	0.026	0.025	0.027*	0.022*
	(0.014)	(0.014)	(0.024)	(0.022)	(0.015)	(0.015)
Rice price	0.040***	0.040***	0.039	0.039	0.058***	0.061***
	(0.007)	(0.007)	(0.023)	(0.024)	(0.007)	(0.008)
Maize price	-0.014***	-0.013***	-0.003	-0.002	-0.029***	-0.027***
	(0.005)	(0.005)	(0.012)	(0.013)	(0.005)	(0.005)
Sorghum price	-0.005	-0.005	-0.007	-0.006	-0.002	-0.003
	(0.003)	(0.003)	(0.010)	(0.009)	(0.003)	(0.003)
QoB1	-0.018	-0.019	-0.006	-0.005	-0.037*	-0.039*
	(0.013)	(0.013)	(0.023)	(0.023)	(0.017)	(0.018)
QoB2	-0.033	-0.033	-0.023	-0.024	-0.037	-0.039*
	(0.021)	(0.021)	(0.032)	(0.032)	(0.024)	(0.024)
QoB4	-0.004	-0.005	0.027	0.026	-0.037*	-0.040*
	(0.014)	(0.014)	(0.023)	(0.022)	(0.020)	(0.020)
Other religion	0.004	0.005	0.067*	0.067*	-0.049	-0.050
	(0.027)	(0.027)	(0.037)	(0.038)	(0.055)	(0.055)
No religion	0.013	0.013	0.011	0.012	0.017	0.019
	(0.011)	(0.010)	(0.021)	(0.021)	(0.015)	(0.014)
Rural	0.007	0.005	-	-	-	-
	(0.016)	(0.017)				
GR	-0.079**	-0.103**	-0.156*	-0.164*	-0.057	-0.095**
	(0.036)	(0.045)	(0.083)	(0.091)	(0.037)	(0.037)
GR*APC	-	0.053	-	0.029	-	0.104
		(0.058)		(0.090)		(0.072)
Observations	1159	1159	656	656	503	503

Notes: Each column provides estimates from separate regressions. The first two columns are for the full sample while the rest are for urban and rural sample. Standard errors are in parenthesis. The omitted category is QoB3 for the quarter of birth and Christianity for religion. QoB: Quarter of birth, APC: Adequate prenatal care utilization, GR: generalized residual. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A2.2: Two-Step IV regression results

	Full (1)	Urban (2)	Rural (3)
APC	0.134*	0.343***	0.108*
	(0.064)	(0.076)	(0.065)
Age	0.017**	0.019*	0.017*
	(0.007)	(0.010)	(0.009)
Age squared	-0.000**	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)
Married	0.043***	0.024	0.076**
	(0.015)	(0.020)	(0.024)
Male	0.025**	0.017	0.028**
	(0.011)	(0.014)	(0.013)
Height	0.003***	0.004***	0.002**
	(0.001)	(0.001)	(0.001)
Pre-secondary	0.016	0.033***	0.022
	(0.015)	(0.011)	(0.013)
Rice price	0.042***	0.057***	0.059***
	(0.014)	(0.018)	(0.007)
Maize price	-0.013*	-0.013	-0.028***
	(0.008)	(0.009)	(0.005)
Sorghum price	-0.005	-0.008***	-0.002
	(0.005)	(0.002)	(0.003)
QoB1	-0.018	0.015	-0.038*
	(0.015)	(0.014)	(0.017)
QoB2	-0.036**	-0.028	-0.040
	(0.015)	(0.019)	(0.025)
QoB4	-0.002	0.044***	-0.040*
	(0.015)	(0.015)	(0.020)
Other religion	0.005	0.086*	-0.052
	(0.028)	(0.051)	(0.049)
No religion	0.012	0.007	0.015
	(0.015)	(0.031)	(0.016)
Rural	0.001	-	-
	(0.014)		
Observations	1159	656	503

Notes: Each column provides results from separate regression. The first column provides regression results for the full sample. The remaining columns are for urban and rural samples, respectively. Standard errors are in parenthesis. Omitted categories are QoB3 for the quarter of birth and Christianity for religion. APC: Adequate prenatal care utilization ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A2.3: Second Stage Regression results for proxy care quality

	Full	Urban	Rural
	(1)	(2)	(3)
APC*Injection	0.232***	0.229***	0.281***
	(0.046)	(0.061)	(0.038)
Age	0.013	0.007	0.015
	(0.008)	(0.011)	(0.012)
Age squared	-0.002*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Married	0.051**	0.040	0.078**
	(0.023)	(0.033)	(0.039)
Male	0.023*	0.018	0.023
	(0.012)	(0.016)	(0.018)
Height	0.003***	0.003***	0.002
	(0.001)	(0.001)	(0.001)
Pre-secondary	0.023	0.011	0.040
	(0.015)	(0.024)	(0.019)
Rice price	0.041***	0.011	0.055***
	(0.007)	(0.011)	(0.010)
Maize price	-0.018***	0.008	-0.037***
	(0.004)	(0.008)	(0.005)
Sorghum price	-0.007*	-0.006	-0.003
	(0.004)	(0.006)	(0.006)
QoB1	-0.018	-0.001	-0.026
	(0.013)	(0.023)	(0.017)
QoB2	-0.030	-0.026	-0.033
	(0.020)	(0.032)	(0.021)
QoB4	-0.003	0.026	-0.037*
	(0.013)	(0.022)	(0.020)
Other religion	0.011	0.076	-0.043
	(0.021)	(0.050)	(0.036)
No religion	0.020	0.017	0.032
	(0.013)	(0.019)	(0.028)
Rural	0.008	-	-
	(0.015)		
Observations	1159	656	503

Notes: Each column provides results from separate regression. The first column provides regression results for the full sample. The remaining columns are for urban and rural samples, respectively. Standard errors are in parenthesis. Omitted categories are QoB3 for the quarter of birth and Christianity for religion. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A2.4: Results for Sample Selection Model of infant birth weight availability

	Full		Urban		Rural	
	Coff.	SE	Coff.	SE	Coff.	SE
Birth weight						
APC	0.039***	0.011	0.032*	0.018	0.048***	0.019
Age	0.015*	0.008	0.014*	0.008	0.018**	0.009
Age squared	-0.000*	0.000	-0.000	0.000	-0.000**	0.000
Married	0.048***	0.015	0.026	0.026	0.070***	0.022
Male	0.029***	0.009	0.022*	0.013	0.030**	0.013
Height	0.003***	0.001	0.004***	0.001	0.002**	0.001
Pre-secondary	0.025*	0.013	-0.000	0.020	0.012	0.011
Rice price	0.028***	0.006	-0.007	0.014	0.062***	0.007
Maize price	-0.008***	0.003	-0.019***	0.006	-0.024***	0.003
Sorghum price	0.000	0.004	-0.003	0.003	-0.003	0.003
QoB1	-0.022	0.013	-0.001	0.022	-0.037**	0.017
QoB2	-0.037*	0.020	-0.029	0.032	-0.037	0.024
QoB4	-0.012	0.013	0.022	0.022	-0.037*	0.020
Other religion	-0.006	0.028	0.062*	0.035	-0.054	0.054
No religion	-0.001	0.012	0.020	0.018	0.016	0.014
Birth weight report						
Age	0.011	0.041	0.027	0.065	-0.024	0.053
Age squared	-0.000	0.001	-0.000	0.001	0.000	0.001
Married	-0.099	0.134	-0.035	0.135	-0.136	0.304
Male	0.015	0.054	-0.005	0.094	-0.022	0.104
Height	0.003	0.005	0.001	0.009	0.007	0.010
Pre-secondary	-0.238**	0.109	0.095	0.198	-0.485***	0.164
Other religion	-0.020	0.166	-0.144	0.157	0.220	0.228
No religion	0.159	0.156	0.701***	0.230	0.038	0.197
Proximity	0.431***	0.112	0.488***	0.100	0.518***	0.118
Care Facility	0.032***	0.010	0.026***	0.009	-0.003	0.016
Rural	-0.421***	0.099				
/athrho	-1.117***	0.146	-0.004	0.079	0.117	0.158

Notes: All standard errors are clustered at fixed effects. Omitted categories are QoB3 for the quarter of birth and Christianity for religion. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level. APC: Adequate Prenatal Care utilization. Coff: coefficient, SE: standard errors.

CHAPTER 3:HOUSEHOLD SIZE, BIRTH ORDER AND CHILD HEALTH IN BOTSWANA

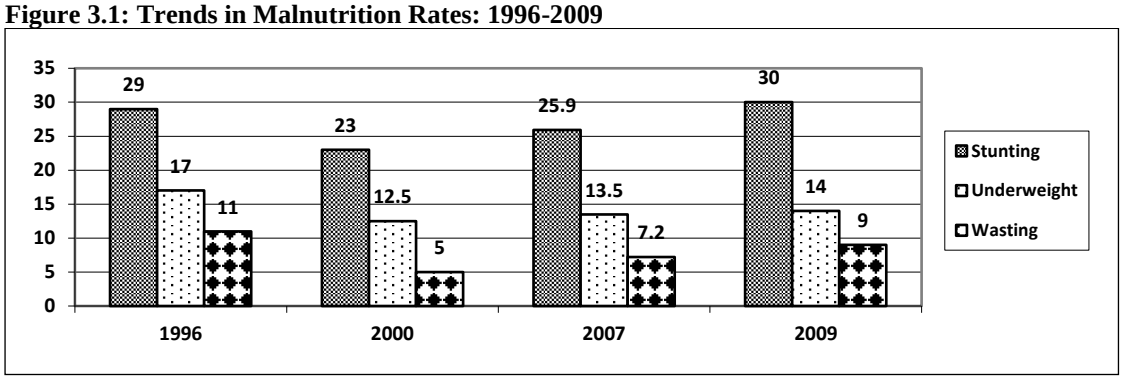
Abstract

One of the theoretical predictions relating to the family size and birth order effect on child capital is resource dilution hypothesis, according to which large sizes and high child`s birth order are likely to have negative effects. However, there are arguments that the assumption of a fixed and narrow flow of resources from parents underpinning the theory may not always hold. In Botswana, children aged 6-60 months are eligible for monthly food ration provided through the health care facilities. Notwithstanding this, child health as measured by the three anthropometric indicators: stunting (low height for age), being underweight (low weight for age), and wasting (low weight for height) deteriorated over time. This chapter, therefore, investigates the child birth order and household size effect on child health. Using the 2009/10 Botswana Core Welfare Indicator Survey (BCWIS) data, we estimate the random effects model to explore the within and between household effect. We find that children of high birth order are likely to fare worse than their lower birth order counterparts in nutrition. Household size is negatively associated with child health, and there are higher variances across than within households. Results further show that higher variances are unexplained by the observed characteristics.

3.1 Introduction

The effect of early life`s health on both subsequent health and life outcomes has been acknowledged in the literature (Black et al., 2007; Victora et al., 2008; Smith, 2009). Despite the acknowledgement that there is a need to address health problems at an early stage in Botswana (Gobotswang, 1993; Tharakan and Suchindran, 1999; Mahgoub et al., 2006; Nnyepi, 2007), statistics show continuing poor child nutrition (stunting[4], wasting and underweight), which is persistent (CSO, 2009a; World Bank, 2015). Figure 3.1 shows that from 2000, rates for stunting (low height for age), being underweight (low weight for age), and wasting (low weight for height) have increased consistently. This is despite the fact that all children aged 6-60 months are eligible for a monthly food ration at health care facilities through the government`s vulnerable group feeding programme (RHVP, 2011)[5].

Figure 3.1: Trends in Malnutrition Rates: 1996-2009



Source: CSO (2009a); World Bank (2015)

Not only have malnutrition rates increased, but inequality (higher malnutrition rates among children from poor households than from well-off households) has been evident as well (Statistics Botswana and UNICEF, 2008). For instance, in 2007, severe (moderate) stunting rates stood at 19 (17) and 11 (12) per cent for lowest and highest quintiles respectively. Similarly, the proportion of underweight children among the poorest households was 16%, compared to 4% among wealthier households (Statistics Botswana and UNICEF, 2008). From a policy viewpoint, the pattern depicted in Figure 3.1 is worrying, given that there seems to be

[4] Of the three indicators, stunting is mostly preferred as it indicates both the past and present nutritional deficiencies (Charmarbagwala et al., 2004)

[5] This could reflect the possibility that the ration is redistributed among other household members and therefore reducing gains by the targeted recipients. A thorough evaluation would need to be done to ascertain this, especially that data on underweight produced by the Ministry of Health and Wellness suggests a downward trend, which has been attributed to the ration programme (BIDPA, 2010).

a reverse of improvements achieved during the 1990s. In this chapter, we set out to examine factors influencing (deteriorating) child health.

Theoretical predictions regarding the effect of household size and child birth order on child outcomes have been provided through the resource dilution hypothesis, hygiene hypothesis and biologically based arguments (De Keyser and Rossem, 2017; Lundborg et al., 2015; Horton, 1988; Blake, 1981). According to the resource dilution hypothesis, large household size and high birth order children imply high costs in terms of caring for children, reducing the parents` likelihood of evenly distributing resources and potentially concentrating resources in those born first (Horton, 1988). Consequently, in a family context, this leads to a trade-off between the number of children and their quality measured by their human capital (either education or health); a relationship formally expressed in the quantity-quality model (Becker, 1960; Becker and Lewis, 1973; Becker and Tomes, 1976). This model implies that a negative relationship is to be expected between the number of children and their quality. The hygiene hypothesis postulates that as the number of children increases in the household, the likelihood of health problems, such as infections, increases as well, which negatively affects child nutrition (Lundborg et al., 2015; Horton, 1988). With regards to the biologically based argument, birth order matters because maternal health deteriorates with time: aging mothers are more likely to produce additional children of poorer health (Horton, 1988). However, a counter-argument is that these children may benefit from their mother`s child care experience (Dasgupta and Solomon, 2018).

To investigate the deterioration in child health (Figure 3.1), we investigate household size and birth order effects on child nutrition. We ask two related questions: could deteriorating child health reflect the effect of household size as theoretically predicted: and how do different birth order children fare in terms of nutrition? Answering these questions is essential because of the following. Firstly, the dilution hypothesis assumes that resources only come from the parents, which has led to a debate in the literature (Desai, 1995; Blake, 1981). However, in the case of Botswana, this may not apply as many children are in the care of extended families so may not rely solely on parental resources. Secondly, although much research on child human capital in both developed and developing countries has focused on education as a measure of quality, evidence on the resource dilution hypothesis is inconclusive (Dasgupta and Solomon, 2018; Zhong, 2017; Lawson and Mace, 2008; Lu and Treiman, 2008; Price, 2008; Desai, 1995; Horton, 1988; 1986). Some (De Haan, 2010; Black et al., 2005) find that consideration of birth

order attenuates the negative effect of family size on child human capital, although Mogstad and Wiswall (2016) argue that this is because of linearity assumption imposed on family size.

While this inconclusiveness could be due to several factors including the nature of data (i.e. cross sectional versus longitudinal), the most cited explanation is that the assumption of a fixed and narrow flow of resources from parents underpinning the theory such as the quantity-quality model, does not hold (Desai, 1995; Blake, 1981)[6]. Indeed, some note that households go through a development cycle and that buffering, either at household or national level, may reduce the predicted negative effect of family size (Bras et al., 2010). For instance, from demographic survey data for Latin America and the Caribbean (Bolivia, Brazil, Colombia, Dominic Republic, Guatemala and Trinidad & Tobago), Asia and North Africa (Egypt, Morocco, Sri Lanka and Thailand) and SSA (Burundi, Ghana, Mali, Senegal, Zimbabwe) countries collected during 1986-1990, Desai (1995) finds that the negative effect of family size is lessened by government assistance. However, the effect is not evident in all the considered countries. In their meta-analysis of studies that consider factors influencing child health, Charmarbagwala et al. (2004) find that household size is positively associated with child nutrition, even though the effect was insignificant in some instances. In their view, failure to account for an alternative household structure could be seen to "overstate" the negative association between household and child nutrition.

Lu and Treiman (2008) find that in China when schooling opportunities were limited and expensive, children (especially girls) from large families obtained less schooling. However, schooling expansion and affordability reduced the negative effect of having many siblings. They note that demographic, socio-economic and political factors external to the household matter in both resource availability and their intra-household allocation. Shavit and Pierce (1991) find that although educational attainment is negatively affected by nuclear family size, this is not evident among extended families in Israel. Therefore, not only do we contribute to this existing debate, but highlight issues for policies and programmes to address child malnutrition.

[6] Behrman (1997) classifies this approach to fall into a category she refers to as "consensus parental preferences model of intrahousehold allocation."

In Botswana, households obtain resources, including income and food from a variety of sources, including government programmes, which may reduce children`s reliance on their biological parents. Moreover, unmarried fathers (and mothers to a certain extent) are unlikely to reside with their children, implying that children may not always depend solely on resources from their biological parents. The effect of alternative family structures on child nutrition has not been explored explicitly in the literature. Allowing for the effect of resources sourced from outside the "nuclear family" relates to the anthropological literature, on the relevant definition (or concept) of the African household in assessing child welfare (see Randall and Coast, 2015 and studies cited therein). This literature argues that a definition of a household as recommended by the United Nations (2008) misses both the social and economic interaction of members residing elsewhere (Bongaarts, 2001). Furthermore, while previous work in Botswana has considered household factors, such as house type and household head`s gender, they have not controlled for (or explored) the possible unobserved heterogeneity within households that might explain inequality in child nutrition. We address this gap.

Our work relates specifically to Botswana`s national population policy, which has a goal of "improved quality of life and standard of living" (MFDP and UNFPA, 2010:2). The policy recognises the need to reduce child malnutrition, which has to be informed by an understanding of influential factors that affect malnutrition. Statistics from various censuses show that population growth rate declined from 4.7 in 1981 to 3.5, 2.4 and 1.9 per cent in 1991, 2001 and 2011 respectively. Moreover, the average household size declined from 4.8 in 1991 to 4.15 in 2001 and further to 3.5 in 2011 (Statistics Botswana, 2014a). The expectation would be that these declines together with those in fertility rates would be coupled with improved child nutritional status, especially in the presence of the government`s food ration programme and the publicly provided health care. The main results are that household size and birth order are negatively associated with child nutrition. We also find that there is higher variation in child nutrition across than within households and that a larger share of the total variation is accounted for by the unexplained component. Given the existence of the food ration programme, results could imply that effectiveness of the programme may be threatened. Moreover, the effect of intrahousehold allocation may not be underestimated given that a larger share of variation is unexplained. Results are also important considering the revised population policy. The chapter proceeds as follows: Section 3.2 reviews related literature, while 3.3 presents data and summary statistics. The empirical strategy is presented in 3.4; results and robustness checks in 3.5 and 3.6, respectively. Section 3.7 concludes.

3.2 Review of related literature

On the basis of the theoretical predictions highlighted above, there have been efforts to empirically investigate such relationships, both in the developed and developing world. However, it is crucial to recognize the approach in defining a family and a household, especially that some (Black et al., 2005; De Haan, 2010) have noted the relationship between child birth order and family size, as high birth order children within a given family will be from larger sized families. In the Botswana context, where many households comprise children from different families (different mothers), the link between birth order and household/family size is weakened. This is also important as it has a bearing on the framework within which an investigation is carried out, such as sources of resource flows as well as the number of people being considered as diluting resources.

According to Burch (1979:174), "family refers quite generally to a group of kin. i.e. person related by blood, marriage or adoption". On the other hand, "a household is usually defined as a group of persons who make provision for food, shelter and other essentials for living" Bongaarts (2001:263-4). In the context of the latter, which is usually used in national surveys and population censuses, United Nations (2008) states that a household may constitute more than one family, may be made of only one person (whereas family must have at least two), and that in a multi-person household, members do not necessarily have to be related. Therefore, a family is seen as a part of the household, and this has implications on both the availability and distribution of resources for the production of child health. While these two terms have been used interchangeably in the literature, we consider a household as defined above by Bongaarts (2001). In the following section, we provide empirical evidence regarding the effect of household (and family) size and birth order on child health.

3.2.1 Household Size and Child Health in Developed Countries

Dasgupta and Solomon (2018) find that family size has no effect on child Body Mass Index in the US, although Price (2008) earlier established (for the US) that the amount of time spent with children decreases with increasing child birth order, predicting the negative effect of family size on child outcomes. Lundborg et al. (2015) find a positive effect of family size on male individual health in families with first born males and 2 or more births and in those with first and second born males and 3 or more births in Sweden. Lawson and Mace (2008) find a negative effect of family size on child height in Britain. What is commonly acknowledged in this literature is that family size is treated as an endogenous variable resulting in the use of

instrumental variable strategy and other approaches such as multi-level models; an approach argued to be able to account for heterogeneity by considering both the context (the environment) and compositional effects in health research (Duncan et al., 1998).

3.2.2 Household Size and Child Health in Developing Countries

Horton (1986) uses cross-sectional data from the Bicol region in the Philippines, to jointly estimate the determinants of fertility (children ever born) and child nutrition (z-score for height for age). She finds that the number of older siblings (brothers and sisters) is negatively associated with the child`s nutrition. However, given that the number of younger siblings had a mixed effect (negative and insignificant effect for younger brothers and significant positive effect for younger sisters), it appears this captured the issue of child`s birth order rather than family size. Indeed, the author makes reference to the effects of birth order in explaining this, concluding that "there are strong effects of birth order on child nutritional status" (pp.174). Using data from China Nutrition and Health survey to estimate urban and rural sample regressions separately, and one-child policy as a source of family size exogenous variation, Zhong (2017) finds that family size has a negative effect on child height for age but not on education. However, separate consideration of urban and rural samples reveals that the negative effect is only evident in rural areas. The author tests explicitly for the resource dilution hypothesis by further dividing the rural sample into three (low, medium and high) income per capita groups and still finds a negative effect of family size on child health.

Mussa (2015) uses the 2006 multiple indicator cluster survey data to investigate intra and inter-household child nutrition inequality in Malawi. Having estimated rural, urban and national regressions, the author finds evidence of intra-household gender differences against boys and a positive association between a child`s nutritional status and the number of younger siblings, suggesting that the argument for resource dilution may not hold in this case. Kabagenyi and Rutaremwa (2013) find that, although large households had a higher infant and child mortality in Uganda, there was no statistical association between them. Through the logistic regression model, they find that children from medium (4-6) sized households are more likely to die than those from small (1-3) sized households. Given that their descriptive statistics show larger proportions of nuclear households in the medium and large sizes than extended households, results could suggest that there is more likelihood of child mortality among nuclear than extended households. In fact, this is confirmed by regression results, although the effect is significant at 10 per cent. Among other variables, Kabubo-Mariara et al. (2008) consider

household size as a determinant of child nutrition in Kenya. Using a pooled sample for the 1998 and 2003 health and demographic surveys, they find a negative association between household size and child height for age.

Ajao et al. (2010) assess the influence of family size and household food security status on child nutrition in Nigeria. Through the binary logistic regression for cross-sectional survey data for mothers of children aged below five, they find that although household food insecurity increases the likelihood of child malnutrition, there is no evidence on the association between family size and malnutrition. On the other hand, Millimet and Wang (2011) find a mixed effect in Indonesia. Through the instrumental variable approach, they find that moving from a two-child family to the one with more than two has no effect on a child`s height for age and body mass index for age. On the other hand, through the estimation of quantile treatment effects, they find that accounting for endogeneity of family size decision results in significant differences only in the distribution of height for age, hence their conclusion that this quantity-quality trade-off is for some quantiles of the distribution.

3.2.3 Birth Order and Child Health in Developed Countries

Lundberg and Svaleryd (2016) investigate the effect of birth order on child health in Sweden using registry data and measuring child health through several indicators. For children aged 0-6 years, they find that compared to first-borns, only the second-borns are less likely to be hospitalised. Moreover, in terms of diagnoses, children of birth order 2, 3 and 4 are less likely to have perinatal and congenital malfunctions. They also find that there is more likelihood of mortality among first born children, a pattern which they say was earlier observed. However, other health indicators (respiratory & eye/ear problems, injuries) are less likely to be evidenced among first born children compared to the second, third and fourth born. They also provide the mechanism through which the effect of birth order on child health manifests. Among them is that mothers` endogenous fertility decision varies with the mortality of the first child. They demonstrate that controlling for this reduces the magnitude of the birth order effects.

Brenøe and Molitor (2018) examine this issue in Denmark, with standardised birth weight (adjusted for gestation period) and Apgar score as child health measures. Through the family fixed effect estimation, they find that second, third and fourth-born children have better health at birth compared to first-borns. Furthermore, these high birth orders reduce the likelihood of both prematurity and the incidence of low birth weight children. The authors demonstrate that

the positive effect of birth order on child health was not due to a change in maternal behaviour. They observe a reduction in women working and the likelihood of being a student as well as decreasing prenatal care use with increasing birth order. By implication, it could not be argued that mothers cared less for their first birth order children.

3.2.4 Birth Order and Child Health in Developing Countries

A different pattern on the effect of birth order effect on child health emerges in the developing world. Outcomes from empirical investigation generally reveal a negative effect of birth order on child health (Jayachandran and Pande, 2017; Howell et al., 2016; Rahman, 2016; Kebede, 2005; Horton, 1988). From a comparison of Indian and Sub Saharan African children, Jayachandran and Pande (2017) find that although Indians had higher height at birth, the difference decreases for children above second birth order. According to them, this reflects parent intra-household allocation decisions, including eldest son preference. Thus, if the first born is a son, there is a higher likelihood to not invest in subsequent children, or that if not, earlier born are disadvantaged by the hope of a son. In fact, Jayachandran and Kuziemko (2011) find that this child gender preference results in girls being breastfed for a shorter duration than boys. In addition to confirming this, Barcellos et al. (2014) further find that in rural Indian households, child care time for an infant boy was relatively more than that for an infant girl. Rubalcava and Contreras (2000) also find evidence of child preference where mothers channel resources towards their daughters and fathers do to their sons in Chile. For some Africa countries, Sahn and Stifel (2002) find that mothers prefer their daughters while fathers prefer their sons. They arrive at this conclusion after investigating the effect of mothers` and fathers` education on nutritional status for girls and boys respectively.

In Bangladesh, Rahman (2016) finds that the likelihood of being stunted among children aged 36 months increases with high birth order. The author`s analysis of the 2011 demographic and health survey reveals that compared to the first borns, children of birth order 3 and above are more than 24% likely to be stunted. The author partly attributes this to two possibilities: first, later born children are more likely to be unwanted, which affects the way they are being cared for. The author cites one study, which established the decreasing likelihood of both ante and postnatal care utilization for children of high birth order. While results also show that child unwantedness increases the likelihood of stunting, it remains unclear for how long the effect of this lasts, especially that results show that stunting is more likely among children aged 12-23 months compared to those with less than 12 months. Secondly, the effect could reflect the

theoretical predictions of resource dilution, through which later-born children are disadvantaged.

Horton (1988) considers the effect of birth order on nutritional status for children aged 15 years and below in the Philippines, through pooled, and family fixed effect estimation, which also allowed for interaction of birth order with some other variables. Results show that birth order is negatively associated with child`s height for age but not weight for height. Through the interaction variables, fixed effect estimation results show that although mothers` education and household assets could reduce the negative effect of birth order, the effect is statistically insignificant. Moreover, while birth order interaction with the number of children has a positive and significant effect on child height for age, the author observes that the magnitude of the effect is less than the negative effect of birth order on child height for age. However, the birth order number of children interaction has a negative effect on child weight for height. Further, from pooled analysis, interacting mother`s education reduces the negative effect of birth order. The overall implication is that children born later are more likely to be stunted than those born earlier, which the author interprets to be suggestive of the effect of resource allocation over time. In particular, the non-significance of birth order on weight for height implies parents` ability to equitably allocate resources in the short run, and their inability to do so in the long run as evidenced in the significant birth order effect on child height for age.

Howell et al. (2016) consider 18 African countries (Botswana excluded) and find that high birth order is positively associated with the incidence of mortality and stunting. They demonstrate this effect for children aged 1-4 and those aged 5-14. They further show that among children aged 12-23, for instance, the incidence of consuming fewer food groups reduced the likelihood of diversified diet and increases with increasing birth order. However, the non-consideration of infants on the basis that the focus is on "potential competition for food and other resources" (pp4), could be argued to have understated the effect of time as an input in child health. It could be the case that spending more time with infants, may compromise the health for those aged 1-14 years, although the counter-argument may be that those nearing 14 years may fend for themselves (Desai, 1995).

Kebede (2005) considers the effect of birth order on child`s (aged ten years and below) height for age from a four-round (panel) of surveys conducted in 1994, 1995 and 1997 in Ethiopia. Compared to children of sixth or more birth order (reference category), children in the first

three birth positions yield significantly higher height for age. No significant differences are evidenced for those in the fourth and fifth positions. The author also finds that height for age declines with the increasing birth position. The author attributes this to biological factors and parental behaviour (i.e possibilities of elder child preference), with an alternative explanation of resource dilution.

Using the demographic and health survey data from 14 African countries, Sahn and Stifel (2002) find that among children aged 3-36 months, birth order is positively associated with child height for age and negatively with child weight for height. Compared to first born children, second and third borns have a higher height for age while fifth and higher birth order have no effect. On the other hand, they also find that later born children are more likely to be wasted. In Kenya, Kabubo-Mariara et al. (2008) find that birth order has no significant effect on child height for age, although the coefficient has a negative sign. However, they find an increased likelihood (through binary regression) of stunting with increasing birth order. Yassin (2000) finds that although high birth order increased the likelihood for child mortality in bivariate analysis, no effect is evidenced through the multivariate analysis in rural Egypt.

It emerges from the existing literature that there are differences in child birth order and household size effects on child health. While some find no and positive effect (Mussa, 2015; Ajao et al., 2010; Charmarbagwala et al., 2004; Yassin, 2000) some find a negative effect (Zhong, 2017; Kabubo-Mariara et al.,2008; Horton, 1986). However, there are arguments that evident negative effect between household size and child health may be due to selection on the structure of the household (Charmarbagwala et al., 2004). Moreover, as we stated at the outset, this assumes a narrow and fixed flow of resources (Desai, 1995; Blake, 1981). With regards to the birth order effects, the pattern generally observed in developing countries is the opposite of evidence in the developed world: high birth order children are likely to fare worse in developing than in developed countries. A similar pattern is observed by De Haan et al. (2014), for studies that consider education. Even on the mechanism that could explain the birth order effect, one study in the US (Price, 2008) find that the incidence of mother's quality time accorded to the child decreases with increasing birth order. On the contrary, the opposite is evidenced in Ecuador (De Haan et al., 2014) and that boys in India have more care time than girls (Barcellos et al. 2014). Moreover, while generally, birth order has a negative effect, there are indications of some positive effect and no effect on child health as well, suggesting that the effects may not be conclusive. Hence our study contributes to this debate.

3.3 Data and Summary Statistics

We use the 2009/10 Botswana Core Welfare Indicator Survey (BCWIS) data, collected by Statistics Botswana during the period from April 2009 to March 2010. This national survey is an improvement of the previous Household Income and Expenditure Surveys as it covers other non-income welfare indicators such as child nutrition which we consider in this chapter. The survey used a two-stage stratified probability sampling, with the first and second stages being the selection of primary sampling units (Enumeration areas) and occupied households, respectively. A total of 7,732 households with 27,308 individuals were covered in the three strata; cities/towns, urban villages and rural areas. The survey sourced information on socio-economic and demographic characteristics of households and their members. For all births that mothers had, they were asked to provide child age, sex and birth order information. The key piece of information that we use in this chapter is knowledge about the biological parents, namely if they were still alive at the date of interview, lived in the same household with their children and if not, whether children received support from them.

From the individual data file, we extracted data for children aged five years and below[7], and merged it with the household data file, resulting in a sample of 3,873 children. To measure child nutrition, we computed the standardized child nutritional indicators through the World Health Organisation (WHO) software for growth and development of the world`s child (WHO, 2011b). This software requires information on child age (in months), sex, height and weight. It then transforms height for age, weight for age and weight for height into z scores, as standardized measures of their deviation from the median of the reference population. Specifically, the approach takes the difference between the individual child`s observation and median reference value and divides it by the standard deviation of the reference population (i.e children of the same age). The composition of the reference population changed over time, from the US National Center for Health Statistics (NCHS) data to the current reference population, which is an outcome of a WHO Multi Growth Reference Study (MGRS) of six countries: Brazil, India, Norway, Oman and the USA (WHO, 2011b). The interpretation is such that a z-score in the range between -2 and +2 standard deviation from the median of the reference population implies a normal growth; otherwise, the child is considered malnourished.

[7] The sample is restricted to children aged five years and below as this is the only group of children for whom anthropometric indicators are monitored (through surveys) over time. Additionally, government through the Ministry of Health and Wellness, monitors child underweight until the child is aged 5 years.

Z scores for weight for height higher than 2 standard deviations from the reference median indicate an increased likelihood of being overweight.

However, for each nutritional indicator, the total sample size varies according to availability. There are cases where a child`s weight is available while height is not, and reasons for the missing information are not provided in the data. Furthermore, some children have standardized z scores outside the WHO`s recommended range. According to WHO (2011b), a range for weight for age, height for age and weight for height is between -6 & 5, -6 & 6 and -5 & 5 respectively, hence any score outside these ranges should be excluded from the analysis. Overall, we are able to analyse a sample range of 2863-3481 children depending on the outcome measure. For robustness checks, we also consider the 2007/08 Botswana Family Health Survey (BFHS) data, also collected by Statistics Botswana. This covers 2532 children aged five years and below. The variables informed by a literature review (although limited by data) are defined in Table 3.1 below. It is important to note the variation in the variable availability between these two data sets. BCWIS, for instance, did not collect information on child breastfeeding while BFHS does not have a birth order variable. We provide definitions for all variables found in both surveys and which variables are available.

Table 3.1: Definitions of variables

Variable	Definition	BFHS	BCWIS
Outcome variables			
WAZ	Standardized z scores for weight for age	√	√
HAZ	Standardized z scores for height for age	√	√
WHZ	Standardized z scores for weight for height	√	√
Child Characteristics			
BO1	An indicator for a first birth order child		√
BO2	An indicator for a second order child		√
BO3	An indicator for a third birth order child		√
BO4+	An indicator for a fourth or higher birth order child		√
MaleC	An indicator of a male child, 1=Yes, 0 otherwise	√	√
Breastfed	Child ever breastfed (1=Yes)	√	
AGEC	Child age in months	√	√
Diarrhoea	Child had diarrhoea two weeks before the survey, 1=Yes	√	√
MOTLIVEH	Child`s mother lives with the child in the same household, 1=Yes	√	√
FATLIVEH	Child`s father lives with the child in the same household, 1= Yes	√	√
FatContribute	Child`s father contributes to his/her support on regular basis, 1=Yes		√
Grand Child	Child is a grandchild to the household head	√	√
Son/Daughter	Child is a son/daughter to the household head	√	√
Other relative	Child has other relation to the household head	√	√
Mother`s Characteristics			
Mother`s age	Mother/carer`s age (years)	√	√
Mother Married	Mother is married	√	√
Pre-Secondary	Mother has pre-secondary education	√	√
Junior Secondary education	Mother has junior secondary education	√	√
Post-JSecondary education	Mother has post-junior secondary education	√	√
Household Characteristics			
MaleHH	An indicator of a male household head, 1=Yes		√
AgeHH	Age of the household head in completed years		√
MarriedHH	Whether the household head is married, 1=Yes		√
EDUHH	Education of the household`s head (Years of schooling)		√
HHSZ	Number of household members	√	√
Children<1 year	Number of children aged less than 1 year		√
Children 1-3 years	Number of children aged 1-3 years		√
Children 4-5 years	Number of children aged 4-5 years		√
Piped Water	Household has piped water in yard (indoors or out) (1=Yes)		
Communal tap	Household water sourced from communal tap (1=Yes)	√	√
Other	Household water sourced from other (1=Yes)	√	√
Own Flush toilet	Household owns a flush toilet (1 =Yes)	√	√
Own Pit latrine	Household owns a pit latrine (1 =Yes)	√	√
Communal flush toilet	Household uses communal flush toilet (1 =Yes)	√	√
Communal pit latrine	Household uses communal pit latrine (1 =Yes)	√	√
Communal VIP	Household uses communal VIP (1 =Yes)	√	√
Neighbour's toilet	Household uses neighbour`s toilet (1 =Yes)	√	√
Rural	Rural area residence	√	√

Summary Statistics

Before presenting the summary statistics, we show the within a household structure by gender of the household head (Figure 3.2). There are higher frequencies for head`s son/daughter within male-headed households than in female-headed households. However, the opposite is evident for grandchildren and other relationships. Moreover, although there is a higher likelihood for fathers (mothers) to reside with their children in male (female) headed households respectively, the pattern provides a basis for parent-child resource flow. Thus, not all parents are heads of households and the lower shares for mother-child resident in male-headed households, and

father-child resident in female-headed households suggest that resources cannot be assumed to be solely from biological parents.

Another factor relating to the household structure is on the child age-birth order link. In a typical (nuclear) family, the expectation is that high birth order children should be younger. However, in the case of household (in the African context), there are possibilities of children of the same age with different birth order or children of the same birth order with different child age. We show these, but only for households with at least two children (n=2292). In such households, a total of about 37% of children have the same birth order, but different ages, while 2% have the same age but different birth order. There are also cases where a lower birth order child in a specific household is younger than a different high birth order child. All these are due to the household nature; different mothers in the household which as argued, affect both the source and distribution of resources for child`s care.

Table 3.2 shows summary statistics. The average height for age Z score of -0.928 implies that the children in this sample are about 0.9 standard deviations below the median of the reference population. First birth position (BO1) accounts for almost half of the children (41%), followed by second and fourth or higher birth order with shares of 27 and 18 per cent respectively. An examination of nutritional indicators by birth order reveals that generally, nutrition worsens with increasing birth order (Table 3.3). Approximately 49 per cent of children are males, and 10 per cent report having had an incidence of diarrhoea two weeks preceding the survey. Although biological parents are alive for more than 95 per cent of children, mothers are more likely than fathers to be staying with them in the household. The pattern might partly be explained by the low rates accounted for by the married individuals in the country, which has been attributed to delayed time to marry due to cohabitation (Mokomane, 2005). Traditionally, in cases of unmarried parents, there is a higher likelihood of children staying with their mothers than their fathers. About 67 per cent of children not living with their fathers do receive regular contributions from their fathers.

Considering the relationship of the children to the household head, the majority, 49 per cent, are grandchildren and 39 per cent are sons/daughters. Other relation (nephews/nieces, step children) account for 12 per cent. An average of 44 per cent of children are from male-headed households while average household head`s age is 49 years. Table 3.2 also shows that around 34 per cent of children are from households whose heads are married. They are followed by

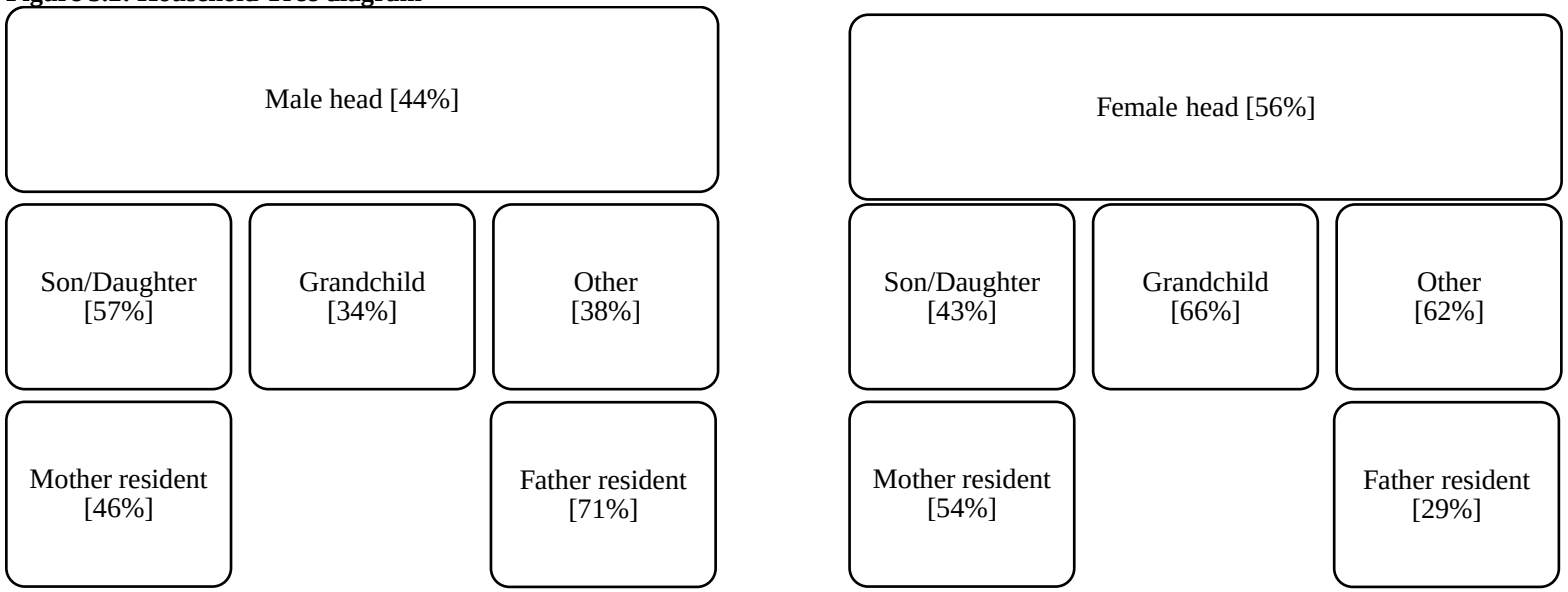
those from households whose heads are living together (24%), never married (23%) and widowed (15%) respectively. The lowest share (1%) is accounted for by those whose heads are separated, while the divorced average about 2 per cent.

On average, household heads have about 5 years of schooling. For biological parents, demographic information is only available for those who were residing with their children. Only 37 per cent of children resided with their biological fathers, implying that fathers' information is not available for the majority of children (non- residing fathers did not provide information). For the sample of children residing with their mothers we find that about 40 per cent are for mothers with junior secondary education, followed by those with post junior secondary and primary education (Table 3.4). Regarding the marital status, a larger share (46%) are never married, 32% living together[8], 20% married and 10% each for separated and widowed. Furthermore, no information on parents` ethnicity was sourced during the survey.

The average size of the household is 6 (Table 3.2). A further investigation into the household structure is through the categorization of the number of children by age. As shown, the majority of children are from households with children aged 1-3 years. About 51% of children are residing in rural areas; 34% and 16% were in urban areas and cities/towns respectively. Table B3.1 in the appendices presents summary statistics from BFHS and shows that on average, the two survey data have a similar pattern.

[8] We do not observe whether the male partners are biological fathers to the children

Figure 3.2: Household Tree diagram



Source: author`s computation from BCWIS data

Table 3.2: Summary Statistics

Variable	Mean	Standard Dev
Dependent		
WAZ	-0.179	2.866
HAZ	-0.928	2.003
WHZ	0.772	4.859
Independent		
Child Characteristics		
BO1	0.408	0.492
BO2	0.274	0.446
BO3	0.135	0.342
BO4+	0.178	0.382
MaleC	0.491	0.500
AGEC	32.850	18.833
Diarrhoea	0.096	0.295
MotliveH	0.784	0.412
FatliveH	0.374	0.484
Fat contributes	0.665	0.472
Grand Child	0.486	0.500
Son/Daughter	0.388	0.487
Other relative	0.116	0.320
Household Characteristics		
MaleHH	0.438	0.496
AgeHH	49.183	16.321
MarriedHH	0.342	0.474
EDUHH	5.151	5.174
HHSZ	6.975	3.473
Children <1 year	0.335	0.541
Children 1-3 years	1.045	0.883
Children 4-5 years	0.850	0.841
Rural	0.506	0.500

Source: author`s computation from BCWIS data

Notes: WAZ: Weight for age z score, HAZ: Height for age z score, WHZ: Weight for height z score. BO: Birth order.

Table 3.3: Nutritional Indicators Summary Statistics by Birth order

Birth Order	WAZ	Underweight	HAZ	Stunted	WHZ	Wasted
1	-0.336	0.123	-0.865	0.268	0.106	0.108
2	-0.391	0.143	-1.008	0.295	0.130	0.109
3	-0.437	0.176	-0.842	0.251	-0.090	0.152
4+	-0.601	0.159	-1.161	0.336	-0.024	0.122

Source: author`s computation from BCWIS data

Table 3.4: Summary Statistics for Child with Mothers in Household

Variable	Mean	Std. Dev.
BO1	0.347	0.476
BO2	0.270	0.444
BO3	0.164	0.370
BO4+	0.213	0.410
MaleHH	0.458	0.498
AgeMother	29.695	7.236
Pre-Secondary	0.279	0.448
Junior Secondary	0.395	0.489
Post Junior Secondary	0.309	0.462
Mother Married	0.199	0.399
Mother Living together	0.321	0.467
Mother Separated	0.010	0.099
Mother Divorced	0.004	0.067
Mother Widow	0.010	0.098
Mother Never married	0.455	0.498
Son/Daughter	0.498	0.500
Grand Child	0.387	0.487
Other Relation	0.107	0.309

Source: author`s computation from BCWIS data
Notes: BO: Birth order

3.4 Model Specification

3.4.1 Theoretical Model

Studies on production of human capital (either health or education) at the household level are usually based on the Becker (1965) model, which incorporates non-market dimensions into household production (Strauss and Thomas, 1995). Time and goods are allocated for the production of commodities, some of which are sold in the market, while some are consumed at home, partly because no market exists for them. In this framework, it is assumed that households have a utility function that represents their preferences and determined by the consumption of commodities X and leisure L, formalised in equation 3.1 below:

$$U = U(X, L) \tag{3.1}$$

The household optimally chooses consumption bundle subject to both budget and time constraints. According to the budget constraint, total market consumption cannot be more than the total income, which includes both non-labour and labour earnings. On the other hand, total time includes leisure, time for home production and time for work in the market. Although the initial assumption on perfect substitutability between market and home-produced goods was maintained in the model for a single person household (Gronau, 1977), Strauss and Thomas (1995) state that this assumption can be relaxed on the basis that most of the human capital

cannot be purchased in the market. In the context of child health, such modification considers the possible integration of biological, demographic and socio-economic factors. Therefore equation 3.1 gets modified to include child health, which in our case is measured by the anthropometric indicators for those aged 5 years and below. In this case, it is assumed that the production of child health is done on its own right and directly enters the household utility function.

3.4.2 Empirical Model

Following the approach in the literature (Strauss and Thomas, 1995), our basic model is specified as follows:

$$N_{ch} = \beta_0 + \beta_1 X_{ch} + \beta_2 Y_{ch} + \varepsilon_{ch} \quad (3.2)$$

In (3.2) N_{ch} is child nutritional status as measured by the standardized z scores of weight for age, height for age and weight for height for child *c* residing in household *h*; X_{ch} represents specific child characteristics while Y_{ch} is a vector for household characteristics. ε_{ch} is the error term. We also estimate binary regressions for which the dependent variable is equal to 1 if the child is malnourished (z score< -2) and 0 otherwise. The binary regression is of the form:

$$Prob(Ma = 1|x) = \beta_0 + \beta_1 X_{ch} + \beta_2 Y_{ch} + \varepsilon_{ch} > 0 \quad (3.3)$$

or in index form:

$$Prob(Ma = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}) \quad (3.4)$$

Where Ma represents malnutrition (z score< -2), **x** and *Y* are vectors of explanatory variables, G is a cumulative distribution function. According to equation (3.4), the probability of child being malnourished is determined by a vector of the considered explanatory variables. We specifically estimate the Probit model, in which case

$$G(z)=\Phi(z)=\int_{-\infty}^{z} \phi(v)dv \quad (3.5)$$

where ø(z) is the standard normal density expressed below:

$$\phi(Z)= (2\Pi)^{-1/2}\exp(-Z^2/2) \quad (3.6)$$

This binary equation is estimated using maximum likelihood estimation. The log-likelihood function for child *i* is as follows:

$$l_i(\boldsymbol{\beta})Ma_{chi} = Ma_{chi}log[G(\mathbf{x_i}\boldsymbol{\beta})] + (1 - Ma_{chi})log[1 - G(\mathrm{x_i}\boldsymbol{\beta})] \quad (3.7)$$

It also follows that the log likelihood for all children is given by a sum of the individual child`s likelihood (Wooldridge, 2010). As our interest is on the impact of the explanatory variables on the probability of malnutrition, we derive the marginal effect.

The literature suggests issues to address in the empirical investigation. One is that family completeness may introduce bias on the effect of birth order (Horton, 1988): a certain birth order for a family whose size is yet to increase may be inappropriately considered as the last, while in fact, that could be a middle one. Horton (1988) notes that the inclusion of mothers' age could partly address this issue, though imperfectly. We do not necessarily compare the first with last, but rather investigate the effect of different birth orders on child nutrition. Thus, we do not assume that the number of children is complete.

Another issue is that of endogeneity of household size. Previous work that considered family size used the instrumental variable approach (De Haan, 2010; Black et al., 2005, Dasgupta and Solomon, 2018) while others used household fixed effect model (Horton, 1988). There seems to be a consensus that household size is influenced by several factors such as fertility, mortality and migration (Hoddinott and Mekasha, 2017; Bongaarts, 2001), which are also determined by other factors. For instance, fertility is determined by the female`s ability to give birth (fecundity), incidence of marriage, as well as female labour force participation; a contributing factor for the declining Total Fertility Rates in Botswana (Mills et al., 2010). Although marriage, in the context of nuclear households, has been viewed as exposure to childbirth (Bongaarts, 1982), in Botswana, there are incidences of pre-marital births. Data shows that during the period 2001-2011, the number of registered marriages was almost constant at about 4000 per year (Statistics Botswana, 2014b). Moreover, marriage decision can be a source of migration in or out of the household. Two studies in Africa (Hoddinott and Mekasha, 2017 in Ethiopia and Hamoudi and Thomas, 2014 in South Africa) find that household size changes in response to public social protection programmes. Hoddinott and Mekasha (2017) demonstrate that the increase in household size was not due to increased fertility since participation in the

Productive Safety Net Programme reduced the likelihood (among adult female members) of fertility. According to them, household size increased because of the increase in the number of females aged 12-18 years, whose outmigration was reduced by their delayed marriage, possibly because they had to take over tasks previously performed by their mothers.

Changes in household size are tracked through population censuses and some national surveys. However, these may not adequately capture the dynamics of household size change since they do not follow the same households over time. Table 3.5 presents some demographic indicators for Botswana to shed light on household size patterns over time. The table shows that after an increase during 1971-1981, both crude death and birth rates consistently declined over time. Total fertility rates also declined, despite the increase in the share of women (out of total women) of reproductive age (15-49). The 2006 Botswana Demographic survey shows that age-specific fertility rates are higher among females aged 20-24 years, a pattern observed in both rural and urban areas (CSO, 2009b). The table shows that the number of households increased over time, and the average household size declined, which seems to suggest a pattern where, from the increased population size, they were new households.

The proportion of those in urban areas increased, implying possibilities of migration. Indeed, data from the 2006 Botswana Demography Survey shows that 45% of migration was from rural to urban areas (CSO, 2009b). Urban to rural area migration stood at 21 per cent, while urban to urban and rural to rural recorded 23 and 10 per cent respectively. Moreover, 36 per cent of internal citizen in-migrants reported that they migrated to join their parents, followed by those who cited job station/transfer (24%). On the other hand, among the non-citizens, high rates were for job-related factors. With regards to out-migration, among cities/towns, it was dominated by visit/vacation/leave while in rural areas, it was mainly for educational purposes. Incidents of twin births (which has been used as an instrument for family size elsewhere) is very low in Botswana. For instance, of the total 39,368 live births registered in 2011, only 397 (about 1 per cent) were twins (Statistics Botswana, 2014b) and twin births are not observed in either the 2007/08 BFHS or 2009/10 BCWIS data.

Table 3.5: Demographic Indicators: 1971-2011

Year	CDR	CBR	TFR	% of W15-49	% Urban	No. of HH	AHHSZ
1971	13.7	45.3	6.5	42.8	9.0		
1981	13.9	47.7	6.6	42.9	17.7	170,833	5.5
1991	11.5	39.3	4.2	46.5	45.7	276,209	4.7
2001	12.4	28.9	3.3	52.4	54.2	404,706	4.2
2006	11.2	29.8	3.2	-	-		
2011	6.25	25.7	2.8	54.4	64.1	550,920	3.7

Source: MFDP (2017:10)
Notes: CDR: Crude death rate, CBR: Crude birth rate, TFR: Total fertility rate, W15-49: women aged 15-49, HH: household, AHHSZ: Average household size

To control for the unobservables within and across households, we follow Mussa (2015), which is an extension of Picard and Wolff (2010)[9]. The approach involves obtaining both the explained and unexplained variation in child nutritional status across and within households in three steps. In the first step, a linear random effect model is estimated using maximum likelihood and fitted values of child nutritional indicators are obtained together with two random residuals for child and household unobservables. In the second step, the mean child nutritional status for each household is computed. The third step, aided by the outcome of the first two steps, involves the calculation of variances, which are used to compute the contribution of each component. This enables us to identify the contribution of the observed child and household characteristics as well as the proportion accounted for by the unobservables. The model is re-specified as follows:

$$N_{ch} = X'_{ch}\beta_1 + Y'_{ch}\beta_2 + \varepsilon_{ch} \qquad c=1\ldots C \qquad h=1\ldots H \tag{3.8}$$

$$\varepsilon_{ch} = u_h + z_{ch} \tag{3.9}$$

Where $N_{ch,}$ X_{ch} and Y_{ch} are child nutritional indicators measured by z scores, specific child characteristics and a vector for household characteristics, respectively, and the parameters represented by β include a constant. The key component in this specification is that the random error term ε_{ch} (equation 3.9) includes two errors: one for the unobserved household heterogeneity (u_h), and the other specific to the child (z_{ch}). These two errors (u_h and z_{ch}) are assumed to be independent and normally distributed. However, it is assumed that only the variance for the within unexplained part, $z_{ch,}$ is normalised to 1 (i.e., z_{ch} ~N (0; 1) and u_h~N (0,

[9] Following work on intrahousehold allocation by Behrman (1997) among others, these authors provide a review of literature on intrahousehold decisions and educational outcomes, and estimate a random effect model, to compute variances for education inequality in Albanian households.

δ^2_u). From equation (3.8) the equation for fitted values is obtained (equation 3.10), which can further be rewritten in mean deviation form (equation 3.11) using an overall mean ($\bar{n}$) and estimated mean ($\bar{\hat{n}}$) of nutritional indicators.

$$N_{ch} = \hat{n}_{ch} + \hat{\varepsilon}_{ch} \tag{3.10}$$

$$(n_{ch} - \bar{n}) = (\hat{n}_{ch} - \bar{\hat{n}}) + \hat{\varepsilon}_{ch} \tag{3.11}$$

Rearranging after some adjustments(adding and subtracting the mean of the estimated health for the hth household to the right of equation (3.11), squaring and summing), , total variation is made of explained and unexplained variations (terms in the first and second brackets respectively) as shown in equation (3.12) below. Furthermore, both variations are composed of between and within explained and unexplained variations. For instance, the between and within unexplained variations are represented by the last two terms (in the last squared bracket) of the entire equation.

$$\sum_{h=1}^{H}\sum_{C=1}^{C}(n_{ch}\text{-}\bar{n})^2 = \left[\sum_{h=1}^{H} C\left(\bar{\hat{n}}_h\text{-}\bar{\hat{n}}\right)^2 + \sum_{h=1}^{H}\sum_{c=1}^{C}\left(\hat{n}_{ch}\text{-}\bar{\hat{n}}_h\right)^2\right] + \left[C\sum_{h=1}^{H}\hat{u}_h^2 + \sum_{h=1}^{H}\sum_{c=1}^{C}\hat{Z}_{ch}^2\right] \tag{3.12}$$

Lastly, the proportions of variances to the total are obtained. For instance, the proportion of explained variation across the households is computed as follows:

$$\frac{\frac{\sum_{h=1}^{H} C(\bar{\hat{n}}_h-\bar{\hat{n}})^2}{h-1}}{\frac{\sum_{h=1}^{H}(\bar{\hat{n}}_h-\bar{\hat{n}})^2}{h-1}+\frac{\sum_{h=1}^{H}\sum_{c=1}^{C}(\hat{n}_{ch}-\bar{\hat{n}}_h)^2}{c-1}+\frac{C\sum_{h=1}^{H}\hat{u}_{ch}^2}{h-1}+\frac{\sum_{c=1}^{C}\sum_{h=1}^{H}\hat{z}_{ch}^2}{c-1}} \tag{3.13}$$

In equation (3.13), the numerator is the explained inequality across households. The denominator is a total of explained inequality across households, explained inequality within households, unexplained inequality in nutrition across households and unexplained inequality within households, respectively. The proportions of other variances are computed in a similar manner. Unlike other approaches that explore the within household variation in child nutrition (Statistics Botswana and UNICEF, 2008), our adopted approach does not necessarily tie the problem to income or socio-economic status, as such is conditioned on the choice of a variable

that classifies population into a social group (Murray et al., 1999). It thus enables us to compute proportions of the total variation both explained and unexplained, which is essential for policy intervention (Bell and Jones, 2015).

3.5 Results and Discussion

Table 3.6 presents OLS and random effect regression results, while probit results are presented in Table 3.7. We discuss them concurrently. OLS results show that compared to the first-order child (reference category), birth orders of 3 and 4 or more are negatively associated with weight for age. Weight for age of child`s birth order 3 and 4+ is 0.216 and 0.363 standard deviations lower than that of first birth order. These are supported by the probit model results, which show that the likelihood of being underweight increases with increasing birth order. Similarly, birth order 4 and above is negatively associated with child weight for height and increases the likelihood of child stunting (Table 3.7). Birth order 3 is negatively associated with weight for height. Overall, these suggest that children of high birth order are more likely to fare worse than their lower birth order counterparts in terms of child nutrition, consistent with the evidence in the Philippines (Horton, 1988), Ethiopia (Kebede, 2005), India (Jayachandran and Pande, 2017) and some other African countries (Howell et al., 2016). However, essential to note here is that in our approach, we do not assume that resources for the child`s growth and development are sourced solely from the child`s biological parents. Therefore, it would appear that in an extended family, under-five children`s nutrition could be disadvantaged by their entering the household late. Household size is negatively associated with weight for age and height for age and positively associated with the child`s likelihood of being both underweight and stunted. This may be interpreted to reflect the prediction of resource dilution in the household.

Male children have an increased likelihood of being underweight, but we find no effect of gender on child stunting. Age of the child is negatively associated with all the three nutritional indicators. Probit results confirm that child`s age is positively associated with the likelihood of child malnutrition, a pattern observed in previous studies in Botswana (Tharakan and Suchindran, 1999), Kenya (Kabubo-Mariara et al., 2008) and Morocco, where the effect was interpreted to reflect possible effects of child weaning (Glewwe, 1999). Contrary to expectation, OLS results show that the presence of a child`s mother in the household is negatively associated with child weight for age and positively associated with the probability of underweight. Perhaps this may have to do with the characteristics of these mothers as opposed to their presence/absence in the household. Or a possibility that more disadvantaged

mothers are more likely to join an extended family and to have lower health quality children. To assess their effect on child nutrition, we later separately present the results for those children who were residing with their mothers in the household. The incidence of father contributing to the support of the child is positively associated with weight for age. However, an interaction of the father's absence with a contribution towards child support does not have any effect on child nutrition.

Compared to grandchildren, being a son/daughter is positively associated with weight for age and height for age, suggesting a possibility that parents are more likely to prefer their biological children. However, children categorised as other relations have better height for age values than grandchildren, which is not easy to interpret. It could be explained by the characteristics of their biological parents (or other unobserved factors). It could be that those grandchildren in an extended family to be looked after by grandparents tend to be less healthy, possibly because their mothers are younger or disadvantaged. Age of the head of household is positively associated with weight for age and height for age. Similarly, from the probit results, Table 3.7 shows that this reduces the probability of child underweight. The number of household head`s years of schooling is positively associated with all the indicators, implying an important role played by household heads` education in child health. Rural area residence is negatively associated with weight for age and positively with the likelihood of child underweight.

The other key results, the decomposed variances and their proportions (in squared brackets), are presented in the bottom of Table 3.6. Overall, it is evident that larger shares of variance are from between than within households. For instance, the weight for age between variance accounts for about 86 per cent, while the rest is for the within variance. Moreover, for all the three indicators, the unexplained components account for larger variance than explained: more than 75 per cent between and more than 13 per cent within households. As for the explained component, there is less within variance than between household variance.

Table 3.6: Regression Results for Child Nutritional Indicators

	OLS			Random Effect		
Independent Variables	**WAZ (1)**	**HAZ (2)**	**WHZ (3)**	**WAZ (4)**	**HAZ (5)**	**WHZ (6)**
BO2	-0.110	-0.001	-0.174	-0.091	-0.006	-0.184
	(0.090)	(0.123)	(0.117)	(0.086)	(0.121)	(0.112)
BO3	-0.216*	-0.020	-0.356**	-0.188	-0.039	-0.390**
	(0.121)	(0.155)	(0.161)	(0.118)	(0.165)	(0.152)
BO4+	-0.363***	-0.356**	-0.135	-0.307***	-0.349**	-0.086
	(0.116)	(0.154)	(0.143)	(0.112)	(0.154)	(0.144)
HHSIZE	-0.027***	-0.033**	0.002	-0.034***	-0.028*	-0.007
	(0.010)	(0.013)	(0.012)	(0.011)	(0.015)	(0.014)
MaleC	-0.109	-0.112	0.038	-0.093	-0.109	0.019
	(0.085)	(0.117)	(0.109)	(0.081)	(0.113)	(0.105)
MaleC*BO2	0.000	-0.277	0.258	-0.003	-0.282	0.272*
	(0.132)	(0.183)	(0.171)	(0.126)	(0.177)	(0.164)
MaleC*BO3	0.130	0.097	0.098	0.131	0.058	0.161
	(0.172)	(0.226)	(0.222)	(0.160)	(0.225)	(0.208)
MaleC*BO4+	0.150	0.133	-0.122	0.127	0.126	-0.139
	(0.151)	(0.200)	(0.166)	(0.144)	(0.203)	(0.185)
AGEC	-0.011***	-0.009***	-0.019***	-0.010***	-0.009***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Diarrhoea	-0.120	0.013	-0.149	-0.144	0.004	-0.156
	(0.100)	(0.145)	(0.131)	(0.095)	(0.133)	(0.124)
MOTLIVEH	-0.145**	0.001	-0.025	-0.107	0.013	-0.016
	(0.073)	(0.098)	(0.095)	(0.075)	(0.101)	(0.094)
FATLIVEH	0.061	0.125	0.142	0.069	0.101	0.216
	(0.135)	(0.169)	(0.155)	(0.137)	(0.184)	(0.172)
FatContribute	0.173**	-0.018	0.124	0.165**	-0.012	0.122
	(0.069)	(0.094)	(0.090)	(0.071)	(0.095)	(0.089)
Fatliveh*FatContribute	-0.076	-0.168	-0.090	-0.107	-0.150	-0.197
	(0.158)	(0.208)	(0.190)	(0.163)	(0.218)	(0.205)
Son/Daughter	0.266**	0.354**	0.094	0.239**	0.369**	0.099
	(0.110)	(0.156)	(0.140)	(0.110)	(0.149)	(0.140)
Other Relative	0.184*	0.359	0.067	0.109	0.348**	0.001
	(0.102)	(0.140)	(0.130)	(0.105)	(0.139)	(0.133)
MaleHH	-0.052	-0.074	-0.073	-0.034	-0.077	-0.058
	(0.071)	(0.095)	(0.092)	(0.078)	(0.101)	(0.098)
AgeHH	0.042***	0.043***	0.011	0.040***	0.045**	0.006
	(0.010)	(0.015)	(0.014)	(0.012)	(0.016)	(0.015)
AgeHHSQ	-0.000***	-0.000**	-0.000	-0.000**	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MarriedHH	0.140**	0.164*	-0.003	0.124	0.161	-0.016
	(0.070)	(0.096)	(0.088)	(0.077)	(0.100)	(0.097)
EDUHH	0.028***	0.030***	0.024***	0.034***	0.031***	0.027***
	(0.007)	(0.009)	(0.009)	(0.008)	(0.010)	(0.010)
Rural	-0.202***	-0.038	-0.091	-0.173***	-0.027	-0.077
	(0.058)	(0.077)	(0.073)	(0.064)	(0.082)	(0.080)
Log Likelihood				-6431.467	-6315.098	-5731.163
Chi^2				143.16	84.98	130.90
Prob> Chi^2				0.000	0.000	0.000
Between Explained				0.314[8.20]	0.290[5.11]	0.258[5.31]
Between Unexplained				2.965[77.47]	4.402[77.45]	3.867[79.53]
Within Explained				0.018[0.47]	0.037[0.65]	0.043[0.893]
Within unexplained				0.531[13.87]	0.953[16.76]	0.694[14.26]
Observations	3481	3013	2863	3481	3013	2863

Notes: Each column lists estimates from separate regressions. The first three columns present OLS results for the three nutritional indicators, while columns 4-6 present the results for the random effect model. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table 3.7: Probit Marginal Effects for Child Malnutrition

Independent Variables	Underweight (1)	Stunting (2)	Wasting (3)
BO2	0.035*	0.033	0.004
	(0.021)	(0.028)	(0.020)
BO3	0.047*	0.055	0.024
	(0.027)	(0.038)	(0.026)
BO4+	0.063**	0.060*	0.004
	(0.025)	(0.033)	(0.025)
HHSIZE	0.005**	0.010***	0.001
	(0.002)	(0.003)	(0.002)
MaleC	0.045**	0.015	0.014
	(0.019)	(0.025)	(0.019)
MaleC*BO2	-0.024	0.093**	0.013
	(0.029)	(0.040)	(0.030)
MaleC*BO3	-0.004	0.013	0.022
	(0.036)	(0.052)	(0.036)
MaleC*BO4+	0.065*	0.042	0.006
	(0.033)	(0.044)	(0.033)
AGEC	0.000	0.004***	0.000
	(0.000)	(0.000)	(0.000)
Diarrhoea	0.027	0.026	0.010
	(0.020)	(0.028)	(0.021)
MOTLIVEH	0.032*	0.031	0.021
	(0.017)	(0.022)	(0.017)
FATLIVEH	0.008	0.004	-0.011
	(0.029)	(0.039)	(0.030)
FatContribute	-0.009	-0.007	-0.007
	(0.015)	(0.020)	(0.015)
Fatliveh*FatContribute	-0.032	-0.010	-0.002
	(0.034)	(0.047)	(0.035)
Son/Daughter	0.024	-0.036	0.027
	(0.024)	(0.032)	(0.024)
Other Relative	0.028	-0.034	0.004
	(0.022)	(0.030)	(0.022)
MaleHH	0.028*	0.005	0.012
	(0.015)	(0.021)	(0.016)
AgeHH	-0.004*	-0.004	0.003
	(0.002)	(0.003)	(0.002)
AgeHHSQ	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
MarriedHH	-0.037**	-0.003	0.000
	(0.015)	(0.021)	(0.015)
EDUHH	-0.003*	-0.008***	-0.002
	(0.002)	(0.002)	(0.002)
Rural	0.035***	-0.008	0.008
	(0.012)	(0.017)	(0.013)
Log Likelihood	-1391.131	-1720.301	-1021.783
Chi^2	64.20	157.73	18.09
$Prob>Chi^2$	0.000	0.000	0.7006
Observations	3481	3013	2863

Notes: Each column lists estimates from separate regressions. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

We also consider the effect of the age structure of children aged 5 years and below. This we do by controlling specifically for the number of children who are infants (aged less than one year),

aged 1-3 and those aged 4-5 years. Since these variables may correlate with household size, we leave out the latter in the estimation. Results are presented in Table 3.8, which shows that the number of children in the age of 0-1 is positively associated with weight for age and weight for height. On the other hand, the number of children aged 1-3 years is negatively associated with weight for age and height for age. This suggests that competition for resources could be introduced by those aged 1-3, while relatively younger ones benefit. Our results could alternatively be explained by the possibility that children in the 0-1 age category are likely to be still breastfed, leaving other resources for older children.

Table 3.9 presents results for the sub-sample of children who resided with their mothers in the household. The table shows that compared to the first birth order, other orders are negatively associated with child nutritional indicators. Compared to children whose mothers have junior secondary education (reference category), pre-secondary education is negatively associated with weight for age and height for age, while post-secondary education is positively associated with child weight for age. This suggests that child nutrition is likely to improve with mother`s education. The incidence of married mothers and mother`s age are positively associated with weight for age and height for age. The positive effect of mother's age may signal experience gained for child health production as mothers age, although there appears to be a pattern for an inverted U shape relationship. Table 3.10 presents the probit marginal effects for children residing with their mothers and generally support a pattern observed in Table 3.9.

We also examined the effect when mothers are heads of household, captured by a dummy variable. However, we maintain that mothers not residing with children are less likely to be heads of households. This is because while some mothers were not residing with their children, we have information for all heads as household residents. Moreover, none of the children`s absent mothers were reported as heads of households. Nonetheless, a consideration of the entire sample of children reveals that although this variable carries a positive sign in all the nutritional indicators, it is statistically insignificant. Similarly, it is insignificant in the subsample of those children residing with mothers. With regards to the latter, several possibilities emerge. When we cross-examine the incidence of married mothers with both household headship and the presence of the father, we conclude that mothers are likely to be household heads when unmarried and in the absence of the father. Since children of married mothers have better nutrition, this suggests that the effect outweighs that of mother`s household headship status.

Table 3.8: Regression results for Child Nutritional Indicators (omitting household size)

	OLS			Random MLE		
Independent Variables	**WAZ (1)**	**HAZ (2)**	**WHZ (3)**	**WAZ (4)**	**HAZ (5)**	**WHZ (6)**
BO2	-0.105	-0.004	-0.170	-0.090	-0.001	-0.184
	(0.091)	(0.123)	(0.118)	(0.087)	(0.122)	(0.113)
BO3	-0.232*	-0.015	-0.346**	-0.205*	-0.012	-0.391**
	(0.120)	(0.152)	(0.160)	(0.118)	(0.164)	(0.152)
BO4+	-0.370***	-0.383**	-0.122	-0.323***	-0.372**	-0.086
	(0.116)	(0.152)	(0.143)	(0.112)	(0.154)	(0.144)
MaleC	-0.113	-0.122	0.037	-0.096	-0.117	0.017
	(0.085)	(0.117)	(0.109)	(0.080)	(0.112)	(0.105)
MaleC*BO2	0.009	-0.247	0.254	0.002	-0.257	0.273*
	(0.132)	(0.183)	(0.172)	(0.126)	(0.176)	(0.164)
MaleC*BO3	0.154	0.127	0.102	0.150	0.082	0.169
	(0.172)	(0.224)	(0.222)	(0.160)	(0.224)	(0.208)
MaleC*BO4+	0.147	0.146	-0.122	0.127	0.136	-0.139
	(0.151)	(0.199)	(0.143)	(0.144)	(0.202)	(0.185)
AGEC	-0.010***	-0.008***	-0.018***	-0.010***	-0.008***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Diarrhoea	-0.126	0.025	-0.170	-0.147	0.015	-0.174
	(0.100)	(0.144)	(0.132)	(0.095)	(0.133)	(0.124)
MOTLIVEH	-0.186***	-0.050	-0.029	-0.146***	-0.028	-0.026
	(0.071)	(0.096)	(0.092)	(0.038)	(0.099)	(0.093)
FATLIVEH	0.055	0.097	0.155	0.066	0.077	0.227
	(0.135)	(0.171)	(0.154)	(0.137)	(0.184)	(0.172)
FatContribute	0.173**	-0.019	0.123	0.167**	-0.011	0.122
	(0.069)	(0.094)	(0.090)	(0.070)	(0.095)	(0.089)
Fatliveh*Contribute	-0.058	-0.120	-0.109	-0.094	-0.113	-0.210
	(0.159)	(0.210)	(0.191)	(0.163)	(0.218)	(0.205)
Son/Daughter	0.300***	0.402***	0.096	0.274***	0.404***	0.114
	(0.107)	(0.151)	(0.138)	(0.108)	(0.145)	(0.137)
Other Relative	0.190*	0.397	0.053	0.109	0.372***	0.006
	(0.103)	(0.140)	(0.130)	(0.105)	(0.139)	(0.133)
MaleHH	-0.083	-0.119	-0.074	-0.060	-0.111	-0.065
	(0.071)	(0.095)	(0.092)	(0.078)	(0.101)	(0.098)
AgeHH	0.039***	0.039**	0.011	0.036***	0.041**	0.006
	(0.010)	(0.015)	(0.014)	(0.012)	(0.016)	(0.015)
AgeHHSQ	-0.000***	-0.000*	-0.000	-0.000**	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MarriedHH	0.147**	0.187*	-0.007	0.127	0.178*	-0.019
	(0.070)	(0.096)	(0.088)	(0.077)	(0.099)	(0.097)
EDUH	0.029***	0.028***	0.025***	0.035***	0.029***	0.028***
	(0.007)	(0.009)	(0.009)	(0.008)	(0.010)	(0.010)
Children <1 year	0.130**	-0.066	0.139*	0.130*	0.072	0.165*
	(0.062)	(0.087)	(0.083)	(0.073)	(0.093)	(0.091)
Children 1-3 years	-0.158***	-0.231***	-0.012	-0.187***	-0.230***	-0.042
	(0.039)	(0.050)	(0.050)	(0.046)	(0.058)	(0.057)
Children 4-5 years	-0.022	0.032	-0.031	-0.022	0.035	-0.018
	(0.038)	(0.049)	(0.045)	(0.042)	(0.053)	(0.052)
Rural	-0.173***	-0.001	-0.092	-0.143**	-0.005	-0.073
	(0.058)	(0.077)	(0.074)	(0.064)	(0.082)	(0.080)
Log likelihood				-6427.646	-6307.322	-5729.599
LR Chi2				150.80	100.53	134.03
Prob>Chi2				0.000	0.000	0.000
Observations	3481	3013	2863	3481	3013	2863

Notes: Each column lists estimates from separate regressions. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table 3.9: Regression results for Children residing with their biological mothers

	OLS			Random Effect		
Independent Variables	**WAZ**	**HAZ**	**WHZ**	**WAZ**	**HAZ**	**WHZ**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
BO2	-0.255**	-0.112	-0.196	-0.237**	-0.100	-0.224
	(0.112)	(0.158)	(0.146)	(0.105)	(0.154)	(0.141)
BO3	-0.391***	-0.058	-0.468**	-0.353**	-0.029	-0.498***
	(0.139)	(0.191)	(0.181)	(0.136)	(0.194)	(0.179)
BO4	-0.547***	-0.474**	-0.260	-0.504***	-0.470**	-0.244
	(0.139)	(0.197)	(0.178)	(0.136)	(0.193)	(0.179)
HHSIZE	-0.009	-0.010	0.001	-0.015	-0.006	-0.002
	(0.012)	(0.017)	(0.016)	(0.014)	(0.018)	(0.017)
MaleC	-0.244**	-0.010	-0.146	-0.238**	-0.017	-0.170
	(0.107)	(0.156)	(0.145)	(0.101)	(0.145)	(0.135)
BO2*MaleC	0.077	-0.318	0.314	0.099	-0.333	0.346*
	(0.159)	(0.230)	(0.213)	(0.152)	(0.219)	(0.202)
BO3*MaleC	0.267	-0.020	0.249	0.266	-0.076	0.285
	(0.191)	(0.254)	(0.249)	(0.176)	(0.254)	(0.234)
BO4*MaleC	0.300*	-0.012	0.085	0.314*	-0.014	0.084
	(0.171)	(0.230)	(0.214)	(0.162)	(0.233)	(0.213)
AGEC	-0.012***	-0.006**	-0.021***	-0.011***	-0.006***	-0.021***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Diarrhoea	-0.144	-0.009	-0.140	-0.183*	-0.021	-0.158
	(0.108)	(0.158)	(0.153)	(0.107)	(0.154)	(0.143)
FATLIVEH	-0.142	-0.044	0.103	-0.033	-0.050	0.099
	(0.193)	(0.222)	(0.243)	(0.205)	(0.280)	(0.262)
FatContribute	0.092	-0.047	0.059	0.108	-0.043	0.067
	(0.084)	(0.117)	(0.110)	(0.085)	(0.117)	(0.110)
Fatliveh*FatContribute	0.126	0.036	-0.065	0.023	0.028	-0.072
	(0.215)	(0.256)	(0.271)	(0.227)	(0.310)	(0.289)
Son/Daughter	0.072	0.061	0.037	0.055	0.076	0.049
	(0.096)	0.137	(0.125)	(0.104)	(0.139)	(0.133)
Other Relative	0.147	0.176	0.248*	0.132	0.175	0.192
	(0.111)	(0.157)	(0.140)	(0.116)	(0.155)	(0.150)
Pre-Secondary	-0.241***	-0.214*	-0.120	-0.211**	-0.199*	-0.151
	(0.086)	(0.113)	(0.109)	(0.088)	(0.120)	(0.113)
Post-Junior Secondary	0.148*	0.086	0.116	0.131	0.074	0.094
	(0.078)	(0.108)	(0.103)	(0.081)	(0.112)	(0.106)
Mother Married	0.271***	0.335***	0.019	0.272***	0.313**	0.049
	(0.093)	(0.127)	(0.118)	(0.098)	(0.132)	(0.126)
Mother Age	0.096**	0.083*	0.044	0.096***	0.085*	0.050
	(0.041)	(0.048)	(0.039)	(0.032)	(0.044)	(0.042)
Mother Age squared	-0.001*	-0.001	-0.000	-0.001***	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Rural	-0.151**	-0.039	-0.058	-0.128*	-0.039	-0.042
	(0.067)	(0.092)	(0.086)	(0.073)	(0.096)	(0.092)
Log likelihood				-4908.869	-4779.608	-4316.238
Chi^2				123.12	57.44	107.78
Prob> Chi^2				0.000	0.000	0.000
Observations	2644	2254	2135	2644	2254	2135

Notes: Each column lists estimates from separate regressions. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. WAZ: Weight for age z score, HAZ: Height for age z score, WHZ: Weight for height z score. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table 3.10: Probit Marginal effects for children residing with their biological mothers

Independent Variables	Underweight (1)	Stunting (2)	Wasting (3)
BO2	0.047*	0.034	0.008
	(0.026)	(0.035)	(0.026)
BO3	0.071**	0.061	0.017
	(0.032)	(0.044)	(0.031)
BO4	0.079**	0.067	0.002
	(0.031)	(0.042)	(0.032)
HHSIZE	0.002	0.005	-0.001
	(0.003)	(0.004)	(0.003)
MaleC	0.060**	0.015	0.019
	(0.024)	(0.032)	(0.025)
BO2*MaleC	-0.029	0.105**	-0.018
	(0.036)	(0.048)	(0.037)
BO3*MaleC	-0.019	0.050	0.009
	(0.040)	(0.058)	(0.041)
BO4*MaleC	-0.087**	0.027	-0.026
	(0.038)	(0.051)	(0.038)
AGEC	0.000	0.003***	0.000
	(0.000)	(0.001)	(0.000)
Diarrhoea	0.031	0.020	0.010
	(0.023)	(0.032)	(0.025)
FATLIVEH	0.021	-0.020	0.028
	(0.041)	(0.059)	(0.042)
FatContribute	-0.009	-0.006	0.001
	(0.018)	(0.025)	(0.019)
Fatliveh*FatContribute	-0.029	-0.005	-0.054
	(0.046)	(0.065)	(0.047)
Son/daughter	-0.002	-0.007	-0.005
	(0.021)	(0.029)	(0.021)
Other relative	-0.013	-0.017	-0.052*
	(0.024)	(0.032)	(0.027)
Pre-Secondary	0.044**	0.070***	0.028
	(0.018)	(0.024)	(0.018)
Post-Junior Secondary	-0.025	-0.044*	-0.017
	(0.018)	(0.024)	(0.018)
Mother Married	-0.048**	-0.046	0.013
	(0.022)	(0.029)	(0.021)
Mother Age	-0.007	-0.017*	0.006
	(0.007)	(0.009)	(0.007)
Mother Age squared	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Rural	0.024*	0.017	0.006
	(0.014)	(0.020)	(0.015)
Log likelihood	-1080.556	-1312.462	-783.299
Chi^2	64.49	111.06	19.42
$Prob>Chi^2$	0.000	0.000	0.558
Observations	2644	2254	2135

Notes: Each column lists estimates from separate regressions. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

3.6 Robustness Checks

A major critique of the random effect model relates to its exogeneity assumption between residual and observable covariates, which has resulted in fixed effect models being considered as a default (Bell and Jones, 2015). Although fixed effects is generally preferred, its application in our context would be to consider a within estimator approach, which involves taking differences on child characteristics within the household (Horton, 1988). We, therefore, compute the household mean values for child characteristics, obtain deviations from their mean and run regressions. In this case, the dependent variables are the mean deviations of child nutritional status. We compare the results with those for which the only considered households are those with at least two children. As Table A3.1 shows, in the within estimator approach child age and gender are statistically significant. However, with this estimation we lose household level variables (Greene, 2008; Bell and Jones, 2015).

On the other hand, Table A3.2 shows that results are largely similar to the previous ones. We caution that our within estimator results need careful interpretation, as they may be sensitive to the sample size. Unlike Horton (1988)[10] who maintains that this was not a problem in the case of the Philippines as less than 10 per cent of families had one child, about 39 per cent of our children come from households with one child and the majority of others have two children.

An alternative approach is the within-between effect model of Bell and Jones (2015), which accounts for the two effects. Unlike the fixed effect, which only considers the within effects, their (random effect) approach encompasses a fixed component[11] and deviation of variables from the household mean. Not only does the approach address endogeneity, but explores its source(s), motivated by the inadequate attention accorded to the reasons why random effect model assumptions may not hold. Results are presented in Table A3.3, which shows that as in the previous estimates, high birth order children fare worse than their lower birth order counterparts and household size is negatively associated with weight for age and height for age.

[10] She however finds that only deviated sex and squared birth order affected height for age, while weight for height is affected by deviated sex and birth order.

[11] It is acknowledged that this component was introduced by Mundlak (1978), who maintained that the choice between fixed and random effect model was arbitrary. While Hausman test is commonly used to aid the choice between the two, Bell and Jones (2015:138) argue that for their approach the test is redundant as it "does not address the decision framework for a wider class of problems"

We also estimated regression for the subsamples of infants, those aged 1-3 years and those aged 4-5 years. The pattern for these subsamples is similar to the one previously presented where the age group 1-3 accounts for a relatively larger share of about 35 per cent. Regression results are presented in the Appendix (Tables A3.4-A3.6). Child birth order 4 and above is negatively associated with weight for age for infants and for those aged 1-3 while birth order 3 is negatively associated with weight for age for those aged 1-3. The incident of diarrhoea is negatively associated with weight for age for infants, while father's contribution is positively associated with infants` weight for age. Education of the head of the household is positively associated with weight for age for those aged 1-3 and 4-5 years and height for age for infants and those aged 1-3 years. Household size is negatively associated with weight for age for those aged 4-5 years, height for age for infants and weight for height for those aged 4-5 years. Rural area residence is negatively associated with infants` weight for age as well as for those aged 4-5 years.

As indicated, we also analysed the Botswana Family Health Survey data. Tables B3.2-B3.4 present regression results for weight for age, height for age and weight for height z scores, respectively. These are the estimation results for equation 3.2 including village and district fixed effects (see columns II and III of all the tables) with clustered standard errors to control for any unobserved heterogeneity at both village and district levels. As seen, being a male child is negatively associated with weight for age and height for age. Child`s age is also negatively associated with all three indicators. Similarly, the incident of diarrhoea is negatively associated with the three indicators, signifying the effect that illness may have on child nutrition. Ever breastfed children are positively associated with weight for age and weight for height. Compared to those whose mothers/carers have secondary education (reference category), mother`s pre-secondary education is negatively associated with child weight for age and height for age. The incidence of sourcing water within the household is positively associated with the three child nutritional indicators. Compared to ownership of pit-latrines, ownership of flush toilet has is positively associated with child nutrition.

3.7 Conclusion

In this chapter, following the deteriorating child nutritional indicators in Botswana, we investigate factors influencing both child nutrition and the likelihood of child malnutrition at the household level. Specifically, we answer the question on the impact of household size and

child`s birth order on child nutrition. Using the 2009/10 Botswana Core Welfare indicator survey, we estimate the random effect model, which allows investigating heterogeneity both across and within households. We find that high birth order children are more likely to fare worse than lower birth order children in terms of child nutrition and that not only do children from larger sized households have lower weight for age and height for age, they are more likely to be malnourished. Consideration of child age structure suggests that competition for resources adversely affects those aged 1-3 perhaps because young infants are still breastfed. A different pattern for some countries in Latin America, Asia and Sub Saharan Africa was observed for children aged less than three years (Desai, 1995). We also find that sons/daughters have better weight for age and height for age, suggesting a possibility that parents are more likely to prefer their biological children. However, children categorised as other relations have a better height for age values than grandchildren. These results may indicate an effect of unobserved factors such as parental genetic factors, as observed in Ethiopia by Kebede (2005). However, the difference is not significant in the Probit estimates, so should not be over-interpreted. Educational level, age and marital status of household head affect child nutrition. Given that household heads are not biological parents to all children under consideration, these imply that the generally held assumption underpinning the theory may not hold in the context of Botswana households.

From variance decomposition analysis, there are higher variances across than within households. Moreover, higher variances are unexplained by the observed characteristics. We derive two main implications from these results. Firstly, while results suggest evidence for the resource dilution hypothesis, the negative effect of high birth order could be a reflection for such children entering the household late. Secondly, given that children aged 6-60 months receive monthly food ration, the negative effect of household size implies that the effectiveness of food ration programme could be threatened. While results show that there is higher variation in nutrition between than within households, the effect of intrahousehold allocation may not be underestimated particularly that we also establish a larger share of variation to be accounted for unobserved factors, partly because of data limitation. We could not control for all ideal variables such as characteristics of the fathers who do not reside with their children due to data limitation. Nonetheless, with such caveats in mind, our results also highlight a threat to attaining the population policy objectives for a healthy life and the essential health package intent of ensuring that these children reach their potential as adults.

APPENDIX FOR CHAPTER 3

Table A3.1: Fixed Effect Regressions

Independent Variables	WAZ (1)	HAZ (2)	WHZ (3)
BO2	-0.010	-0.010	-0.039
	(0.054)	(0.080)	(0.072)
BO3	-0.035	-0.035	-0.163*
	(0.072)	(0.097)	(0.089)
BO4+	-0.022	-0.030	-0.016
	(0.058)	(0.087)	(0.076)
AGEC deviation	-0.007***	-0.015***	-0.018***
	(0.002)	(0002)	(0.002)
Sex deviation	0.017	-0.221**	0.129*
	(0.061)	(0.086)	(0.077)
Diarrhoea deviation	-0.143	-0.178	-0.118
	(0.111)	(0.212)	(0.172)
Motherliveh deviation	0.034	0.148	-0.049
	(0.101)	(0.131)	(0.148)
Fatliveh deviation	-0.098	-0.094	-0.004
	(0.105)	(0.164)	(0.148)
Father Contribute deviation	0.058	0.006	-0.135
	(0.094)	(0.140)	(0.132)
Observations	2043	1711	1610

Notes: Each column lists estimates from separate regressions. Columns 1, 2 and 3 present regression results for weight for age, height for age and weight for height z scores, respectively. The dependent variables are the mean deviations of these nutritional indicators. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A3.2: Regression Results when Excluding One under 5 child Household

Independent Variables	WAZ (1)	HAZ (2)	WHZ (3)
BO2	-0.073	-0.064	-0.062
	(0.116)	(0.151)	(0.151)
BO3	-0.253	-0.247	-0.131
	(0.156)	(0.205)	(0.205)
BO4+	-0.399***	-0.504***	-0.035
	(0.144)	(0.107)	(0.178)
HHSIZE	-0.025**	-0.054***	0.015
	(0.012)	(0.017)	(0.016)
MaleC	-0.059	-0.259*	-0.295**
	(0.116)	(0.108)	(0.144)
MaleC*BO2	-0.006	-0.009	0.042
	(0.169)	(0.236)	(0.221)
MaleC*BO3	0.070	0.226	0.249
	(0.224)	(0.296)	(0.281)
MaleC*BO4+	0.170	0.245	-0.400*
	(0.192)	(0.248)	(0.234)
AGEC	-0.009***	-0.013***	-0.019***
	(0.002)	(0.003)	(0.003)
Diarrhoea	-0.079	-0.132	-0.108
	(0.126)	(0.182)	(0.171)
MOTLIVEH	-0.074	0.057	-0.020
	(0.089)	(0.122)	(0.118)
FATLIVEH	0.139	0.146	0.155
	(0.173)	(0.228)	(0.216)
FatContribute	0.148*	-0.004	0.012
	(0.087)	(0.115)	(0.110)
FatliveH*FatContribute	-0.199	-0.183	-0.164
	(0.204)	(0.274)	(0.257)
Son/Daughter	0.145	0.310	0.011
	(0.138)	(0.192)	(0.174)
Other Relative	0.336***	0.522***	0.188
	(0.119)	(0.180)	(0.147)
Male HH	-0.023	-0.076	-0.018
	(0.088)	(0.121)	(0.117)
AgeHH	0.040***	0.049**	0.011
	(0.014)	(0.021)	(0.018)
AgeHHSQ	-0.000***	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)
MarriedHH	0.136	0.186	0.000
	(0.089)	(0.126)	(0.114)
EDUHH	0.088	0.027**	0.011
	(0.009)	(0.012)	(0.011)
Rural	-0.240***	-0.075	-0.091
	(0.074)	(0.097)	(0.096)
Observations	2043	1711	1610

Notes: Each column lists estimates from separate regressions. The dependent variables are Weight for age z-scores(WAZ), Height for age z-scores(HAZ) Weight for Height z-scores (WHZ). Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A3.3: Within-Between Effect Regression Results for Child Nutritional Indicators

Independent Variables	WAZ (1)	HAZ (2)	WHZ (3)
BO2	-0.052	-0.045	-0.053
	(0.078)	(0.113)	(0.103)
BO3	-0.162	-0.080	-0.288**
	(0.106)	(0.150)	(0.140)
BO4+	-0.271**	-0.367**	-0.130
	(0.110)	(0.148)	(0.140)
HHSIZE	-0.031**	-0.056***	-0.016
	(0.015)	(0.018)	(0.019)
Sex Deviation	-0.010	-0.238**	0.136
	(0.074)	(0.111)	(0.099)
AGEC deviation	-0.008***	-0.013***	-0.019***
	(0.002)	(0.003)	(0.003)
Diarrhoea deviation	-0.127	-0.148	-0.086
	(0.153)	(0.238)	(0.208)
MOTLIVEH deviation	0.058	0.153	-0.032
	(0.122)	(0.182)	(0.157)
FATLIVEH deviation	-0.072	-0.106	-0.007
	(0.135)	(0.202)	(0.176)
FatContribute deviation	0.052	-0.006	-0.133
	(0.113)	(0.172)	(0.148)
Son/Daughter	0.081	0.339*	0.039
	(0.133)	(0.177)	(0.170)
Other Relative	0.236*	0.529**	0.066
	(0.133)	(0.172)	(0.168)
MaleHH	-0.022	-0.114	-0.011
	(0.100)	(0.124)	(0.126)
AgeHH	0.033*	0.054**	0.004
	(0.017)	(0.021)	(0.021)
AgeHHSQ	-0.000*	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)
MarriedHH	0.113	0.233*	-0.073
	(0.107)	(0.132)	(0.134)
EDUHH	0.013	0.028**	0.015
	(0.011)	(0.014)	(0.014)
Rural	-0.226**	-0.066	-0.083
	(0.068)	(0.109)	(0.110)
Log Likelihood	-3701.481	-3502.614	-3140.499
Chi^2	46.87	72.79	60.77
Prob> Chi^2	0.002	0.000	0.000
Observations	2043	1711	1610

Notes: Each column lists estimates from separate regressions. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A3.4: Regression Results for Weight for age Z score

	< 1 year	1-3 years	4-5 years
BO2	-0.198	-0.024	-0.093
	(0.190)	(0.142)	(0.136)
BO3	-0.136	-0.377*	-0.068
	(0.219)	(0.195)	(0.213)
BO4+	-0.491**	-0.394**	-0.133
	(0.223)	(0.198)	(0.186)
HHSIZE	-0.025	0.006	-0.067***
	(0.021)	(0.016)	(0.015)
MaleC	-0.232	-0.250*	0.132
	(0.176)	(0.144)	(0.137)
BO2*MaleC	0.148	0.119	-0.184
	(0.265)	(0.215)	(0.212)
BO3*MaleC	0.039	0.343	0.158
	(0.310)	(0.280)	(0.305)
BO4*MaleC	0.096	0.282	0.154
	(0.282)	(0.259)	(0.244)
Diarrhoea	-0.252*	-0.007	0.129
	(0.144)	(0.171)	(0.240)
MOTLIVEH	-0.057	-0.146	-0.158
	(0.165)	(0.113)	(0.120)
FATLIVEH	-0.119	0.071	0.072
	(0.332)	(0.224)	(0.182)
FatContribute	0.272**	0.089	0.136
	(0.133)	(0.113)	(0.115)
FatliveH*FatContribute	-0.219	0.117	-0.049
	(0.370)	(0.258)	(0.235)
Son/Daughter	0.361*	0.072	0.254
	(0.206)	(0.181)	(0.188)
Other relative	0.017	0.097	0.369**
	(0.203)	(0.149)	(0.177)
MaleHH	-0.064	0.065	-0.121
	(0.141)	(0.108)	(0.118)
AGEHH	0.027	0.028	0.057***
	(0.019)	(0.017)	(0.017)
AGEHHSQ	-0.000	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)
MarriedHH	0.007	0.219**	0.190*
	(0.137)	(0.111)	(0.116)
EDUHH	0.011	0.045***	0.021*
	(0.013)	(0.011)	(0.011)
Rural	-0.201**	-0.142	-0.278***
	(0.108)	(0.096)	(0.096)
Observations	1162	1244	1075

Notes: Each column lists estimates for a sub-sample of children. In all the three samples, the dependent variable is weight for age z scores. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A3.5: Regression Results for Height for age z scores for the subsamples

	< 1 year	1-3 years	4-5 years
BO2	0.249	-0.012	-0.143
	(0.261)	(0.189)	(0.195)
BO3	0.065	0.155	0.012
	(0.322)	(0.220)	(0.258)
BO4	-0.450	-0.273	-0.185
	(0.285)	(0.252)	(0.258)
HHSIZE	-0.052*	-0.026	-0.025
	(0.027)	(0.021)	(0.020)
MaleC	-0.268	-0.040	-0.042
	(0.258)	(0.187)	(0.173)
BO2*MaleC	-0.249	-0.265	-0.265
	(0.390)	(0.277)	(0.294)
BO3*MaleC	0.151	0.109	-0.017
	(0.464)	(0.332)	(0.368)
BO4*MaleC	0.229	0.022	0.191
	(0.412)	(0.322)	(0.306)
Diarrhoea	-0.058	0.350	0.134
	(0.217)	(0.233)	(0.318)
MOTLIVEH	0.594***	-0.330**	-0.062
	(0.219)	(0.150)	(0.158)
FATLIVEH	-0.464	0.099	0.391*
	(0.444)	(0.261)	(0.258)
FatContribute	-0.069	-0.013	-0.060
	(0.194)	(0.141)	(0.150)
FatliveH*FatContribute	0.094	-0.001	-0.231
	(0.503)	(0.320)	(0.318)
Son/Daughter	0.449	0.129	0.352
	(0.304)	(0.249)	(0.255)
Other Relative	0.276	0.233	0.375
	(0.283)	(0.199)	(0.233)
MaleHH	0.043	0.028	-0.260*
	(0.196)	(0.145)	(0.152)
AGEHH	0.037	0.031	0.036
	(0.031)	(0.024)	(0.023)
AGEHHSQ	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
MarriedHH	0.084	0.241*	0.175
	(0.194)	(0.144)	(0.160)
EDUHH	0.056***	0.017	0.021
	(0.019)	(0.015)	(0.015)
Rural	-0.021	0.057	-0.182
	(0.157)	(0.116)	(0.124)
Observations	953	1075	985

Notes: Each column lists estimates for a sub-sample of children. In all the three samples, the dependent variable is height for age z scores. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table A3.6: Regression Results for Weight for height z scores for the subsamples

	< 1 year	1-3 years	4-5 years
BO2	-0.342	-0.043	-0.225
	(0.236)	(0.193)	(0.189)
BO3	-0.377	-0.425*	-0.311
	(0.310)	(0.242)	(0.291)
BO4	-0.148	-0.190	-0.084
	(0.269)	(0.233)	(0.241)
HHSIZE	0.015	0.025	-0.036*
	(0.025)	(0.021)	(0.020)
MaleC	-0.146	-0.016	0.234
	(0.229)	(0.188)	(0.161)
BO2*MaleC	0.499	0.101	0.357
	(0.349)	(0.282)	(0.284)
BO3*MaleC	0.037	0.210	0.177
	(0.434)	(0.346)	(0.373)
BO4*MaleC	-0.102	-0.020	-0.208
	(0.370)	(0.307)	(0.298)
Diarrhoea	-0.263	-0.238	0.212
	(0.175)	(0.244)	(0.356)
MOTLIVEH	-0.068	-0.009	0.022
	(0.223)	(0.143)	(0.149)
FATLIVEH	0.038	0.256	0.042
	(0.380)	(0.260)	(0.214)
FatContribute	0.192	0.005	0.179
	(0.177)	(0.146)	(0.171)
FatliveH*FatContribute	-0.049	-0.202	0.030
	(0.439)	(0.315)	(0.285)
Son/Daughter	0.179	0.185	-0.120
	(0.276)	(0.219)	(0.229)
Other relative	-0.119	-0.137	0.105
	(0.259)	(0.204)	(0.208)
MaleHH	-0.198	-0.006	-0.016
	(0.187)	(0.144)	(0.157)
AGEHH	0.001	-0.004	0.037*
	(0.028)	(0.021)	(0.021)
AGEHHSQ	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
MarriedHH	-0.039	0.034	-0.020
	(0.175)	(0.137)	(0.153)
EDUHH	0.032*	0.029**	0.010
	(0.018)	(0.015)	(0.014)
Rural	-0.049	-0.080	-0.143
	(0.144)	(0.118)	(0.118)
Observations	881	1057	931

Notes: Each column lists estimates for a sub-sample of children. In all the three samples, the dependent variable is weight for height for age z scores. Robust Standard Errors are in parenthesis. Omitted categories are BO1 for birth order and grandchild for child relation with the household head. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table B3.1: Summary Statistics from 2007/08 BFHS data

Variable	Mean	Std. Dev.
Dependent		
WAZ	-0.729	1.262
HAZ	-1.067	1.625
WHZ	-0.031	1.484
Independent		
Child Characteristics		
MaleC	0.504	0.500
AGEC	28.926	16.722
Diarrhoea	0.094	0.292
Breastfed	0.763	0.426
Rural	0.499	0.500
Household Characteristics		
Mother Age	33.441	12.546
No Education	0.130	0.337
Primary	0.271	0.445
Secondary	0.584	0.493
Non-Formal	0.013	0.112
Piped indoors	0.154	0.361
Tap in yard	0.412	0.492
Communal tap	0.305	0.460
Bouser/Tanker	0.004	0.062
Well	0.016	0.125
Borehole	0.051	0.219
River/stream	0.014	0.115
Dam	0.006	0.078
Rainwater tank	0.001	0.034
Bottled water from stores	0.005	0.071
Neighbours	0.033	0.179
Own Flush toilet	0.171	0.377
Own Ventilated Improved Pit latrine	0.198	0.399
Own Pit latrine	0.312	0.463
Own Environment loo	0.002	0.048
Communal flush toilet	0.003	0.055
Communal Ventilated Improved Pit latrine	0.003	0.052
Communal pit latrine	0.028	0.165
Neighbour’s toilet	0.043	0.203
No Toilet	0.238	0.426

Source: author`s computation from 2007/08 BFHS data

Table B3.2: Regression results for Weight for age z scores from 2007/08 BFHS Data

Variable	I (Baseline)	II (Village Effect)	III (District Effect)
MaleC	-0.084*	-0.092***	-0.079**
	(0.048)	(0.033)	(0.036)
AGEC	-0.014***	-0.014***	-0.014***
	(0.002)	(0.001)	(0.001)
Diarrhoea	-0.387***	-0.396***	-0.393***
	(0.082)	(0.070)	(0.061)
Breastfed	0.140**	0.149***	0.144**
	(0.058)	(0.050)	(0.062)
Rural	-0.059	-0.019	-0.061
	(0.058)	(0.065)	(0.092)
Mother Age	0.004*	0.004***	0.004
	(0.002)	(0.002)	(0.003)
Pre-Secondary	-0.131**	-0.129**	-0.117
	(0.063)	(0.053)	(0.095)
Fatliveh	-0.071	-0.079***	-0.071
	(0.049)	(0.026)	(0.048)
Water indoors	0.271***	0.235***	0.278***
	(0.066)	(0.058)	(0.065)
Own flush toilet	0.409***	0.394***	0.421***
	(0.076)	(0.044)	(0.103)
No toilet	0.010	0.023	0.016
	(0.076)	(0.076)	(0.101)
Other toilet	-0.021	-0.016	-0.012
	(0.086)	(0.097)	(0.017)
HHSZ	-0.010	-0.007	-0.012
	(0.007)	(0.006)	(0.008)
Observations	2524	2524	2524

Notes: Each column lists estimates from separate regressions. The second and third columns control for village and district effects, respectively. Clustered Standard Errors are in parenthesis. Omitted category for toilet ownership is neighbour`s toilet. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table B3.3: Regression results for Height for age z scores from 200708 BFHS Data

Variable	I (Baseline)	II (Village Effect)	III (District Effect)
MaleC	-0.218***	-0.219***	-0.223***
	(0.064)	(0.057)	(0.055)
AGEC	-0.007***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)
Diarrhoea	-0.345***	-0.360***	-0.321***
	(0.116)	(0.111)	(0.113)
Breastfed	0.080	0.080	0.083
	(0.080)	(0.074)	(0.093)
Rural	-0.051	-0.058	-0.046
	(0.085)	(0.099)	(0.123)
Mother Age	0.004	0.004	0.004
	(0.003)	(0.003)	(0.004)
Pre-Secondary	-0.240***	-0.229***	-0.228*
	(0.080)	(0.056)	(0.113)
Fatliveh	0.001	0.000	0.007
	(0.065)	(0.058)	(0.068)
Water Indoors	0.233**	0.180**	0.220**
	(0.092)	(0.079)	(0.095)
Own flush toilet	0.377***	0.382***	0.331***
	(0.095)	(0.065)	(0.114)
No toilet	0.050	0.053	0.095
	(0.104)	(0.105)	(0.112)
Other toilet	-0.158	-0.179	-0.137
	(0.122)	(0.125)	(0.148)
HHSZ	-0.023**	-0.016	-0.025*
	(0.009)	(0.010)	(0.014)
Observations	2524	2524	2524

Notes: Each column lists estimates from separate regressions. The second and third columns control for village and district effects, respectively. Clustered Standard Errors are in parenthesis. Omitted category for toilet ownership is neighbour`s toilet ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

Table B3.4: Regression results for Weight for height z scores from 2007/08 BFHS Data

Variable	I (Baseline)	II (Village Effect)	III (District Effect)
MaleC	-0.032	-0.044	-0.021
	(0.057)	(0.043)	(0.045)
AGEC	-0.025***	-0.024***	-0.025***
	(0.002)	(0.002)	(0.002)
Diarrhoea	-0.187*	-0.187**	-0.209
	(0.103)	(0.083)	(0.129)
Breastfed	0.123*	0.135**	0.124
	(0.070)	(0.054)	(0.091)
Rural	0.005	0.060	-0.034
	(0.074)	(0.073)	(0.108)
Mother Age	0.002	0.002	0.002
	(0.003)	(0.002)	(0.003)
Pre-Secondary	0.042	0.030	0.052
	(0.074)	(0.056)	(0.079)
FATLIVEH	-0.096*	-0.108**	-0.087
	(0.058)	(0.048)	(0.060)
Water Indoors	0.192**	0.185**	0.209**
	(0.082)	(0.077)	(0.072)
Own flush toilet	0.273***	0.255***	0.337**
	(0.091)	(0.051)	(0.119)
No toilet	-0.050	-0.023	-0.077
	(0.090)	(0.109)	(0.120)
Other toilet	0.078	0.101	0.066
	(0.108)	(0.139)	(0.110)
HHSZ	0.002	-0.000	0.002
	(0.009)	(0.007)	(0.009)
Observations	2524	2524	2524

Notes: Each column lists estimates from separate regressions. The second and third columns control for village and district effects, respectively. Clustered Standard Errors are in parenthesis. Omitted category for toilet ownership is neighbour`s toilet. ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level.

CHAPTER 4: HIV RISK AND SEXUAL BEHAVIOUR IN BOTSWANA

Abstract

This chapter investigates whether sexual behaviour (indicators of activity and use of a condom) is responsive to both objective risk (HIV prevalence rate) and perceived risk (concern on the likelihood of being infected by HIV due to ART). Using three of the four waves of the Botswana AIDS Impact Survey (BAIS) collected by Statistics Botswana, we account for the endogeneity of both risk indicators through the method of instrumental variables. HIV prevalence rate is instrumented with the distance from respondent`s area of residence to the district with the highest HIV rate, while the perceived risk is instrumented with duration (in months) between antiretroviral therapy (ART) rollout and the survey date. Results suggest there is more responsiveness to subjective rather than objective risk indicators. Compared to those who indicate to have not changed their level of concern about being infected by HIV due to ART, the less (more) concerned are less likely to abstain and use a condom (more likely to use condoms). The HIV prevalence rate has no effect on both the number of sexual partners and use of condoms. From a policy perspective, the non-behavioural response to HIV prevalence could be a cause for concern if there are unobserved fatal behavioural patterns. Thus, because programmes such as ART prolongs the lives of the infected and affects perception about the likelihood of infection risk, there could be some negative externalities as revealed by the results.

4.1 Introduction

Epidemiological models predict that higher HIV prevalence subsequently results in increasing numbers of those who get infected, an approach argued to be treating both the sexual contact rate and virus transmission probability as constant (Hauck, 2018). Contrary to this prediction, there are arguments that high prevalence may motivate susceptible individuals to increase their self-protective behaviour to reduce exposure to risk (Laxminarayan and Malani, 2011; Geoffard and Philipson, 1996; Philipson and Posner, 1993). It is also established that responding to the risk of infection may be affected by prior beliefs/perceptions about the likelihood of being infected (Boozer and Philipson, 2000; Baird et al., 2014; Paula et al., 2014), whose direction of the effect is not obvious (Gong, 2015). This implies that the effect of prevalence on behaviour is ambiguous and possibly context-specific, a factor that has to be considered in evaluating the effects of the epidemic (Gersovitz and Hammer, 2003). This chapter assesses the evidence for the context in Botswana, which provides a useful case given a policy to promote awareness about HIV for behavioural change.

Empirical evidence on sexual behaviour in the presence of HIV is mixed (Guillon and Thuilliez, 2015; Oster, 2012; Tenkorang et al., 2011). Lakdawalla et al. (2006) find in the US that HIV positive individuals receiving treatment increase their demand for sexual partners. However, Auld (2006) finds that responsiveness to prevalence in the US is not homogenous: an increasing prevalence led to a larger decrease in risky behaviour among low-risk individuals than among risky ones. In Tanzania and Kenya, Gong (2015) finds that those surprised by HIV positive test results increase their risky sexual behaviour, while those surprised by negative HIV test results reduce their risky behaviour. In contrast, Delavande and Kohler (2012) find, in Malawi, that learning of an HIV positive result reduces the likelihood of risky behaviour and has no observable effect on either the reported number of sexual partners or condom use among the HIV negative individuals. Among others, the mixed evidence may be attributed to factors such as inadequate consideration of endogeneity between prevalence (Oster, 2012), individual`s beliefs (Paula et al., 2014) and sexual behaviour.

This chapter provides an assessment of the responsiveness to the risk of HIV infection in Botswana. The focus on sexual behaviour is based on the consensus that it plays an important role in HIV infection (Oster, 2005; Francis, 2008). We particularly consider the impact of HIV prevalence and perceived likelihood of HIV infection due to antiretroviral therapy (ART) on demand for sex (indicators for activity and the number of partners) and the use (or non-use) of

condoms given its effectiveness in protecting against HIV (Davis and Weller, 1999).Since Botswana`s first AIDS case in 1985, monitoring of the prevalence has been through the Sentinel Surveillance, which mainly covered pregnant women utilising antenatal care services (UNDP, 2000). During the period from 1992 (the year of first HIV surveillance) to 2003, national median HIV prevalence rate increased from 18% to 38%, and since 1995 the rate was consistently above 30% (NACA, 2003). Furthermore, during the period from 1992 to 1997, AIDS-related deaths rose from 2.4% to 9% of the total hospital deaths for those aged 15-44, and the virus has been among the top three indirect causes of maternal deaths (UNDP, 2000; MOH, nd; MOHW, nd). The first national HIV/AIDS assessment survey was in 2001, although no blood tests were carried out. Subsequent surveys in 2004, 2008 and 2013 reveal prevalence rates of 17.1%, 17.6% and 18.5%[12] respectively (NACA and CSO, 2005; CSO and NACA, 2009; Statistics Botswana and NACA, 2016).

In Botswana, previous empirical work that investigated influential factors for sexual behaviour only considered respondents that reported being sexually active (Keetile and Kgosidintsi, 2018; Letamo et al., 2017; Balekang and Dintwa, 2016; Keetile and Letamo, 2015; Keetile, 2014; Malema, 2012), resulting in potential endogeneity (Wooldridge, 2010). Moreover, these studies have not addressed the potential endogeneity of both HIV and perceptions about the likelihood of its infection. For instance, through a logistic regression approach, Keetile & Letamo, (2015) and Letamo et al. (2017) find opposing results on the effect of perception about ART on sexual behaviour. Agiya & Letamo (2011) and Balekang & Dintwa (2016) use a similar approach to investigate the impact of circumcision on sexual behaviour and do not account for the self-selection into circumcision. In investigating the effect of highly active antiretroviral therapy on sexual behaviour, a study by Keetile and Kgosidintsi (2018) does not account for the fact that assignment to treatment is not random.

In addition to policy debates resulting from the mixed empirical evidence, the need to carry out this investigation is motivated by the following. Firstly, family health surveys in Botswana show increasing rates of contraceptive use from 16% to 67% over 1984-2017 (CSO, 1999; 2001; 2009a; Statistics Botswana, 2018). While this pattern may have facilitated the policy

[12] These rates may not be comparable to the ones obtained through surveillance reports. The sentinel surveys estimated prevalence for males by multiplying the female prevalence by a male-female infection ratio, and further by the male population (NACA, 2003).

objective to reduce the total fertility rate, it might have led to a neglect in controlling HIV/AIDS infection, which has been increasing over this period.

Secondly, insights from AIDS impact survey reports suggest a behavioural pattern that warrants further investigation to ascertain responsiveness to HIV risk. For instance, for the surveyed 15-24 age cohort there was an increase from 3.5% to 4.4% between 2008 and 2013 in the proportion of those who had sex before age 15 and a decline in the use of condoms at "last sex" with a non-regular partner from 78% to 65% during the same period (NACA, 2015). Widening the age cohort to 15-49 this latter statistic dropped from 90% to about 82%, implying a drop in self-protection against infection. In 2001 it was reported that while 77% of respondents indicated that they ceased sexual intercourse upon infection with a sexually transmitted disease (STD), 17% reported continuing to engage in unsafe sex (CSO, 2002). In 2004, 9% indicated that they continued having unprotected sex while on treatment for sexually transmitted infections (STIs) (NACA and CSO, 2005). One study (Schaan et al., 2014) documents that among women living with HIV in Botswana, the proportion of those who became pregnant increased from 49.5% in 2007 to 69% in 2009 and to 70% in 2011. The question is, does the above suggest a pattern for fatal behaviour as theoretically predicted (Kremer, 1996; Sterck, 2013). Answers to such questions form an important part in assisting response measures to curb the spread of the virus.

We find that those less concerned about the likelihood of being infected reduce the likelihood of condom use and abstinence. While results show that majority of the respondents were knowledgeable about HIV, sexual behavior is not responsive to HIV prevalence. This could be a cause for concern if there are unobserved fatal behaviour, as the programmes such as ART may bring negative externalities. While we also establish that the effect of HIV on sexual behaviour could be indirect through changing perception, if the effect on behaviour is outweighed by the effect of concern, it might be the case that negative externalities could be evident. Hence, efforts should address these unintended consequences. The rest of the chapter is outlined as follows. Section 4.2 presents an overview of related literature, while data and descriptive statistics are presented in section 4.3. The empirical strategy is discussed in section 4.4 and results presented in section 4.5. Robustness checks are presented in section 4.6. Section 4.7 discusses the results in relation to literature, and 4.8 concludes.

4.2 Related Literature

4.2.1 Concept and Measurement of HIV Risk

HIV risk can be measured both objectively or subjectively and at either individual or population level (Guillen and Thuilliez, 2015).[13] Objective measures at population level include (i) HIV prevalence rate, which is "the proportion of those affected by the virus at any point in time"; (ii) incidence, which is "the number of new cases in a specified time" and (iii) mortality rates, the number of people who died as a result of the disease (Bonita et al., 2006:19-22). At an individual level, the objective measure is the confirmed HIV status (Godlonton and Thornton, 2013). On the other hand, subjective measures include the perceived likelihood of being infected and anxiety due to the disease, which are mainly measured at an individual level (Guillen and Thuilliez, 2015). Such perceptions are determined by several factors, including known previous sexual behaviour, perceived partner`s behaviour, and the existence of programmes/initiatives aimed at reducing the likelihood or severity of the objective risk, such as antiretroviral therapy and male circumcision (Godlonton et al., 2016; Wilson et al., 2014; Paula et al., 2014; Godlonton and Thornton, 2013; Delavande and Kohler, 2012; Smith and Watkins, 2004; Akwara et al., 2003; Gregson et al., 1998).

Efforts to objectively inform people about their HIV status are believed to be able to encourage less risky subsequent behaviour as this reduces uncertainty about one`s status (Delavande and Kohler, 2012). However, as Philipson and Posner (1993:41) argue, such a policy objective "assigns no weight to the utility that people derive from risky activities". As we discuss in the next sub-section, evidence has shown that the effect on behaviour may be affected by one`s prior belief about the likelihood of infection (see for example Boozer and Philipson, 2000). Similarly, although perceptions could be due to incorrect information (Maughan-Brown and Venkataramani, 2018; Delavande and Kohler, 2012), it is believed that higher perceived likelihood of infection may induce people to demand less risky sexual behaviour (Becker and Joseph, 1988; Tenkorang et al., 2011). In this study, we consider two measures of risk: HIV prevalence rate at the district level and individual`s perceived likelihood of infection as a result of the antiretroviral therapy in Botswana.

[13] Guillen and Thuilliez (2015) have an additional category of indirect risk measures. However, based on their definition of "variables that could directly influence a person`s perceived or objective risk of infection from the disease (pp 14), we restrict the classification to two: objective and perceived (or subjective).

Assessment of the risk-behaviour relationship is carried out within a framework which considers, in the case of HIV, two states: Susceptible and Infected (Hauck, 2018). Accordingly, individuals are either infected or susceptible to infection and may affect future infections through their behaviour (Kremer, 1996). As stated in the introduction, the effect on future infections has remained a debatable topic. The susceptibles may reduce the likelihood of infection through abstinence and condom use (Geoffard and Philipson, 1996; Philipson and Posner, 1993). The infected could reduce the severity of infection through strategies such as a reduction in the number of partners and enrolling in antiretroviral therapy, which reduces the probability of transmission rate. This framework is important to note in view of the third category of risk classification by Guillen and Thuilliez (2015). Thus, the sexual behaviour induced by antiretroviral therapy, for instance, could be due to the direct effect of reduced viral load among the infected or its perceived effect among the susceptibles.

4.2.2 Evidence on HIV Risk and Sexual behaviour

This chapter relates to the literature on the responsiveness of individuals to the risk of HIV infection. Whether the risk is objectively or subjectively measured, research efforts on the relationship between risk and sexual behaviour are faced with two main challenges (Hauck, 2018). Firstly, studies rely on self-reported sexual behaviour, which has been found to be susceptible to reporting bias, either under-reporting risky behaviour or over-reporting protective behaviour (Minnis et al., 2009; Akwara et al., 2003). Secondly, there is reverse causality between sexual behaviour and HIV risk. Thus, objectively measured risk (positive HIV status or higher prevalence rate) and perceived risk could have been due to previous risky sexual behaviour, and the risk affects sexual behaviour as well (Paula et al., 2014; Oster, 2012).

In what follows, we present an overview of empirical evidence distinguishing between objective and subjective measures of risk. Table B4.1 in the appendix provides other details. We start with objective risk measures: being in HAART (highly active antiretroviral therapy, an indicator for HIV positive status), HIV prevalence rate and individual HIV status. We then proceed to subjective risk measures: ART-induced perceptions, male circumcision-induced perceptions and perceptions informed by other factors such as knowing someone infected/dead from HIV. It must be noted that the risk classification may not be unambiguous. This is because objectively measured risk, such as HIV prevalence could affect perception about the likelihood of infection. It is for this reason that under subjective risk, we include studies that consider both

risks. Empirical methods have varied from behavioural experiments to the consideration of longitudinal data (see Table B4.1).

Objective risk indicators

A meta-analytic review (of studies for which 64% are in the US) by Crepaz, Hart and Marks (2004) finds no association between receiving highly active antiretroviral therapy (HAART) and unprotected sex as well as between undetectable viral load and unprotected sex. On the other hand, Lakdawalla et al. (2006) show that receipt of HAART in the US increases the number of sexual partners (by an additional 4). To avoid the problem of selection bias, Mechoulan (2007) considers testing for HIV and sexual behaviour in San Francisco and finds that risky behaviour (sex without condom) is more likely among testers than non-testers. In Botswana, enrolment in HAART has been found to have no association with risky behaviour (Keetile and Kgosidintsi, 2018). However, one potential threat to this study`s results is that assignment to HAART is not random; it is preceded by testing decisions, which have been found to have a mixed effect (Laxminarayan and Malani, 2011; Baird et al., 2014; Gong, 2015). Moreover, the inability to observe the exact time of being enrolled in the programme may lead to incorrect attribution of behaviour to HAART. While their results could be considered to suggest the absence of evidence on risk compensation in Botswana, issues surrounding their method should be the basis for carefully interpreting their results given that elsewhere accounting for endogeneity has a different effect (Lakdawalla et al., 2006; Mechoulan, 2007). Another study, which only focuses on women on ART finds that perceived stigmatization by both health workers and family members reduces the likelihood of childbearing in Botswana (Schaan et al., 2014).

Other studies consider HIV prevalence rate and sexual behaviour. For instance, in the US, Ahituv et al. (1996) find that increasing HIV prevalence led to increased use of condoms. Their results show that an increase in HIV by one per cent leads to about 8 per cent increase in the use of condoms. Using a cluster-group level data from the demographic and health surveys for 14 African countries (Burkina Faso, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Niger Senegal, Swaziland, Zambia, Zimbabwe) Oster (2012) finds that an increase in HIV prevalence reduces risky behaviour among married individuals and those with higher non-HIV life expectancy. To address the endogeneity of HIV prevalence at the country level, she employs an instrumental variable strategy, with the distance between each cluster and the country of HIV`s origin (Democratic Republic of Congo) as an instrument.

Two other studies in Botswana investigate the effect of having ever tested for HIV (Keetile, 2014; Malema, 2012). It should be noted that while testing for HIV might produce an objective measure of HIV risk, these two studies do not observe such test results in their analysis. Keetile (2014) finds that having ever tested for HIV/AIDS increases the likelihood of inconsistent condom use, while Malema (2012) finds that among females the employed (professionals and legislators, associate professionals and technicians) are less likely to use condoms than the unemployed, while among males only professionals and legislators are less likely to use condoms. The author partly attributes this finding to the observation that the employed reported having had more HIV tests. However, the absence of test results from these studies makes them to have inadequately addressed the question of how sexual behaviour responds to the risk of infection since they are threatened by potential omission bias. Moreover, given that not all respondents reported having sexual intercourse during the past 12 months prior to the survey, and that subsequent questions are conditional on such response, it is important to account for this censoring effect.

Subjective Risk Indicators

Another strand considers the impact of general perception/beliefs about the likelihood of being infected on sexual behaviour. However, others are specific about the source of perception, such as initiatives like ART and male circumcision. Hence our review starts with studies that consider ART induced perception, followed by those considering male-circumcision induced perception. We also include studies that considered both objective and perceived measures of risk with the view to demonstrate how perceptions may affect the effect of objectively measured risk on sexual behaviour. Lastly, we review those that consider the effect of knowing someone infected with HIV/AIDS and those that consider the general perception about the likelihood of being infected with HIV.

In Scotland, analysis of repeated (1996, 1992 and 2002) cross section data reveals that homosexual men who reported less anxiety about being infected with HIV as a result of improved treatments are more likely to engage in risky sexual behaviour (unsafe sex with casual partners and with more than one partner) (Hart and Williamson, 2005). When they split the sample into those who ever tested for HIV and those never, risky behaviour is evident among those who tested. In the non-tested sample, the incidence of unprotected sex is higher among those without treatment optimism. Similarly, a meta-analysis by Crepaz, Hart and Marks (2004) finds that people who believe HAART reduces HIV transmission and those less

concerned about engaging in unsafe sex were likely to engage in unprotected sex. However, the authors note that endogeneity between being in HAART (and belief about HAART) and sexual behaviour was not addressed in the studies they review. In emphasising this potential endogeneity, they note, "a person may come to strongly believe that HAART reduces HIV transmission after engaging in risky sex, in which case belief serves as a psychological coping mechanising to reduce stress" (pp234).

Two studies (Keetile & Letamo, 2015; Letamo et al., 2017) exogenously consider the impact of belief about the efficacy of ARV on condom use in Botswana have opposing results. While Keetile & Letamo (2015) find believing that ARV cure AIDS to be positively associated with an increased likelihood of inconsistent condom use, Letamo et al. (2017) find that those who believe that ARV cure HIV/AIDS (and those that were unsure) are less likely to report inconsistent condom use. Although the difference could be attributed to the fact that they used different samples, what is common in both is that they do not account for the endogeneity of belief.

Studies that consider the impact of male circumcision address the issue of self-selection through behavioural experiments. For instance, in Kenya, Wilson et al. (2014) find that believing that circumcision reduces the transmission risk (interacted belief-circumcision) reduces the number of partners and the likelihood of having multiple sexual partners. While they emphasize the issue of belief, they also find suggestive evidence on the increased number of sexual partners and the likelihood of multiple partners as well as less use of condoms. According to the authors, the latter could possibly be due to increased demand for circumcised partners since they are viewed to be less risky in transmitting HIV. Godlonton et al. (2016) find that in Malawi, randomly provided information that male circumcision reduces risk transmission induces uncircumcised men to reduce risky sexual behaviour (reduced number of sex partners, increased use of condoms), increases the likelihood of their willingness to have their male dependents circumcised and that circumcised men updated their beliefs about the risk reduction effect of male circumcision.

Two studies consider the effect of male circumcision on sexual behaviour in Botswana, both using a logistic regression approach (Balekang and Dintwa, 2016; Agiya and Letamo, 2011). Balekang and Dintwa (2016) find that compared to those with tertiary education, non-circumcised men with secondary education or below (including no education) are less likely to

have more than one sexual partner, while never married uncircumcised men are more likely to report inconsistent use of condoms and have sex under the influence of alcohol than the ever married. Primary education is only positively associated with an increased likelihood of having more than one sexual partner among circumcised men. On the other hand, using the same survey data, Agiya and Letamo (2011) find no evidence that circumcision affects the use of condoms in Botswana. Both studies do not account for the possible selection bias. Two qualitative studies (Sabone et al., 2013; Katisi and Daniel, 2015) reveal cultural factors and perceptions as prohibiting the impact of circumcision in Botswana, suggesting that its effect cannot be assessed by a comparative approach that has so far been used.

As indicated, some simultaneously consider both objective and subjective risk measures. Oster (2012) finds that individuals` risk perception measured through their knowledge on HIV transmission and related misconceptions has no effect on sexual behaviour in the 14 Sub-Saharan African countries. Thornton (2008) finds that in Malawi, those who learnt their HIV positive status are more likely to purchase condoms and on average, bought two more condoms than those who do not know their status. However, she notes that "condom purchases may not reflect the true demand for safe sex; if knowledge of HIV status increases abstinence, the demand for condoms could fall in response to obtaining test results" (pp1846). On the other hand, there is no effect among those who learnt their HIV negative status as well as on beliefs about the likelihood of being infected. Still in Malawi, Delavande and Kohler (2012) find that learning of HIV positive status reduces the likelihood of risky behaviour and no effect is observed for either the reported number of sexual partners or condom use among the HIV negative individuals. However, they also find that there were some individuals informed of their positive status in 2004 who expressed a low likelihood of being infected by HIV in 2006.

On the other hand, in Kenya and Tanzania, Gong (2015) find that the effect of testing for HIV is evident among those surprised by their test results. Those surprised by HIV positive test results increase their risky sexual behaviour, while those surprised by negative HIV test reduced their risky behaviour, a pattern also previously observed in Malawi (Baird et al. 2014). Accounting for the possibility that beliefs about the likelihood of being infected with HIV can be updated over time, Paula et al. (2014) through the semi-parametric panel estimator, find that in Malawi, men who believe they have a higher likelihood of HIV infection are less likely to report either being in extramarital sex relationships or having multiple sex partners, while those with low belief are more likely to report risky behaviour.

We end the subsection with those that consider the effect of factors such as knowing someone living with or dead from HIV as well as the general perception about the likelihood of being infected by the virus. In the US, Francis (2008) finds that having a relative (living or dead) with HIV/AIDS reduces the likelihood of reported homosexual behaviour more among men than among women. His findings indicate that men were 4.8 percentage points less likely to report not only being attracted to women, while women were 6.2 percentage points more likely to report not only being sexually attracted to men. In South Africa, Tenkorang et al. (2011) find that knowing someone who died of HIV reduces the likelihood of risky sexual behaviour among men and increases it among women.

In Kenya, a study that compares the effect of the randomly provided information on risk avoidance to information on the relative risk of being infected with HIV to teenagers found that risk reduction information led to a reduction in teenage pregnancy, while risk avoidance has no effect (Dupas, 2011). This is despite the observation that risk avoidance information improved the extent to which HIV education was taught at schools. Moreover, self-reported sexual behaviour reveals a pattern in which teenagers appear to have shifted from older partners to their age mates. While the provision of relative risk information also led to an increase in sexual activity, this is coupled with a reduction in unsafe sex activities.

The South African study also found that those who perceive themselves to be more susceptible to infection are less likely to report risky behaviour (Tenkorang et al., 2011). The study considers reported behavioural change between the first (2002) and third (2005) wave and form a transitional risk category variable for those who were sexually inactive in 2002. In Kenya, Akwara et al. (2003) find that higher perceived risk is positively associated with less risky behaviour and is stronger among men than women. However, the observation that men underestimated their likelihood might have driven the results. For instance, while 45% of men compared to 11% for women experienced risky sex 12 months before the survey, 5% of men (compared to 10% for women) perceived themselves to be at greater risk of infection (pp395-396).

In sum, as shown above, evidence on the effect of risk on sexual behaviour is mixed. Specifically, for Botswana, no study considers endogeneity between risk and sexual behaviour in assessing behavioural responsiveness to both objective and subjective risk of infection. We also observe that previous studies only consider those who report sexual activity, suffering

from potential selectivity bias, which we address by considering those who report sexual inactivity. Our analysis adds to the literature that addresses endogeneity between sexual behaviour and objectively measured risk (Oster, 2012) and updated perceived risk (Paula et al., 2014, who only considered men). As our data is not panel, for the perceived risk, we are unable to adopt the methods used by Paula et al. (2014).

4.3 Data and Descriptive Statistics

We use individual data from three of the four waves of the Botswana AIDS Impact Survey (BAIS) conducted by the National AIDS Coordinating Agency and Statistics Botswana (2004, 2008 and 2013). The first wave (2001) is not used because it has no information on the risk measures we consider. These surveys aim to provide statistical information to gauge the extent of knowledge on HIV/AIDS and related diseases to inform necessary interventions. They collect information on individual demographics, alcohol consumption and drug use, historical sexual behaviour, male circumcision and reported incidents of STDs, knowledge about HIV/AIDS, attitudes towards those affected by the virus, childbearing and antenatal care (for women only) as well as the availability of social and medical services (NACA and CSO, 2005; CSO and NACA, 2009; Statistics Botswana and NACA, 2016). Noteworthy is that under the sexual history and behaviour component, while respondents were asked about the number of people with whom they had a sexual relationship in the past 12 months before the survey, some information (such as the type of relationship, partner`s age difference, use of a condom in the first and last time of intercourse) were only collected for the last three sexual partners. In addition, the measure of the perceived risk (respondents` level of concern about the likelihood of being infected with HIV due to ART) is feasible only with the 2008 and 2013 waves.

The use of any data on such a sensitive subject needs to be acknowledged. There are practical issues, including variation in sample composition, but also the nature of the questions asked, and respondents' sensitivity in answering (Glick and Sahn, 2008). These are relevant for the Botswana surveys. First, they are repeated cross sections not longitudinal panels, with different sampling frames over time. Second, sensitivity in individual responses to personal questions is likely to change over time, possibly due to changes in publicity and respondent`s experience. For instance, the 2001 survey report indicates that "two questions related to sexual behaviour, especially sexual partners, were dropped due to the high rate of refusals experienced" (CSO, 2002:22). This suggests that some sensitive questions may have produced non or unreliable

responses. To control for the possible vertical transmission of HIV (Oster, 2005), we restrict the sample to those aged 15 years and above.

Tables 4.1 and 4.2 present descriptions of variables considered and descriptive statistics ,respectively. Across all the waves, the majority of respondents (more than 50%) are females and on average in their early thirties. The sample is dominated by young[14] people with those aged between 15 and 34 accounting for the largest share of the sample. The pattern for education is also similar across the waves; almost half have secondary education, followed by those with primary education. As for religion, the majority are Christians. The never married account for more than 50% across the three waves. Cohabiting partners account for the second-highest share. While more than 80% report having had sexual intercourse, a lower proportion report being sexually active during the 12 months before the survey.

Conditional on reporting being sexually active in the past 12 months before the survey, information was sourced on the three most recent partners. For those that report being sexually active, a clear majority report only one partner. We only report responses to questions regarding the first most recent partner. Table 4.2 shows that the relationship type is dominated by *boyfriend/girlfriend* and the lowest share is accounted for by *other relationship* type (for example a casual acquaintance or someone whom the respondent may have paid for sex). The reported use of a condom at first and last sex with the partner appears to have increased, although the share always using a condom appears to have declined. The latter indicator may be preferable to capture the frequency of use between first and last sex (Tenkorang et al., 2011). However, we consider both sets of responses to gauge consistency in reporting. This is because it might be easier for the respondent to over-report the use of condoms as established in the literature (Minnis et al., 2009). There appears to have been a notable increase in concern about HIV between 2008 and 2013. In 2008 42 per cent reported that they have not changed their level of concern about the likelihood of HIV since the introduction of ART and 26 per cent were more concerned. However, in 2013, 34 per cent were more concerned (and 50 per cent did not change the level of concern).

[14] According to the Revised national youth policy for Botswana "youth refers to persons of ages 15-35 years" (Republic of Botswana, 2010c:5)

The HIV prevention knowledge variable is constructed by adding the binary responses for two questions: (i) can people reduce their chances of getting HIV by using condoms correctly every time they have sex, and (ii) can people reduce their chances by having only one uninfected partner who has no other partners. A similar approach is used for the HIV misconceptions for the four questions (whether it is possible for a healthy-looking person to have HIV, HIV being transmitted through mosquitoes, sharing a meal with an infected person and through witchcraft), with a core of 1 for each misconception rejected. In both cases, scores range from zero to 2 for the first and zero to 4 for the second set of questions. Zero implies either no correct HIV prevention knowledge or no true rejection of the misconceptions. Although these questions were only asked of those who indicated ever having heard of HIV, that represents 97% of the sample. For HIV prevention knowledge, Table 4.2 shows a mean score of 1.8 with a mean reject misconceptions score of 3; respondents had good prevention knowledge and were alert to misconceptions.

Table 4.1: Description of Variables

Variable	Description
Dependent Variables	
Had sex relations	Whether the individual reports sexually active 12 months before survey
Sexual partners	Indicators for the reported number of sexual partners in the past 12months 0 for no partners, 1 for 1 partner and 2 for more than 1 partner
CUFS	Indicator for condom use at first sex (1=Yes)
CULS	Indicator for condom use at last sex (1=Yes)
Explanatory Variables	
Age	Age in years
Age squared	Age in years squared
Male	An indicator for a male respondent (1=Yes)
Education	
No education	An indicator for no education (1=Yes)
Non formal	An indicator for non-formal education (1=Yes)
Primary	An indicator for primary education (1=Yes)
Secondary	An indicator for secondary education (1=Yes)
Higher	An indicator for higher education (1=Yes)
Christianity	An indicator for Christianity as main religion (1=Yes)
Current Marital Status	
Married	Whether an individual is married (1=Yes)
Never married	Whether an individual has never been married (1=Yes)
Cohabitation	Whether an individual is cohabiting (1=Yes)
Other marital status	Whether an individual is widowed, divorced, separated (1=Yes)
Area of residence	
Rural	An indicator for rural area residence (1=Yes)
Relationship to the most recent partner	
Husband/Wife	Whether the individual is husband/wife to the most recent sexual partner (1=Yes)
Cohabitation	Whether the individual is cohabiting with the most recent sexual partner (1=Yes)
Girlfriend/Boyfriend	Whether the individual is a boy/girl friend to the most recent sexual partner (1=Yes)
Other relation type	Indicator for other relationship type (casual acquaintance, paid sex) with most recent sexual partner (1=Yes)
Perceived Unfaithfulness	Whether an individual believes the partner has other partners (1=Yes)
Partner drunk at last sex	Whether partners were drunk at last sex (1=Yes)
Personal concern on likelihood due to ART changed	
No change	No change in the level of concern about the likelihood of HIV infection
Less concerned	Less concerned about the likelihood of HIV infection (1=Yes)
More concerned	More concerned about the likelihood of HIV infection (1= Yes)
HIV prevalence rate	Percentage shares of those HIV positive in a district
Prevention Knowledge	Sum of two dummies on HIV prevention
Correctly reject misconceptions	Sum of four dummies on the true rejection of HIV misconceptions
Instrumental Variables	
Distance	Distance to district with highest HIV prevalence rate
ARVMBS	Number of months for ART rollout before the survey

Table 4.2: Descriptive Statistics

Variable	2004		2008		2013		Pooled	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A. Demographics								
Male	0.450	0.497	0.445	0.497	0.453	0.498	0.449	0.497
Age in years	31.652	12.463	32.445	12.464	33.243	12.520	32.312	12.494
15-24 (1 if Age is between 15 and 24)	0.357	0.479	0.321	0.467	0.294	0.456	0.329	0.470
25-34 (1 if Age is between 25 and 34)	0.284	0.451	0.302	0.459	0.299	0.458	0.294	0.456
35-44 (1 if Age is between 35 and 44)	0.179	0.383	0.184	0.387	0.202	0.401	0.186	0.389
45-54 (1 if Age is between 45 and 54)	0.116	0.320	0.125	0.331	0.130	0.336	0.122	0.328
55-64 (1 if Age is between 55 and 64)	0.064	0.245	0.068	0.252	0.076	0.265	0.068	0.252
Education Level								
No Education	0.136	0.342	0.118	0.322	0.087	0.281	0.118	0.322
Non-Formal	0.009	0.093	0.021	0.145	0.006	0.079	0.013	0.113
Primary	0.250	0.433	0.211	0.408	0.180	0.384	0.219	0.414
Secondary	0.466	0.499	0.493	0.500	0.524	0.499	0.490	0.500
Higher	0.140	0.347	0.156	0.363	0.203	0.402	0.160	0.367
Main Religion								
Christianity	0.808	0.394	0.703	0.457	0.862	0.345	0.781	0.414
Current Marital Status								
Married	0.187	0.390	0.171	0.377	0.189	0.391	0.181	0.385
Never Married	0.537	0.499	0.531	0.499	0.532	0.499	0.533	0.499
Cohabitation	0.225	0.418	0.251	0.433	0.244	0.429	0.239	0.426
Other Marital Status	0.051	0.220	0.047	0.212	0.036	0.185	0.046	0.210
Residence								
Rural	0.377	0.485	0.388	0.487	0.364	0.481	0.379	0.485
Panel B. Behaviour								
Ever had sexual intercourse (1=Yes)	0.830	0.376	0.855	0.352	0.861	0.346	0.846	0.361
Had sex in the past 12 months	0.669	0.471	0.688	0.463	0.829	0.376	0.709	0.454
No of sexual partners								
One Partner (1=Yes)	0.883	0.322	0.851	0.356	0.855	0.352	0.864	0.343
Two partners (1=Yes)	0.087	0.282	0.114	0.317	0.093	0.291	0.099	0.298
Three or more (1=Yes)	0.030	0.172	0.036	0.185	0.051	0.221	0.037	0.190
Relation to the first most recent sexual partner								
Husband/Wife	0.248	0.432	0.200	0.400	0.238	0.426	0.228	0.420
Cohabiting	0.281	0.450	0.328	0.469	0.289	0.453	0.300	0.458
Girlfriend/Boyfriend	0.449	0.497	0.450	0.498	0.442	0.497	0.448	0.497
Other relationship type	0.021	0.143	0.022	0.148	0.032	0.175	0.024	0.153

Behaviour related to sex with the first most recent sexual partner								
Always use condom (1=Yes)	0.872	0.334	0.609	0.488	0.591	0.492	0.678	0.467
Condom used at first sex with partner	0.619	0.486	0.766	0.423	0.823	0.382	0.722	0.448
Condom used at last sex with partner	0.600	0.490	0.699	0.459	0.692	0.462	0.658	0.474
Partner drunk at last sex	0.062	0.241	0.106	0.307	0.065	0.247	0.078	0.269
Perceived partner unfaithfulness	0.223	0.416	0.260	0.438	0.271	0.444	0.248	0.432
Panel C. Knowledge on HIV and level of access to interventions								
Ever heard of HIV (1=Yes)	0.956	0.204	0.973	0.161	0.979	0.142	0.968	0.176
HIV Prevention knowledge	1.741	0.538	1.773	0.505	1.844	0.419	1.777	0.502
Correctly rejects misconceptions	2.780	1.131	3.024	1.038	3.110	0.995	2.948	1.075
Personal concern on likelihood due to ART changed[15]								
No change			0.419	0.493	0.502	0.500	0.452	0.498
Less concerned			0.320	0.467	0.162	0.368	0.257	0.437
More concerned			0.260	0.439	0.336	0.473	0.290	0.454
Observations	**12664**		**12035**		**7249**		**31948**	

[15] This is only for 2008 and 2013. Total number of observations is 17333

Table 4.3: Mean differences by Sexual activity report

	Had no sex in the past 12m	Had sex in the past 12m	Differences (No-Yes)
Male	0.424	0.455	-0.031***
Age	31.246	33.329	-2.083***
No education	0.154	0.107	0.047***
Primary	0.240	0.215	0.025***
Secondary	0.509	0.468	0.041***
Higher	0.083	0.197	-0.114***
Christianity	0.783	0.775	0.008
Married	0.066	0.236	-0.170***
Never Married	0.766	0.419	0.347***
Cohabitation	0.064	0.321	-0.257***
Other Marital Status	0.104	0.025	0.079***

Notes: Means based on the pooled sample. ***significant at the 1 per cent level

Table 4.2 shows that about 70% report having had sexual intercourse in the past 12 months. The surveys do not source information on the reasons behind sexual inactivity. Table 4.3 compares means for selected variables according to sexual activity: the proportion of males (slightly), highly educated, married and cohabiting individuals is higher for those who report having sexual relations in the past 12 months than those who report no sex. While we do not establish the reasons, the proportion of never married and less than higher education reporting no sexual activity is significantly higher than the sexually active proportion.

4.4 Model Specification

4.4.1 Theoretical

The general theoretical/conceptual framework within which we conduct this investigation is based on Philipson and Posner (1993), who challenged the epidemiological view of behavioural responses to prevalence. According to this framework, HIV incidence increases during the early stages of prevalence since the *subjective* probability of infection is arguably low. However, as prevalence increases over time, the cost of risky sexual behaviour such as unprotected sex rises, resulting in a possible reduced demand for risky sexual activities. This is because increasing prevalence can depreciate sexual capital, which, according to Michael (2004: 644) is the "present value of the flow of benefits from sexual enjoyment over the remaining lifetime". With the objective to maximise expected utility from sexual activities, an individual runs the risk of contracting HIV. Probability of infection from one sexual act is presented as follows (Michael, 2004):

$$P_r(1) = PIE \tag{4.1}$$

In (4.1), *P*, *I*, and *E* are the probability that the partner is infectious, the infectivity rate of the disease, and the effectiveness of the practice in preventing disease transmission, respectively. Should the individual be involved in *N* sexual activities with different partners (*S*), the probability of being infected becomes:

$$P_r(N,S) = 1 - (1 - PIE)^N \tag{4.2}$$

Key in this framework is that behavioural decisions can minimise the probability of contracting HIV (4.2). Hence, based on Philipson and Posner (1993), we investigate whether and how individuals respond to the risk of contracting HIV/AIDS through their behaviour. Behaviour, in this case, includes both risky and non-risky actions which can be measured through for instance rate of partner change (Kremer, 1996) as well as use and non-use of condoms (Geoffard and Philipson, 1996). The framework also provides for the effect of perceived likelihood of infection, which in our case is due to the existence of ART. Following Paula et al. (2014), the perception/concern about the likelihood of being infected with HIV due to ART can be represented as follows.

$$C_1 = C_0 + (1 - C_0)h(ART) \tag{4.3}$$

Where C_0 and C_1 are the previous and updated levels of concern, respectively, and *h(ART)* captures the fact that the source of concern is ART.

4.4.2 Empirical Strategy

Following Oster (2012), our basic model is specified as follows:

$$S_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{4.4}$$

In (4.4) S_i is sexual behaviour (measured as indicators for sexual partners[16] and reported condom use at first and last sex), X_i is a vector of controls. Given the binary nature of reported condom use, we specify a probit regression model (Greene, 2008) in equation (4.5) below. In this case the dependent variable is equal to 1 when the respondent reports the use of condom and 0 otherwise.

[16] Our consideration of this variable is in line with what Francis (2008) refers to as economics of sexuality. He argues that this approach differs from the biological theories according to which sexual behaviour is exogenous.

$$Pr(Si = 1) = \beta_0 + \beta_1 X_i + \varepsilon_i > 0 \quad (4.5)$$

Previous work that examines sexual behaviour only considers those who report being sexually active. However, given the possible underreporting of risky behaviour and over reporting of protective behaviour (Corno and Paula, 2019; Gong, 2015; Paula et al., 2014; Tennekoon and Roseman, 2014; Minnis et al., 2009; Palen et al., 2008), consideration of only those that report being sexually active may bias the results. The decision to engage in a sexual activity may be considered as participating in a market for sexual partners (Philipson and Posner, 1995;1993; Glick and Sahn, 2008). Hence, the potential unobservable effect surrounding the reported sexual inactivity (i.e abstinence) should be explored. It is worth noting that unlike elsewhere, in Botswana there is non-use of biomarkers to validate responses provided by respondents[17]. To address the potential selection bias, therefore, we adopt two approaches. Firstly, we estimate an ordered probit model for the number of sexual partners, which allows us to include those who report having abstained from sexual relations (Glick and Sahn, 2008). Thus, the dependent variable differentiates those with zero, one and more than one reported sexual partner. The reason to have indicator variables for the number of sexual partners is to minimise the impact of errors that may have been brought by bias in reporting the true number of sexual partners (Oster, 2012). The specified ordered probit model may be obtained from a latent variable model (Greene, 2008; Wooldridge, 2010). In our case if we have latent variable

$$S^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon, \qquad \varepsilon|x \sim N(0,1) \quad (4.6)$$

We observe the following:

$$\begin{aligned} &S = 0 \ \ if \ \ S^* \leq 0 \\ &S = 1 \ \ if \ \ 0 < \ \ S^* \leq \alpha_1 \\ &S = 2 \ \ if \ \ \alpha_1 < \ S^* \leq \alpha_2 \end{aligned} \quad (4.7)$$

Where αs are parameters (cut points) to be estimated. The standard normality assumption for the error term, ε, makes it possible for the derivation of conditional distribution of S given the

[17] The use of biomarkers is not however, without non-monetary costs. A negative test for an STI for instance, does not mean there was no risky behaviour (Corno and Paula, 2019; Paula et al. 2014).

explanatory variables. Such is achieved by obtaining the following response probabilities (Wooldridge, 2010):

$$P(S=0|\mathbf{x}) = P(\mathbf{x}\boldsymbol{\beta} + \varepsilon \leq \alpha_1|\mathbf{x}) = \Phi(\alpha_1 - \mathbf{x}\boldsymbol{\beta})$$
$$P(S=1|\mathbf{x}) = P(\alpha_1 < S^* \leq \alpha_2|\mathbf{x}) = \Phi(\alpha_2 - \mathbf{x}\boldsymbol{\beta}) - \Phi(\alpha_1 - \mathbf{x}\boldsymbol{\beta}) \quad (4.8)$$
$$P(S=2|\mathbf{x}) = 1 - \Phi(\alpha_2 - \mathbf{x}\boldsymbol{\beta})$$

The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are obtained through the maximum likelihood method, with the following individual likelihood function.

$$li\ (\boldsymbol{\alpha}, \boldsymbol{\beta}) = 1[S_i = 0]\log[\Phi(\alpha_1 - \mathrm{x}_i\boldsymbol{\beta})] + 1[S_i = 1]\log[\Phi(\alpha_2 - \mathrm{x}_i\boldsymbol{\beta}) - \Phi(\alpha_1 - \mathrm{x}_i\boldsymbol{\beta})] +$$
$$1[S_i = 2]\log\,[1 - \Phi(\alpha_2 - \mathrm{x}_i\boldsymbol{\beta}) \quad (4.9)$$

Our interest, however, is in the marginal effects, computed as follows:

$$\frac{\partial P(S=0|\mathrm{x})}{\partial \boldsymbol{x}} = -\beta\phi(\alpha_1 - \mathbf{x}\boldsymbol{\beta})$$

$$\frac{\partial P(S=1|\mathrm{x})}{\partial \boldsymbol{x}} = \beta[\phi(\alpha_2 - \mathbf{x}\boldsymbol{\beta}) - \phi(\alpha_1 - \mathbf{x}\boldsymbol{\beta})] \quad (4.10)$$

$$\frac{\partial P(S=2|\mathrm{x})}{\partial \boldsymbol{x}} = \beta\phi(\alpha_2 - \mathrm{x}\boldsymbol{\beta})$$

Secondly, we estimate the following selection model for robustness

$$Si = x_i\beta + \varepsilon_1 > 0$$

$$SR_i = 1(z_i\alpha + \varepsilon_3 > 0) \quad (4.11)$$

Where SR_i is the sex report by the respondent, x_i are variables that affect sexual behavior and z_i are those that affect reporting sexual activity. In this model, we test the hypothesis that there is no relationship between the unobservables that affect the decision to report being sexually active and the unobservables that affect the decision to use condoms (Cameron and Trivedi,

2010). If such a relationship which is represented by *rho* (the relationship between the error terms, ε_1 and ε_3) is statistically significant, then failing to account for it may bias the results.

We use religion for identification. As surveys in Botswana have shown that majority are christians (even the BAIS), we argue that religion could affect the likelihood of participating in sexual activity but may not have direct effect on the use of condoms. It may discourage some behaviours such as premarital intercourse and promiscuous relations. Using data for 38 sub Saharan African countries, Gray (2004) finds that the proportion of Muslims in such countries is negatively associated with HIV prevalence. This among others, has been attributed to adhering to some tenets of the Muslim religion.

Empirically, evidence on whether individuals respond to HIV risk is complicated by the challenge of reverse causality between these risks and sexual behaviour (Guillon and Thuilliez, 2015; Oster, 2012). Thus, HIV prevalence rates could be the outcome of more cases of risky sexual relations. Similarly, the reported sexual behaviour may also be due to the HIV prevalence rates. According to Oster (2012), inadequate control of such reverse causality may be one factor explaining the mixed evidence on the behavioural response to HIV. Guillon and Thuilliez (2015) state that HIV risk endogeneity has been addressed by some through the use of indirect measures of risk such as knowing someone infected by HIV (Tenkorang et al., 2011), having a relative with AIDS (Francis, 2008) and the type of relationship (Gerber and Berman, 2008). However, these factors (except for the case where the dead person is a relative) could be argued not to be exogenous. In our case, the endogenous variables are HIV prevalence rate (Oster, 2012) and the updated concern about the likelihood of HIV infection due to ART (Paula et al., 2014). To accommodate the endogeneity of objective risk (HIV prevalence rate), we follow Oster (2012) and consider a two-stage estimation (structural equation and reduced form equation) as follows:

$$S_i = \alpha_0 + \alpha_1 X_i + \alpha_2 HIV_i + \varepsilon_i \quad (4.13)$$

$$HIV_i = \alpha_0 + \alpha_1 X_i + \alpha_2 Z_i + u_i \quad (4.14)$$

Where *Si* is the reported sexual behaviour, measured by indicators of sexual partners and condom use at first and last sex. *X* represents a vector of explanatory variables and HIV is the

prevalence rate, while ε_i and u_i are error terms. On the other hand, addressing the endogeneity of perceived risk results in the following:

$$S_i = \alpha_0 + \alpha_1 X_i + \alpha_2 concerned_i + \varepsilon_i \quad (4.15)$$

$$Concerned_i = \alpha_0 + \alpha_1 X_i + \alpha_2 Z_i + v_i \quad (4.16)$$

In equations (4.14) and (4.16) Z_i is an instrumental variable, while *concerned* in (4.15) and (4.16) represents those who updated their concern level (either less concerned or more concerned) since the introduction of ART. While equation (4.14) can be considered an instrumental variable probit (IVProbit) model, such is not the case with equation (4.16) since the endogenous regressor is binary (Lewbel et al., 2012). For efficiency purposes, we estimate the IV probit model by maximum likelihood as opposed to the two-step procedure (Wooldridge, 2010). Equations 4.15 and 4.16 are estimated through the 2SLS method.

We instrument for HIV prevalence rate with distance (in kilometres) from the respondent`s area of residence to the single HIV hot spot (i.e area with the highest HIV prevalence rate in Botswana). This variable may not be directly related to sexual behaviour (Oster, 2012). It could be argued that migration may affect the validity of this instrument. For instance, a study in Kenya established that the effect of migration on AIDS varies by destination area and gender of respondents: migrants from urban to rural areas were more likely to engage in risky behaviour than non-migrants (Brockerhoff and Biddlecom, 1999). To address this, we control for rural/urban residence since in Botswana migration is mainly from rural to urban areas (CSO, 2009b), the latter consistently have higher HIV prevalence rates in all three waves (2004, 2008, 2013). We also estimate regressions for the rural and urban subsamples separately.

The perceived risk information was sourced with reference to the introduction of ART (provision of ARV drugs). In the two waves being considered, the question asked is "how has your personal concern about getting HIV changed since the introduction of ARV", with four broad responses of less concerned, more concerned, no change in concern level and do not know. Although no change could include several possibilities, in particular, still concerned and still not concerned, we combine them as our focus is on the impact of having reported an update of one`s perception. The way respondents theoretically formulate their levels of concern is beyond the scope of this chapter. Indeed, some literature on health risk perception documents

complexities of risk formulation. It is acknowledged generally that risk perception may be individualistic or context-specific (Ferrer and Klein, 2015; Waters et al., 2014). This notwithstanding, we maintain that such is the basis for accounting for endogeneity of risk perception. In our context, theoretically, the effect of ARV on behaviour remains ambiguous (Auld, 2003; De Walque et al., 2012).

Therefore, we use ARV programme rollout dates in various districts to instrument for the perceived risk. At its inception, the programme was first rolled out in 2002 in 4 sites, 6 in 2003 and 19 in 2004 (Republic of Botswana, 2009). In the context of our data, rollout dates were before the first national AIDS impact survey that conducted blood tests. The earlier the rollout, the more we expect the extent of knowledge on the existence of ARV, and its impact on risk perception, although the direction of effect would remain ambiguous. Moreover, we do not expect this date to be directly correlated with sexual behaviour, but rather the perception about the likelihood of being infected with HIV. While it could be the case that its introduction may increase HIV prevalence rate, this may not necessarily be through incidence (new infection), but through prolonged life of the infected (Lakdawalla et al., 2006; Bonita et al., 2006). Assessing the impact on incidence will require the analysis to distinguish between old and new infections (Friedman, 2018). There could be a concern that the non-random placement of ARV facilities may introduce omitted variable bias. However, our examination of the previous HIV prevalence data does not suggest that placement was informed by prevalence. For instance, Selebi-Phikwe, which is arguably the hot spot (according to reports on HIV prevalence) is not among the first to have received the programme. To address a potential issue that earlier rollout could have been influenced by non-rurality of the districts, we control for rural-urban status to address such a concern (Friedman, 2018). We do not directly use rollout dates but consider the number of months before the survey.

The above notwithstanding, it could still be the case that our results would need careful interpretation because of the following. Firstly, we do not observe whether such an initiative was expected prior to its inception. Secondly, we do not observe the exact time during which the respondent accessed information on ARV (or the drug itself for those enrolled). While the perception is said to have been induced by the introduction of ARV, it could be related to the time during which the respondent heard of the programme or accessed it, information for which is not available in the data.

4.5 Results and Discussions

We present results separately for responsiveness to both the objective and perceived risk indicators. In both, we present ordered probit marginal effects, binary probit marginal effects and results when endogeneity issues are accounted for. Having established an association between education and HIV knowledge indicators through cross-tabulations (Table 4.4), we interact these two in our estimations. As seen in Table 4.4, HIV prevention knowledge and correct rejection of misconceptions, increase with increasing levels of education.

Table 4.4: Cross tabulation of knowledge and education variables

Correct Prevention Knowledge	**None**	**Primary**	**Secondary**	**Higher**
0	343(10.31)	315(4.75)	464(3.03)	62(1.23)
1	689(20.72)	1018(15.35)	2156(14.07)	523(10.40)
2	2294(68.97)	5300(79.90)	12701(82.90)	4443(88.37)
Total	3326	6633	15321	5028
Pearson Chi2 (2)	562.991***	22.570***	71.409***	212.108***
Correctly reject misconceptions	**None**	**Primary**	**Secondary**	**Higher**
0	320(9.66)	232(3.50)	113(0.74)	5(0.10)
1	899(27.14)	1083(16.34)	705(4.61)	100(1.99)
2	1034(31.22)	1980(28.87)	2330(15.23)	441(8.76)
3	747(22.55)	1900(28.66)	5083(33.22)	1324(26.30)
4	312(9.42)	1434(21.63)	7071(46.21)	3164(62.85)
Total	3312	6629	15302	5034
Pearson Chi2 (4)	3400***	1700***	1800***	1700***

Notes: ***Significant at the 1 per cent level

4.5.1 Objective risk and sexual behaviour

Table 4.5 presents ordered probit marginal effects by the number of sexual partners for HIV rate, which we use to proxy for objective risk. Although the HIV prevalence rate carries a negative sign for one or more partners, its effect is statistically insignificant. Males are 7.9 percentage points less likely to report no partner and about 3 percentage points more likely to report more than one sexual partner than females. A similar pattern is observed for education; compared to those with no education, being educated increases the likelihood of having one or more partners.

Christianity is associated with an increase in the likelihood of abstinence and reduction in the likelihood of an increased number of partners. Compared to married individuals, cohabitation is associated with an increase in the likelihood of participating in sexual relations with one or more partners. HIV prevention knowledge and correct rejection of HIV misconceptions are associated with a reduced likelihood of more than one sexual relation. Only the interaction of

primary education and correctly rejecting HIV misconceptions is significant with a negative coefficient suggesting that those with primary education who correctly reject misconceptions are less likely to have one or more partner (Table 4.5).

Table 4.5: Marginal Effects (at means) for indicators of sexual partners (Ordered Probit)

Variables	0 Partner	1 partner	>1 partner
HIV Rate	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Male	-0.079***	0.045***	0.033***
	(0.004)	(0.002)	(0.002)
Age	-0.052***	0.030***	0.022***
	(0.002)	(0.001)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.065***	0.037**	0.027***
	(0.025)	(0.014)	(0.010)
Secondary	-0.081***	0.047***	0.034***
	(0.021)	(0.012)	(0.009)
Higher	-0.095***	0.055***	0.040***
	(0.035)	(0.020)	(0.015)
Christianity	0.042***	-0.024***	-0.018***
	(0.005)	(0.003)	(0.002)
Cohabitation	-0.021***	0.012***	0.009***
	(0.008)	(0.005)	(0.003)
Never married	0.158***	-0.091***	-0.067***
	(0.012)	(0.007)	(0.005)
Other Marital status	0.294***	-0.170***	-0.124***
	(0.021)	(0.013)	(0.009)
Rural	0.011	-0.006	-0.005
	(0.009)	(0.005)	(0.004)
HIV Prevention Knowledge	-0.036***	0.021***	0.015***
	(0.011)	(0.007)	(0.005)
Corectlyrejectmis	-0.012**	0.007**	0.005**
	(0.006)	(0.003)	(0.002)
Primary*HIV Knowledge	0.009	-0.005	-0.004
	(0.011)	(0.006)	(0.005)
Secondary*HIV Knowledge	-0.012	0.007	0.005
	(0.013)	(0.008)	(0.006)
Higher*HIV Knowledge	-0.009	0.005	0.004
	(0.016)	(0.009)	(0.007)
Primary*Correcrejecmis	0.014**	-0.008**	-0.006**
	(0.007)	(0.004)	(0.003)
Secondary* Correcrejecmis	0.013	-0.007	-0.005
	(0.008)	(0.004)	(0.003)
Higher*Correcrejecmis	0.003	-0.002	-0.001
	(0.009)	(0.005)	(0.004)
Wave 3 (2008)	-0.017**	0.010***	0.007**
	(0.007)	(0.004)	(0.003)
Wave 4 (2013)	-0.078***	0.045***	0.033***
	(0.009)	(0.005)	(0.004)
District FE	Yes	Yes	Yes
Observations	29601	29601	29601

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, married for marital status. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table 4.6 presents the binary probit marginal effects for condom use at both first and last sexual intercourse. It shows that HIV prevalence rate increases the likelihood of condom use at first sexual intercourse, but by less than 1 per cent. Males are more likely to report the use of a condom at both times, and the likelihood of using a condom increases with increasing levels of education. The nature of relationship also matters; there is less likelihood of condom use among married and cohabiting partners, while "other relation type" (such as casual acquaintance or paying someone for sex) is associated with an increased likelihood of condom use at last sexual intercourse. The less use of condoms among the married and those cohabiting could partly be due to the confidence/trust in a relationship, or fatalism as predicted by Sterck (2013). A cross examination of marital status and partner perceived unfaithfulness shows that within the married, 9% indicated that they do not trust their wives/husbands, while among the cohabiting 28% said they do not trust their partners. .

Those who perceive their partners to be unfaithful are more likely to report condom use at last sexual intercourse. However, given that the same variable (i.e perceived unfaithfulness) carries a negative sign for condom use at first sexual intercourse, this could suggest some risky behaviour. It could suggest that, depending on other factors such as the duration of the relationship (which we do not observe), perceived unfaithfulness has a positive compensation effect at a later period. While HIV knowledge variables increase the likelihood of condom use, results also show that when interacted with the education variables, the likelihood of condom use is reduced. Those who report having had sex when drunk are less likely to use a condom at last sex.

Table 4.6: Probit Marginal Effects for condom use

Variables	CUFS	CULS
	(1)	(2)
HIV Rate	0.004**	0.003
	(0.002)	(0.002)
Male	0.052***	0.067***
	(0.007)	(0.006)
Age	-0.017***	-0.003
	(0.002)	(0.002)
Age squared	0.000***	-0.000
	(0.000)	(0.000)
Primary	0.096**	0.083**
	(0.041)	(0.038)
Secondary	0.221***	0.345***
	(0.050)	(0.050)
Higher	0.336***	0.338***
	(0.057)	(0.060)
Christianity	0.012	0.012
	(0.008)	(0.011)
Husband/Wife	-0.225***	-0.265***
	(0.007)	(0.016)
Cohabiting	-0.085***	-0.125***
	(0.007)	(0.008)
Other relationship type	0.017	0.088***
	(0.020)	(0.029)
Perceived Unfaithfulness (Yes)	-0.013**	0.042***
	(0.005)	(0.004)
Rural	-0.049***	-0.058***
	(0.004)	(0.008)
One Partner	0.014	-0.014
	(0.009)	(0.009)
PK	0.045**	0.066***
	(0.018)	(0.017)
CRM	0.023***	0.027**
	(0.006)	(0.011)
Primary*PK	-0.003	0.009
	(0.020)	(0.022)
Secondary*PK	-0.020	-0.077***
	(0.024)	(0.019)
Higher*PK	-0.040	-0.046
	(0.029)	(0.029)
Primary*CRM	-0.013*	-0.008
	(0.008)	(0.011)
Secondary*CRM	-0.014**	-0.020
	(0.006)	(0.013)
Higher*CRM	-0.025***	-0.032**
	(0.009)	(0.014)
Partner drunk at last sex	-	-0.094***
		(0.007)
Wave 3(2008)	0.095***	0.066***
	(0.006)	(0.010)
Wave 4 (2013)	0.173***	0.066***
	(0.010)	(0.011)
District FE	Yes	Yes
Observations	20493	20475
Loglikelihood	-8341.716	-11057.096

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relation type. CUFS: Condom use at first sex, CULS: Condom use at last sex.PK: HIV Prevention Knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Tables 4.7 and 4.8 show results when endogeneity of HIV prevalence is accommodated. While the main results are for the IV probit model, we also report the 2SLS results for robustness. Worth noting is that columns 3 and 4 of Table 4.8 shows the rescaled coefficients from an approximate conversion factor (Amemiya, 1981). Specifically, the 2SLS coefficients have been multiplied by 2.5 for the purpose of comparison following the established relation between coefficients for linear and probit models (Amemiya, 1981; Madala, 1983). This is because an application of linear regression on a binary dependent variable has a drawback that the fitted probabilities can be outside the 0-1 range, contrary to the probability requirements (Lewbel et al., 2012; Amemiya, 1981).

In Both IV Probit and 2SLS, we assess the relevance and strength of the instrument as well as potential endogeneity of HIV prevalence. Regarding the relevance of the considered instrument, it is evident in Table 4.7 that the considered instrument (distance to HIV hot spot) is statistically significant, implying that it is relevant to the endogenous variable. In testing for the strength of the instrument, the null hypothesis is that the instrument is weak. We obtain an F statistic of 13.323, which suggests a non-weak instrument since it is above the threshold of 10 (Staiger and Stock, 1997). In the case of the endogeneity test, the null hypothesis is that the variable is exogenous. The Wald test from IV Probit yields Chi2 (1) = 4.67, with the P>Chi2 =0.0307 implying that we reject the null hypothesis of exogenous HIV prevalence rate (Cameron and Trivedi, 2010). For 2SLS we consider the robust Durbin-Wu Hausman test, which yields an F (1,25) =3.082, P=0.0914, also leading to the rejection of the null hypothesis, although at 10 per cent level (Cameron and Trivedi, 2010).

Table 4.8 shows that the effect of HIV prevalence rate on condom use is insignificant. Being male or more educated increases the likelihood of condom use while there is less use among married and cohabiting individuals. Other relationship types and perceived partner unfaithfulness increase the likelihood of condom use at last sex. Compared to those who report more than one partner during the past 12 months, those with one partner are less likely to report the use of a condom at last sex only. Similarly, alcohol consumption reduces the likelihood of the reported use of condoms.

Table 4.7: First stage results for objective risk (HIV Prevalence)

Variables	IV Probit		2SLS	
	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
Male	0.178**	0.179**	0.178**	0.179**
	(0.081)	(0.082)	(0.081)	(0.082)
Age	0.037	0.037	0.037	0.037
	(0.032)	(0.032)	(0.032)	(0.032)
Age squared	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	-0.016	-0.034	-0.016	-0.034
	(0.319)	(0.319)	(0.319)	(0.319)
Secondary	0.600	0.593	0.600	0.593
	(0.651)	(0.654)	(0.651)	(0.654)
Higher	0.440	0.440	0.440	0.440
	(0.443)	(0.445)	(0.443)	(0.445)
Christianity	0.223	0.216	0.223	0.216
	(0.140)	(0.139)	(0.140)	(0.139)
Husband/Wife	0.152	0.150	0.152	0.150
	(0.125)	(0.127)	(0.126)	(0.127)
Cohabiting	0.331***	0.334***	0.331***	0.334***
	(0.114)	(0.114)	(0.114)	(0.115)
Other relationship type	0.346**	0.358**	0.346**	0.358**
	(0.173)	(0.171)	(0.173)	(0.171)
Perceived Unfaithfulness (Yes)	0.213***	0.217***	0.213***	0.217***
	(0.076)	(0.077)	(0.076)	(0.077)
Rural	-0.338	-0.342	-0.338	-0.342
	(0.399)	(0.398)	(0.399)	(0.398)
One Partner	-0.156**	-0.163**	-0.156**	-0.163**
	(0.072)	(0.072)	(0.072)	(0.072)
PK	-0.181	-0.182	-0.181	-0.182
	(0.114)	(0.114)	(0.114)	(0.115)
CRM	0.047	0.046	0.047	0.046
	(0.089)	(0.090)	(0.089)	(0.090)
Primary*PK	-0.008	-0.002	-0.008	-0.002
	(0.147)	(0.147)	(0.147)	(0.147)
Secondary*PK	-0.039	-0.041	-0.039	-0.041
	(0.186)	(0.187)	(0.186)	(0.187)
Higher*PK	0.186	0.182	0.186	0.182
	(0.166)	(0.168)	(0.166)	(0.168)
Primary* CRM	0.136	0.138	0.136	0.138
	(0.084)	(0.084)	(0.084)	(0.084)
Secondary*CRM	-0.016	-0.013	-0.016	-0.013
	(0.109)	(0.110)	(0.109)	(0.110)
Higher*CRM	-0.145	-0.144	-0.145	-0.144
	(0.090)	(0.091)	(0.090)	(0.091)
Partner drunk at last sex	-	-0.086	-	-0.086
		(0.129)		(0.129)
Distance	-0.008***	-0.008***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	20493	20475	20493	20475

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relation type. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV Prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table 4.8: Second Stage Results for Condom Use

	IV Probit		2SLS	
Variables	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
HIV Rate	-0.010	0.010	-0.007	0.008
	(0.011)	(0.009)	(0.003)	(0.003)
Male	0.224***	0.218***	0.133***	0.161***
	(0.031)	(0.017)	(0.007)	(0.006)
Age	-0.066***	-0.004	-0.015***	0.003
	(0.008)	(0.007)	(0.002)	(0.002)
Age squared	0.000***	-0.000	0.000*	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.360**	0.248**	0.311***	0.192**
	(0.152)	(0.122)	(0.046)	(0.037)
Secondary	1.043***	1.136***	0.972***	0.977***
	(0.200)	(0.171)	(0.057)	(0.054)
Higher	1.571***	1.143***	1.249***	0.988***
	(0.219)	(0.201)	(0.059)	(0.062)
Christianity	0.044	0.016	0.022	0.012
	(0.033)	(0.035)	(0.008)	(0.010)
Husband/Wife	-0.971***	-0.875***	-0.702***	-0.757***
	(0.030)	(0.058)	(0.011)	(0.019)
Cohabiting	-0.341***	-0.400***	-0.186***	-0.315***
	(0.027)	(0.032)	(0.007)	(0.010)
Other relationship type	0.133	0.302***	0.071**	0.161***
	(0.086)	(0.090)	(0.014)	(0.017)
Perceived Unfaithfulness (Yes)	-0.026	0.143***	-0.008	0.106***
	(0.021)	(0.014)	(0.005)	(0.004)
Rural	-0.194***	-0.185***	-0.115***	-0.144***
	(0.025)	(0.029)	(0.006)	(0.009)
One Partner	0.038	-0.049*	0.028	-0.024
	(0.038)	(0.026)	(0.007)	(0.007)
PK	0.212***	0.226***	0.156***	0.179***
	(0.070)	(0.054)	(0.020)	(0.016)
CRM	0.131***	0.099***	0.106***	0.086***
	(0.022)	(0.034)	(0.007)	(0.012)
Primary*PK	0.020	0.037	0.025	0.051
	(0.074)	(0.071)	(0.022)	(0.022)
Secondary*PK	-0.078	-0.247***	-0.091	-0.195***
	(0.096)	(0.061)	(0.025)	(0.018)
Higher*PK	-0.180	-0.164*	-0.144**	-0.135*
	(0.114)	(0.193)	(0.028)	(0.028)
Primary*CRM	-0.052*	-0.025	-0.046*	-0.021
	(0.031)	(0.035)	(0.010)	(0.012)
Secondary*CRM	-0.059**	-0.064	-0.071***	-0.060*
	(0.027)	(0.040)	(0.008)	(0.014)
Higher*CRM	-0.113***	-0.102**	-0.096***	-0.090**
	(0.038)	(0.046)	(0.010)	(0.015)
Partner drunk at last sex	-	-0.286***	-	-0.222***
		(0.026)		(0.007)
Observations	20493	20475	20493	20475
Exogeneity Test	Chi2(1) =4.67	Prob>Chi2=0.031	F(1,25) =3.082	P=0.091
Instrument Strength			F(1,25)=13.335	P>F 0.0012

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relation type. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

4.5.2 Perceived risk and sexual behaviour

Table 4.9 presents ordered probit marginal effects by the number of sexual partners for our perceived risk. Perceived risk variables are dummies taking the value of 1 for either the less concerned or more concerned about the likelihood of infection after the introduction of ART. Results reveal that compared to those who have not changed their concern levels, the less concerned about the likelihood of HIV infection are less likely to abstain. They are also approximately a percentage point more likely to have one or more partners. We observe similar effects for males, age and Christianity as in Table 4.5.

Table 4.10 presents the probit marginal effects of perceived risk on the use of a condom at both first and last sexual intercourse. For those whose level of concern has increased, there is an increased likelihood of condom use, which could imply that higher levels of risk perception may result in precautionary sexual behaviour. However, increased likelihood of condom use is also evident for those reporting they are less concerned, possibly suggesting that although they have increased likelihood of one or more sexual partners (see Table 4.9), they continue to use condoms. Education and Christianity increase the likelihood of condom use, while there is less likelihood of use among the married and cohabiting partners as well as those living in rural areas. Although HIV knowledge variables increase the likelihood of condom use, their interaction with education appears to reduce the likelihood. This could suggest that perhaps education may increase the expected utility of unsafe sex, which may undermine the value of knowledge.

Table 4.9: Marginal Effects (at means) from Ordered Probit (Perceived Risk)

Variables	0 Partner	1 Partner	>1 Partner
Less Concerned	-0.015**	0.007**	0.008**
	(0.007)	(0.003)	(0.004)
More Concerned	-0.004	0.002	0.002
	(0.006)	(0.003)	(0.003)
Male	-0.090***	0.041***	0.049***
	(0.007)	(0.003)	(0.004)
Age	-0.044***	0.020***	0.024***
	(0.002)	(0.001)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.001	0.000	0.000
	(0.050)	(0.023)	(0.028)
Secondary	-0.026	0.012	0.014
	(0.037)	(0.017)	(0.020)
Higher	-0.047	0.021	0.026
	(0.054)	(0.024)	(0.029)
Christianity	0.032***	-0.014***	-0.017***
	(0.007)	(0.003)	(0.004)
Cohabitation	-0.022**	0.010**	0.012**
	(0.010)	(0.004)	(0.005)
Never married	0.116***	-0.053***	-0.064***
	(0.013)	(0.006)	(0.008)
Other Marital status	0.222***	-0.101***	-0.121***
	(0.021)	(0.009)	(0.012)
Rural	0.016**	-0.007**	-0.009**
	(0.008)	(0.004)	(0.004)
PK	-0.023	0.010	0.012
	(0.016)	(0.007)	(0.009)
CRM	0.000	-0.000	-0.000
	(0.006)	(0.003)	(0.004)
Primary*PK	-0.018	0.008	0.010
	(0.021)	(0.009)	(0.011)
Secondary*PK	-0.029	0.013	0.016
	(0.019)	(0.009)	(0.011)
Higher*PK	-0.011	0.005	0.006
	(0.024)	(0.011)	(0.013)
Primary*CRM	0.008	-0.004	-0.004
	(0.009)	(0.004)	(0.005)
Secondary*CRM	0.005	-0.002	-0.003
	(0.008)	(0.004)	(0.004)
Higher*CRM	-0.009	0.004	0.005
	(0.011)	(0.005)	(0.006)
Wave 4 (2013)	-0.058***	0.026***	0.032***
	(0.007)	(0.003)	(0.004)
District FE	Yes	Yes	Yes
Observations	16256	16256	16256

Notes: Robust standard errors are in parenthesis. Omitted categories are none for education, married for marital status, and no change in concern level for change in concern. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table 4.10: Probit Marginal Effects for Perceived Risk on the use of a condom

Variables	Condom use at first sex	Condom use at last sex
Less Concerned	0.020**	0.016
	(0.008)	(0.010)
More Concerned	0.020***	0.037***
	(0.008)	(0.009)
Male	0.048***	0.062***
	(0.007)	(0.008)
Age	-0.011***	0.000
	(0.002)	(0.003)
Age squared	0.000	-0.000
	(0.000)	(0.000)
Primary	0.134***	0.095
	(0.047)	(0.066)
Secondary	0.215***	0.395***
	(0.045)	(0.060)
Higher	0.287***	0.358***
	(0.064)	(0.078)
Christianity	0.018**	0.023***
	(0.008)	(0.010)
Husband/Wife	-0.204***	-0.273***
	(0.009)	(0.011)
Cohabitation	-0.079***	-0.113***
	(0.008)	(0.010)
Other relationship type	0.012	0.130***
	(0.027)	(0.032)
Perceived Unfaithfulness (Yes)	-0.015**	0.038***
	(0.008)	(0.010)
Rural	-0.051***	-0.068***
	(0.009)	(0.011)
One Partner	0.004	-0.021*
	(0.011)	(0.012)
PK	0.039**	0.069***
	(0.018)	(0.025)
CRM	0.023***	0.032***
	(0.009)	(0.012)
Primary*PK	-0.022	-0.006
	(0.023)	(0.032)
Secondary*PK	-0.026	-0.106***
	(0.021)	(0.029)
Higher*PK	-0.043	-0.070**
	(0.028)	(0.035)
Primary*CRM	-0.016	-0.010
	(0.011)	(0.015)
Secondary*CRM	-0.015	-0.034**
	(0.010)	(0.014)
Higher*CRM	-0.020	-0.041**
	(0.013)	(0.017)
Partner drunk at last sex	-	-0.095***
		(0.014)
Wave 4 (2013)	0.067***	-0.010
	(0.007)	(0.009)
District FE	Yes	Yes
Observations	11704	11675
Loglikelihood	-4461.099	-6192.283
Prob>chi2	0.000	0.000

Notes: Robust standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type, no change in concern level for change in concern.PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Tables 4.11 and 4.12 present results for first and second stage estimations (equations 4.15 and 4.16) for the perceived risk, respectively. In this case, updating the level of concern (being less or more concerned) is considered endogenous (Paula et al., 2014) and is instrumented with months of ART enrolment before the survey. Table 4.11 shows that the instrument for the perceived risk carries a negative sign for the less concerned. This implies that the longer the duration before the survey, the less likely respondents indicate they are less concerned about the likelihood of being infected. For those who increased their concern level, the instrument carries a positive sign. A possible explanation could be that if there is a belief that ARV suppresses the virus, then risky sexual behaviour may increase as was the case documented in the US (Lakdawalla et al. 2006). This could make respondents increase their level of concern. In both cases, the statistical significance of the instrument implies that it is relevant to the endogenous regressor. Moreover, diagnostic tests for the endogeneity of perceived risk reject the null hypothesis that perceived risk is exogenous. For the less concerned, the Chi2 =42.521 (p=0.000), while it is 41.126 (p=0.000) for the more concerned. Tests for the strength of the instrument show that the considered instrument is not weak. For instance, with the less concerned F statistic is 164.634 (P>F 0.000), while it is 17.769 (P>F 0.000) for the more concerned. Since both statistics are greater than 10, we conclude that the instrument is not weak (Staiger and Stock, 1997).

On the other hand, second stage results show that those less concerned are less likely to use a condom at first sex, while the more concerned have an increased likelihood of condom use at first sex (Table 4.12). We observe that for the less concerned, the accompanying sign is the opposite of that in Table 4.10 (for binary probit), suggesting that not accounting for the endogeneity may bias the results. No effect of concern level is evidenced for condom use at last sex.

Table 4.11: First Stage results for the change in the level of concern about HIV infection

	Condom at first sex		Condom at last sex	
Variables	Less Concerned	More Concerned	Less Concerned	More Concerned
Male	-0.015*	0.004	-0.015*	0.004
	(0.008)	(0.009)	(0.008)	(0.009)
Age	-0.000	0.004	-0.000	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.021	0.094	0.021	0.098
	(0.069)	(0.064)	(0.069)	(0.064)
Secondary	0.053	0.159***	0.051	0.161***
	(0.061)	(0.056)	(0.061)	(0.056)
Higher	0.019	0.302***	0.021	0.307***
	(0.076)	(0.080)	(0.076)	(0.079)
Christianity	-0.028***	0.026***	-0.028***	0.026***
	(0.010)	(0.010)	(0.010)	(0.010)
Husband/Wife	0.007	-0.024*	0.008	-0.026*
	(0.012)	(0.013)	(0.012)	(0.013)
Cohabitation	0.035***	0.002	0.034***	0.002
	(0.010)	(0.011)	(0.010)	(0.011)
Other relationship type	0.009	-0.006	0.007	-0.006
	(0.026)	(0.027)	(0.026)	(0.027)
Perceived unfaithfulness (Yes)	0.008	0.023**	0.008	0.023**
	(0.010)	(0.010)	(0.010)	(0.010)
Rural	-0.028***	0.018*	-0.028***	0.018*
	(0.009)	(0.010)	(0.009)	(0.010)
One Partner	0.002	0.016	0.002	0.015
	(0.012)	(0.013)	(0.012)	(0.013)
PK	-0.025	0.067***	-0.026	0.063***
	(0.027)	(0.023)	(0.027)	(0.023)
CRM	0.037***	-0.002	0.038***	-0.001
	(0.013)	(0.013)	(0.013)	(0.013)
Primary*PK	0.029	-0.040	0.030	-0.037
	(0.035)	(0.033)	(0.035)	(0.033)
Secondary*PK	0.004	-0.029	0.006	-0.027
	(0.031)	(0.027)	(0.031)	(0.027)
Higher*PK	0.007	-0.048	0.008	-0.046
	(0.035)	(0.035)	(0.035)	(0.035)
Primary*CRM	-0.031**	0.012	-0.031**	0.008
	(0.016)	(0.016)	(0.016)	(0.016)
Secondary*CRM	-0.030**	-0.002	-0.031**	-0.004
	(0.015)	(0.015)	(0.015)	(0.015)
Higher*CRM	-0.044***	-0.029	-0.045***	-0.032*
	(0.017)	(0.019)	(0.017)	(0.019)
Partner drunk at last sex	-	-	0.010	-0.001
			(0.015)	(0.015)
ARVMBS	-0.002***	0.001***	-0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	11704	11704	11675	11675

Notes: Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type.PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ARVMBS: ARV months before the survey. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table 4.12: IV (2SLS) Regression results for condom use

Variables	CAFS	CAFS	CALS	CALS
Less Concerned	-0.427***	-	-0.021	-
	(0.074)		(0.081)	
More Concerned	-	1.113***	-	0.055
		(0.307)		(0.211)
Male	0.042***	0.044***	0.060***	0.060***
	(0.008)	(0.012)	(0.008)	(0.008)
Age	0.001	-0.004	0.003	0.003
	(0.003)	(0.004)	(0.003)	(0.003)
Age squared	-0.000***	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.251***	0.138	0.135*	0.129*
	(0.076)	(0.103)	(0.074)	(0.077)
Secondary	0.425***	0.226**	0.453***	0.443***
	(0.063)	(0.096)	(0.062)	(0.072)
Higher	0.483***	0.138	0.434***	0.417***
	(0.077)	(0.149)	(0.081)	(0.105)
Christianity	0.010	-0.007	0.019*	0.018
	(0.010)	(0.017)	(0.010)	(0.012)
Husband/Wife	-0.248***	-0.224***	-0.313***	-0.312***
	(0.013)	(0.020)	(0.013)	(0.014)
Cohabitation	-0.053***	-0.070***	-0.110***	-0.111***
	(0.009)	(0.014)	(0.010)	(0.010)
Other relationship type	0.020	0.023	0.081***	0.081***
	(0.020)	(0.033)	(0.017)	(0.017)
Perceived unfaithfulness (Yes)	-0.008	-0.038**	0.036***	0.034***
	(0.008)	(0.015)	(0.009)	(0.010)
Rural	-0.054***	-0.062***	-0.070***	-0.070***
	(0.008)	(0.013)	(0.009)	(0.009)
One Partner	0.012	-0.006	-0.013	-0.014
	(0.010)	(0.017)	(0.011)	(0.011)
PK	0.060**	-0.004	0.084***	0.081***
	(0.029)	(0.042)	(0.028)	(0.031)
CRM	0.055***	0.041**	0.037**	0.036**
	(0.015)	(0.020)	(0.015)	(0.015)
Primary*PK	-0.026	0.006	-0.016	-0.014
	(0.038)	(0.051)	(0.037)	(0.038)
Secondary*PK	-0.058*	-0.027	-0.118***	-0.117***
	(0.031	(0.042)	(0.030)	(0.031)
Higher*PK	-0.077**	-0.027	-0.091**	-0.089**
	(0.036)	(0.053)	(0.037)	(0.038)
Primary*CRM	-0.041**	-0.041*	-0.012	-0.011
	(0.018)	(0.024)	(0.018)	(0.017)
Secondary*CRM	-0.045***	-0.030	-0.039**	-0.038**
	(0.016)	(0.022)	(0.016)	(0.016)
Higher*CRM	-0.056***	-0.004	-0.048***	-0.046**
	(0.019)	(0.028)	(0.019)	(0.020)
Partner drunk at last sex	-	-	-0.094***	-0.094***
			(0.015)	(0.015)
Observations	11704	11704	11675	11675
Exogeneity test for less concern	Chi2(1)=42.521	P=0.000		
Exogeneity for more concerned	Chi2(1)=41.126	P=0.000		
Instrument strength	F=164.634	P>F 0.000		

Notes: Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. Instrumented: Less concerned and more concerned. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

4.6 Robustness checks

Controlling for the endogenous HIV risk measures, we arrive at the same conclusion through both IV probit and two-stage least squares. The instruments are relevant and not weak according to the conventional tests. Standard tests also confirm that both the objective risk measure (HIV prevalence rate) and perceived risk (concern about the likelihood of infection due to ART) are endogenous. As a robustness check, we consider potential bias due to those included who report they are not sexually active and explore the effect across some subsamples. Results for robustness checks are presented in the appendix.

4.6.1 Is there evidence of sample selection?

We have alluded to the potential effect of non-participation in the sexual market within the framework of Philipson and Posner (1995). In this sub-section, we check for this effect, through the sample selection model (Heckman, 1979). The hypothesis which we test is that there is no relationship between the unobservables that affect the decision to report being sexually active and the unobservables that affect the decision to use condoms (Cameron and Trivedi, 2010). If such a relationship which is represented by *rho* is statistically significant, then failing to account for it may bias the results.

Tables A4.1 and A4.2 in the appendix present the results for objective and perceived risk, respectively. The effect of other control variables is largely the same as in the main specifications. Key to note is that the Wald test rejects the null hypothesis of independence between the unobservables (*rho*=0). In both tables, a positive coefficient of *rho* (0.455, 0.390, 0.339 and 0.310) suggests a positive relationship between unobservables affecting the reported activity and those determining the use of condoms. We, therefore, conclude that the approach that only focuses on those with reported activity could introduce bias. Notably, these results are consistent with the ordered probit model results emphasising the need to include those who reported being not sexually active, as we did. Given that the HIV/AIDS campaigns in the country used to (and still) emphasize the importance of abstaining from sexual activities through the slogan *ABC* (abstain, be faithful and condomise), consideration of the sexually inactive is important.

4.6.2 Subsample Effect

We also investigate the effect for the subsamples of rural vs urban areas, and males vs females. In these subsamples, we repeat the estimation process: ordered probit for sexual relations,

binary probit, and account for endogeneity for both objective and perceived risks through an IV probit model and the 2SLS estimator. When controlling for endogeneity as in the main estimations, we test for both potential endogeneity and the strength of the instruments. The test results largely confirm what we find in the main estimation; HIV prevalence rate and the perceived risk relating to the likelihood of being infected due to ART are indeed endogenous, and that the instruments are strong and relevant based on Staiger and Stock (1997).

Rural vs urban areas

For the number of sexual relations (ordered probit results), the effect of gender, age, education and Christianity are the same in both rural and urban areas (Tables A4.3-A4.6), as observed at the national level. HIV rate remains insignificant. Whereas the secondary education and correct rejection of HIV misconceptions interaction reduces the likelihood of 1 or more sexual relations in rural areas, in the urban sample such an effect is observed among those with primary education level. Those who report having reduced their concern level due to ART are more likely to report being active with one or more partners in urban areas (Table A4.6). Probit results show that HIV rate increases the likelihood of condom use at first sex and last sex by less than one percentage point in urban areas (Table A4.7) Similarly, compared to those who have not changed their level of concern, the more concerned are more likely to report the use of a condom at both first and last sex in urban areas, and only at last sex in rural areas. On the other hand, those less concerned are more likely to report condom use in rural areas only (Table A4.8).

IV probit results presented in Table A4.9 show that the HIV rate increases the likelihood of condom use at last sex by about 2 percentage points in urban areas only. Perceived partner unfaithfulness increases the likelihood of condom use at last sex in both urban and rural areas. The second stage of the 2SLS results presented in Tables A4.10 and A4.11 show that those less concerned about being infected are less likely to report condom use at first and last sex in rural areas, and only at first sex in urban areas, while the more concerned are more likely to report condom use at first sex in both rural and urban sub-samples.

Males vs Females

Education increases the likelihood of non-abstinence more among males than females (see Tables A4.12 and A4.13). Among females, it is only the primary education level that increases the likelihood of sexual relations, while among males primary, secondary and higher education

increase the likelihood compared to those with no education. Similarly, the less concerned males are less likely to abstain than females (Tables A4.14 and A4.15). Probit marginal effects show that the HIV rate increases the likelihood of condom use at both first and last sex among males, and only at first sex among females (Table A4.16). On the other hand, Table A4.17 shows that the more concerned males and females are more likely to report use of condom at first and last sex, while being less concerned has no effect. Perceived partner unfaithfulness increases condom use at last sex among both males and females. Accounting for endogeneity results in a non-significant effect of HIV prevalence on condom use (Table A4.18) as also observed with the total sample. Lower use of condoms among the married and cohabiting partners is still evident, while perceived partner unfaithfulness increases the likelihood of condom use at last sex. Both the less concerned males and females are less likely to report condom use at first sex, and the more concerned are more likely to use condoms at first sex. No effect of changed level of concern is evident on condom use at last sex (see Tables A4.19 and A4.20).

One vs More than one sexual partner

For these, we construct an indicator variable to proxy consistent use of a condom, although we do not observe its frequency of use. We create a dummy variable =1 if the respondent reports condom use at both first and last sex, otherwise 0. We only report the second stage results for both IV Probit and 2SLS in Table A4.21. As in the previous analysis, the HIV rate does not significantly affect condom use among those with one partner and those with more than one partner.

HIV prevalence and perceived risk

Having established that HIV prevalence has no effect on sexual behaviour, one question could be whether prevalence has any effect on perception. Based on the empirical review, we check for this potential effect considering only those who indicated to have increased their level of concern. Table A4.22 presents the probit marginal effects of factors affecting an increase in the level of concern on the likelihood of being infected with HIV/AIDS. HIV prevalence rate is positively associated with an increased likelihood of being more concerned. Thus, while information on the perceived risk was explicitly sourced in relation to the existence of ART, this result could suggest that the effect of HIV prevalence on sexual behaviour might be indirectly through perception. Table A4.22 further shows that the likelihood of being more concerned also increases with increasing levels of education and Christianity.

4.7 Discussion and Relation to Literature

Results show that while HIV prevalence rate does not affect the reported sexual relations (although carries an ideal expected sign), probit results show that it increases the likelihood of condom use at last sexual intercourse but by less than 1 per cent. However, when endogeneity is accounted for, the HIV prevalence rate has no significant effect on condom use. One study (Keetile and Kgosidintsi, 2018) in Botswana finds no significant association between being on HAART (an indicator of being HIV positive) and risky behaviour. Compared to those who indicated not to have changed their level of concern about being infected with HIV due to ART, the less concerned are more likely to report more than one sexual relation, a pattern similar to studies in Scotland (Hart and Williamson, 2005) and Malawi (Paula et al., 2014), suggesting possibilities of negative externalities of policy interventions. Although probit results show that both the more and less concerned have a higher likelihood of condom use, accounting for endogeneity reveals that the less concerned are less likely to use condoms and more concerned are more likely to use a condom, but only at first sex.

Results further show that males are more likely to report more than one partner than females. They are also more likely to report using condoms. The likelihood for more than one partner increases with increasing level of education. A mixed pattern on the effect of education has been observed previously. For instance, Fortson (2008) finds that education is positively associated with HIV status in five Africa countries (Ghana, Burkina Faso, Kenya, Cameroon and Tanzania). Using the same data, although De Walque (2009) finds that education has no effect on HIV infection (with only reduced effect in urban areas), his results also show that more years of education increase the likelihood for extramarital sex and reduce the likelihood for abstinence. A negative association between education and HIV status has been found in Botswana (De Neve et al., 2015) and Zimbabwe (Aguero and Bharadwaj, 2014), where it also reduces the number of sexual partners. Other studies in Sub-Saharan African countries (15 countries considered by Bingenheimer, 2010 and 32 countries by Lucas and Wilson, 2018) find that education for males is associated with multiple partnership. Therefore, our results seem to portray a similar pattern observed in some Sub-Saharan African countries.

There is lower use of condom among married and cohabiting partners, while other relation type (casual acquaintance, paying someone for sex) increases the likelihood of condom use at last sex, consistent with the literature surveyed by Guillon and Thuilliez (2015). Although a

commonly provided explanation for the less use of condoms among married and cohabiting partners is confidence/trust in partners (Guillon and Thuilliez, 2015), possibilities of relationship fatalism as theoretically predicted by Sterck (2013) may not be ruled out. Given that the 2008 and 2013 survey statistical reports show higher HIV prevalence among cohabiting and married than among the never married persons, this could possibly imply relationship fatalism, rather than trust. About 28 per cent of cohabiting partners said they do not trust their partners, while the share is 9 per cent for the married. Fatalism is highly likely in the case of a serodiscordant relationship (i.e where one partner is HIV positive, while the other is not), especially when the HIV transmission rate is overestimated (Sterck, 2013). Sterck provides evidence in which the held beliefs that this discordant relationship is less likely, resulted in unsafe sex. Thus, if one is infected, there is a belief that another partner is infected as well, hence the less likelihood of condom use. Although with a small sample size (46), Baratedi (2013) provides similar insights that couples in some purposefully selected areas in Botswana did not believe that it is possible to be in a discordant relation. In such an instance, the likelihood of condom use is reduced.

Results further suggest that the less use of condoms among married and cohabiting is evident even among those who report more than one sexual partner, which could signal a risky pattern of behaviour. When we consider only those who reported being married during the survey (that is, currently married) and relate this to the type of relationship with the most reported partner, we find that about 3 per cent of the most reported partner for the married was not a husband/wife, 2 per cent were boyfriend/girlfriend, while about 0.3 per cent were other relation. Those in marriage/cohabiting relation were asked about the relationship`s duration. Relating the average duration of about 11 years for those who provided information to the above pattern, this could further signal a trend towards risky behaviour of extra marital sexual intercourse.

4.8 Conclusion

This chapter investigates whether sexual behaviour (indicators of sexual relations and the use of a condom) is responsive to HIV prevalence and perceived risk of HIV infection due to the existence of a treatment programme. The existing literature is inconclusive on this matter, and previous studies for Botswana do not account for endogeneity between sexual behaviour and HIV risk, a gap we address. Furthermore, these studies only consider those who report being sexually active, which introduces issues of potential selection bias. Therefore, we account for

the endogeneity of risk (Oster, 2012; Paula et al., 2014). HIV prevalence is instrumented with distance to HIV hot spot, while the perceived risk is instrumented with the number of months between ART rollout and the surveys. To account for those who report not being sexually active during the 12 months prior to the survey, we estimate an ordered probit model (Glick and Sahn, 2008).

Results show that males and those with higher education levels are more likely to report being sexually active, and more likely to have more than one partner. They are also likely to use condoms. However, the interaction between education and knowledge of HIV prevention reduces the likelihood of condom use in some instances. This pattern is observed for secondary and higher education, possibly suggesting that these variables may unintentionally reduce protective behaviour. While information promotion is useful, some (Oster, 2012) argue that high rates of knowledge among respondents as appears to be the case in Botswana, may, in the long run, render further efforts ineffective. We also find there is less likelihood to use condoms among married and cohabiting partners. While this is commonly attributed to trust, there are some insights from our results that some married partners report sexual relations with non-wives/husbands and that some married/cohabitating couples did not trust their partners.

Overall, results show that people are more responsive to perceived than objective risk indicators. The HIV prevalence rate has no effect on either the number of partners or use of a condom, and this is robust to subsample (rural vs urban, male vs female) investigations. Compared to those who indicated not to have changed their concern level of being infected with HIV due to ART, the less concerned are more likely to report more than one sexual partner, implying possible negative externalities of the policy. From a policy perspective, the non-behavioural response to the HIV prevalence rate could be a cause for concern if there are unobserved fatal behavioural patterns. Thus, because programmes such as ART prolongs the lives of the infected and affects perception on the likelihood of infection risk, there could be some negative externalities as revealed by the results.

Future work should account for one`s known HIV status to be able to assess subsequent behaviour, findings for which will reveal the extent of either fatalism or altruism by those who are infected. This will further guide intervention tools in relation to the Public Health Act, 2013. The Act recognises that the assumption that people are altruistic, which some state as the

basis for public health interventions (Hauck, 2018), may not hold. However, measures such as detaining HIV positive individuals who knowingly spread the virus could be seen to be in contradiction to the revised 2012 national HIV policy objective of reducing HIV and AIDS-related stigma and discrimination. Future research ascertaining the effects of known individual HIV status should assist in guiding policy in this matter. So far, one study (Keetile and Kgosidintsi, 2018) has considered the effect of enrolment in ART on sexual behaviour. However, the study did not address the endogeneity of being in HAART and the effect of testing decisions. The inability to observe the timing of enrolment makes such study inadequate.

Other STD`s biomarker may be considered as indicators for risky behaviour. Currently, surveys are only limited to HIV/AIDS, yet there are indications that some respondents reported risky behaviour while on STD treatment. Such consideration, we argue, should help to enhance the effectiveness of policy intervention. Lastly, results suggest that the approach that only focuses on those with reported activity could introduce bias: there is a positive relationship between unobservables affecting the reported activity and those determining the use of condoms, highlighting the need to include the sexually inactive in understanding sexual behavioural responses to HIV risks.

APPENDIX FOR CHAPTER 4

Table A4.1: Heckman Sample Selection results for HIV Rate impact on condom use

Dependent Variables:	Condom use at first sex		Condom use at last sex	
Controls	Coefficient	CSE	Coefficient	CSE
HIV Rate	0.001	0.001	0.002	0.002
Male	0.060***	0.006	0.072***	0.006
Age	0.008**	0.003	0.016***	0.003
Age squared	-0.000***	0.000	-0.000***	0.000
Primary	0.157***	0.051	0.102***	0.039
Secondary	0.392***	0.061	0.411***	0.051
Higher	0.477***	0.069	0.405***	0.063
Husband/Wife	-0.208***	0.010	-0.239***	0.016
Cohabiting	-0.022**	0.010	-0.074***	0.009
Other relation status	0.017	0.014	0.063***	0.018
Perceived unfaithfulness (Yes)	-0.013**	0.005	0.038***	0.004
Rural	-0.049***	0.006	-0.064***	0.010
One Partner	0.012*	0.007	-0.012	0.008
PK	0.072***	0.023	0.081***	0.017
CRM	0.038***	0.008	0.034***	0.012
Primary*PK	-0.003	0.025	0.014	0.023
Secondary*PK	-0.042	0.028	-0.080***	0.019
Higher*PK	-0.057*	0.030	-0.055*	0.029
Primary*CRM	-0.024**	0.011	-0.012	0.012
Secondary*CRM	-0.031***	0.008	-0.028**	0.013
Higher*CRM	-0.034***	0.011	-0.035**	0.015
Partner drunk at last sex	-	-	-0.094***	0.007
Wave 3 (2008)	0.097***	0.007	0.061***	0.010
Wave 4 (2013)	0.203***	0.010	0.081***	0.010
Selection Dependent variable:		Report of being sexually active		
HIV Rate	-0.003	0.006	-0.001	0.007
Male	0.108***	0.019	0.094***	0.019
Age	0.229***	0.007	0.226***	0.006
Age squared	-0.003***	0.000	-0.003***	0.000
Primary	0.176*	0.102	0.179*	0.100
Secondary	0.351***	0.092	0.352***	0.087
Higher	0.109	0.168	0.082	0.162
Christianity	-0.143***	0.019	-0.153***	0.020
Cohabitation	-0.128***	0.036	-0.107***	0.036
Never married	-1.295***	0.045	-1.268***	0.043
Other Marital status	-1.615***	0.082	-1.588***	0.081
Rural	-0.031	0.034	-0.031	0.034
PK	0.125***	0.046	0.135***	0.044
CRM	0.054*	0.029	0.051*	0.028
Primary*PK	0.014	0.051	0.007	0.047
Secondary*PK	0.025	0.055	0.020	0.054
Higher*PK	0.071	0.072	0.062	0.071
Primary*CRM	-0.047	0.035	-0.044	0.036
Secondary*CRM	-0.046	0.034	-0.041	0.033
Higher*CRM	-0.047	0.044	0.061	0.042
Wave 3 (2008)	-0.076***	0.024	-0.088***	0.025
Wave 4 (2013)	0.343***	0.037	0.323***	0.037
/athrho	0.491***	0.043	0.391***	0.044
/lnsigma	-0.974***	0.009	-0.830***	0.007
rho	0.455	0.034	0.372	0.038
Observations	28855		28837	
Selected	20489		20471	
Non selected	8366		8366	
Wald test chi2(Prob>chi2)	131.44(0.000)		70.52(0.000)	

Notes: CSE=Clustered Standard Error. Omitted categories are none for education, boyfriend/girlfriend for relationship type. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.2: Heckman Sample Selection results for Perceived Risk impact on condom use

Dependent Variables:	Condom use at first sex		Condom use at last sex	
Controls	Coefficient	SE	Coefficient	SE
Less concerned	0.022**	0.009	0.019*	0.011
More concerned	0.019**	0.008	0.037***	0.009
Male	0.052***	0.007	0.064***	0.007
Age	0.013**	0.004	0.016***	0.004
Age squared	-0.000***	0.000	-0.000***	0.000
Primary	0.247***	0.062	0.132**	0.059
Secondary	0.400***	0.064	0.459***	0.066
Higher	0.466***	0.082	0.438***	0.078
Husband/Wife	-0.209***	0.013	-0.268***	0.017
Cohabiting	-0.034***	0.008	-0.075***	0.009
Other relation status	0.014	0.014	0.086***	0.018
Perceived unfaithfulness (Yes)	-0.015*	0.008	0.033***	0.010
Rural	-0.054***	0.007	-0.073***	0.011
One Partner	0.005	0.006	-0.018*	0.009
PK	0.075**	0.031	0.090***	0.030
CRM	0.039***	0.013	0.036**	0.014
Primary*PK	-0.037	0.033	-0.010	0.036
Secondary*PK	-0.057*	0.032	-0.114***	0.035
Higher*PK	-0.080**	0.036	-0.093**	0.041
Primary*CRM	-0.029*	0.017	-0.012	0.016
Secondary*CRM	-0.032***	0.012	-0.039**	0.015
Higher*CRM	-0.032**	0.015	-0.042**	0.017
Partner drunk	-	-	-0.096***	0.009
Wave 4 (2013)	0.093***	0.007	0.009	0.008
Selection Dependent variable:		Report of being sexually active		
Less concerned	0.074**	0.034	0.071**	0.034
More concerned	0.034	0.033	0.032	0.035
Male	0.124***	0.026	0.115***	0.026
Age	0.233***	0.009	0.230***	0.009
Age squared	-0.003***	0.000	-0.003***	0.000
Primary	0.095	0.204	0.091	0.201
Secondary	0.181	0.162	0.179	0.158
Higher	-0.035	0.232	-0.067	0.229
Christianity	-0.141***	0.022	-0.149***	0.023
Cohabitation	-0.138**	0.058	-0.097*	0.054
Never married	-1.206***	0.059	-1.165***	0.053
Other Marital status	-1.507***	0.088	-1.446***	0.084
Rural	-0.057	0.042	-0.058	0.042
PK	0.097	0.062	0.103*	0.060
CRM	0.012	0.036	0.009	0.036
Primary*PK	0.054	0.086	0.046	0.084
Secondary*PK	0.080	0.084	0.079	0.082
Higher*PK	0.056	0.111	0.047	0.108
Primary*CRM	-0.034	0.047	-0.027	0.048
Secondary*CRM	-0.024	0.043	-0.020	0.042
Higher*CRM	0.102**	0.046	0.115**	0.046
Wave 4 (2013)	0.430***	0.031	0.430***	0.031
/athrho	0.355***	0.039	0.326***	0.044
/lnsigma	-1.033***	0.011	-0.847***	0.011
rho	0.341	0.034	0.315	0.039
Observations	15650		15621	
Selected	11700		11671	
Non selected	3950		3950	
Wald test chi2(Prob>chi2)	83.76(0.000)		55.99(0.000)	

Notes: CSE=Clustered Standard Error. Omitted categories are none for education, boyfriend/girlfriend for relationship type.
***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.3: Ordered Probit Marginal Effects (at means) for sexual partners in rural sample

Variables	0 Partner	1 partner	>1 partner
HIV Rate	0.001	-0.001	-0.000
	(0.002)	(0.002)	(0.001)
Male	-0.083***	0.056***	0.027***
	(0.007)	(0.005)	(0.003)
Age	-0.046***	0.031***	0.015***
	(0.002)	(0.002)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.049*	0.033*	0.016*
	(0.026)	(0.018)	(0.009)
Secondary	-0.141***	0.094***	0.046***
	(0.027)	(0.019)	(0.009)
Higher	-0.185**	0.124**	0.061*
	(0.094)	(0.062)	(0.032)
Christianity	0.036***	-0.024***	-0.012***
	(0.007)	(0.005)	(0.002)
Cohabitation	0.006	-0.004	-0.002
	(0.007)	(0.005)	(0.002)
Never married	0.236***	-0.159***	-0.078***
	(0.011)	(0.009)	(0.004)
Other Marital status	0.363***	-0.224***	-0.120***
	(0.022)	(0.017)	(0.007)
PK	-0.039**	0.026**	0.013**
	(0.016)	(0.011)	(0.015)
CRM	-0.015**	0.010**	0.005**
	(0.007)	(0.005)	(0.002)
Primary*PK	0.011	-0.007	-0.004
	(0.016)	(0.011)	(0.005)
Secondary*PK	-0.002	0.001	0.001
	(0.017)	(0.011)	(0.005)
Higher*PK	-0.029	0.020	0.010
	(0.041)	(0.027)	(0.013)
Primary*CRM	0.008	-0.005	-0.003
	(0.008)	(0.005)	(0.003)
Secondary*CRM	0.024**	-0.016**	-0.008**
	(0.011)	(0.008)	(0.004)
Higher*CRM	0.030	-0.020	-0.010
	(0.024)	(0.016)	(0.008)
Wave 3 (2008)	-0.005	0.004	0.002
	(0.009)	(0.006)	(0.003)
Wave 4 (2013)	-0.059***	0.040***	0.020***
	(0.011)	(0.008)	(0.004)
District FE	Yes	Yes	Yes
Observations	11009	11009	11009

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, married for marital status.PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.4: Ordered probit Marginal Effects (at means) for sexual partners in urban sample

Variables	0 Partner	1 partner	>1 partner
HIV Rate	-0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)
Male	-0.077***	0.040***	0.037***
	(0.005)	(0.003)	(0.003)
Age	-0.057***	0.030***	0.028***
	(0.002)	(0.001)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.099**	0.051**	0.048**
	(0.047)	(0.024)	(0.022)
Secondary	-0.067*	0.035*	0.032*
	(0.039)	(0.020)	(0.019)
Higher	-0.079	0.041	0.038
	(0.049)	(0.025)	(0.023)
Christianity	0.047***	-0.024***	-0.022***
	(0.006)	(0.003)	(0.003)
Cohabitation	-0.030***	0.016***	0.014***
	(0.009)	(0.005)	(0.004)
Never married	0.118***	-0.061***	-0.057***
	(0.012)	(0.006)	(0.007)
Other Marital status	0.253***	-0.131***	-0.122***
	(0.024)	(0.013)	(0.012)
PK	-0.037*	0.019*	0.018*
	(0.020)	(0.010)	(0.010)
CRM	-0.014	0.007	0.007
	(0.009)	(0.005)	(0.004)
Primary*PK	0.008	-0.004	-0.004
	(0.020)	(0.010)	(0.010)
Secondary*PK	-0.014	0.007	0.007
	(0.024)	(0.013)	(0.012)
Higher*PK	-0.002	0.001	0.001
	(0.025)	(0.013)	(0.012)
Primary*CRM	0.024**	-0.012**	-0.011**
	(0.010)	(0.005)	(0.005)
Secondary*CRM	0.011	-0.006	-0.005
	(0.011)	(0.006)	(0.005)
Higher*CRM	-0.000	0.000	0.000
	(0.011)	(0.006)	(0.005)
Wave 3 (2008)	-0.024***	0.013***	0.012***
	(0.009)	(0.005)	(0.004)
Wave 4 (2013)	-0.091***	0.047***	0.044***
	(0.010)	(0.005)	(0.005)
District FE	Yes	Yes	Yes
Observations	18592	18592	18592

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, married for marital status. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.5: Ordered Probit Marginal Effects (at means) for sexual partners Rural sample (perceived Risk)

Variables	0 Partner	1 partner	>1 partner
Less concerned	0.000	-0.000	-0.000
	(0.012)	(0.007)	(0.005)
More concerned	-0.018	0.011	0.007
	(0.012)	(0.007)	(0.005)
Male	-0.090***	0.053***	0.037***
	(0.010)	(0.006)	(0.005)
Age	-0.040***	0.024***	0.017***
	(0.003)	(0.002)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	0.030	-0.018	-0.013
	(0.056)	(0.033)	(0.023)
Secondary	-0.048	0.028	0.020
	(0.056)	(0.033)	(0.023)
Higher	-0.161	0.094	0.066
	(0.123)	(0.072)	(0.051)
Christianity	0.027**	-0.016**	-0.011**
	(0.012)	(0.007)	(0.005)
Cohabitation	0.010	-0.006	-0.004
	(0.013)	(0.008)	(0.005)
Never married	0.197***	-0.115***	-0.081***
	(0.015)	(0.010)	(0.006)
Other Marital status	0.304***	-0.179***	-0.126***
	(0.029)	(0.019)	(0.012)
PK	-0.025	0.014	0.010
	(0.019)	(0.011)	(0.008)
CRM	0.007	-0.004	-0.003
	(0.010)	(0.006)	(0.004)
Primary*PK	-0.025	0.015	0.011
	(0.028)	(0.016)	(0.012)
Secondary*PK	-0.029	0.017	0.012
	(0.026)	(0.015)	(0.011)
Higher*PK	-0.012	0.007	0.005
	(0.047)	(0.027)	(0.019)
Primary*CRM	0.001	-0.000	-0.000
	(0.013)	(0.008)	(0.006)
Secondary*CRM	0.007	-0.004	-0.003
	(0.014)	(0.008)	(0.006)
Higher*CRM	0.009	-0.005	-0.004
	(0.027)	(0.016)	(0.011)
Wave 4 (2013)	-0.041***	0.024***	0.017***
	(0.011)	(0.007)	(0.005)
District FE	Yes	Yes	Yes
Observations	5869	5869	5869

Notes: Standard errors are in parenthesis. Omitted categories are none for education, married for marital status, no change in concern for concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.6: Ordered Probit Marginal Effects (at means) for indicators of sexual partners urban sample (Perceived Risk)

Variables	0 Partner	1 partner	>1 partner
Less concerned	-0.021***	0.008***	0.013***
	(0.008)	(0.003)	(0.005)
More concerned	0.003	-0.001	-0.002
	(0.008)	(0.003)	(0.005)
Male	-0.089***	0.033***	0.056***
	(0.007)	(0.003)	(0.005)
Age	-0.048***	0.018***	0.031***
	(0.002)	(0.001)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.065	0.024	0.041
	(0.066)	(0.024)	(0.042)
Secondary	-0.055	0.020	0.035
	(0.061)	(0.022)	(0.038)
Higher	-0.055	0.020	0.035
	(0.071)	(0.026)	(0.045)
Christianity	0.035**	-0.013**	-0.022**
	(0.009)	(0.003)	(0.006)
Cohabitation	-0.030***	0.011***	0.019***
	(0.008)	(0.003)	(0.005)
Never married	0.080***	-0.029***	-0.051***
	(0.009)	(0.004)	(0.005)
Other Marital status	0.179***	-0.066***	-0.113***
	(0.020)	(0.009)	(0.012)
PK	-0.024	0.009	0.015
	(0.028)	(0.010)	(0.018)
CRM	-0.016	0.006	0.010
	(0.012)	(0.004)	(0.008)
Primary*PK	-0.011	0.004	0.007
	(0.034)	(0.012)	(0.021)
Secondary*PK	-0.027	0.010	0.017
	(0.030)	(0.011)	(0.019)
Higher*PK	-0.010	0.004	0.006
	(0.034)	(0.012)	(0.021)
Primary*CRM	0.025	-0.009*	-0.016*
	(0.014)	(0.005)	(0.009)
Secondary*CRM	0.017	-0.006	-0.011
	(0.013)	(0.005)	(0.008)
Higher*CRM	0.001	-0.000	-0.001
	(0.015)	(0.006)	(0.010)
Wave 4 (2013)	-0.068***	0.025***	0.043***
	(0.008)	(0.003)	(0.005)
District FE	Yes	Yes	Yes
Observations	10387	10387	10387

Notes: Standard errors are in parenthesis. Omitted categories are none for education, married for marital status, no change in concern level for concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level

Table A4.7: Probit Marginal Effects for condom use: (Objective risk)

Variables	Urban		Rural	
	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
HIV Rate	0.006**	0.004*	0.002	0.001
	(0.003)	(0.002)	(0.002)	(0.003)
Male	0.050***	0.063***	0.055***	0.076***
	(0.009)	(0.006)	(0.009)	(0.012)
Age	-0.015***	-0.001	-0.020***	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Age squared	0.000*	-0.000	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.008	0.016	0.112**	0.101**
	(0.062)	(0.059)	(0.046)	(0.040)
Secondary	0.088	0.237***	0.275***	0.363***
	(0.061)	(0.052)	(0.068)	(0.066)
Higher	0.167**	0.195***	0.511***	0.459***
	(0.072)	(0.067)	(0.078)	(0.128)
Christianity	0.015	0.007	0.002	0.011
	(0.009)	(0.011)	(0.013)	(0.019)
Husband/Wife	-0.213***	-0.279***	-0.242***	-0.225***
	(0.006)	(0.015)	(0.016)	(0.022)
Cohabiting	-0.071***	-0.114***	-0.109***	-0.142***
	(0.006)	(0.013)	(0.014)	(0.011)
Other relationship type	0.029	0.091***	0.004	0.087*
	(0.026)	(0.028)	(0.032)	(0.049)
Perceived Unfaithfulness (Yes)	-0.011**	0.043***	-0.015	0.038***
	(0.005)	(0.006)	(0.010)	(0.010)
One Partner	0.021*	-0.009	0.003	-0.018
	(0.011)	(0.011)	(0.011)	(0.014)
PK	-0.016	0.009	0.074***	0.088***
	(0.025)	(0.022)	(0.018)	(0.018)
CRM	0.014*	0.028*	0.027***	0.025*
	(0.008)	(0.015)	(0.008)	(0.013)
Primary*PK	0.042*	0.067**	-0.014	-0.011
	(0.024)	(0.030)	(0.027)	(0.023)
Secondary*PK	0.034	-0.021	-0.033	-0.097***
	(0.027)	(0.025)	(0.035)	(0.024)
Higher*PK	0.025	-0.002	-0.095**	-0.008
	(0.035)	(0.037)	(0.047)	(0.042)
Primary*CRM	-0.011	-0.019	-0.010	0.004
	(0.011)	(0.016)	(0.010)	(0.013)
Secondary*CRM	-0.007	-0.027*	-0.015	-0.008
	(0.008)	(0.017)	(0.012)	(0.016)
Higher*CRM	-0.016	-0.025	-0.027*	-0.062***
	(0.011)	(0.019)	(0.015)	(0.022)
Partner drunk at last sex	-	-0.109***	-	-0.066***
		(0.010)		(0.018)
Wave 3 (2008)	0.092***	0.067***	0.101***	0.067***
	(0.007)	(0.011)	(0.011)	(0.016)
Wave 4 (2013)	0.174***	0.059***	0.172***	0.083***
	(0.014)	(0.011)	(0.015)	(0.017)
District FE	Yes	Yes	Yes	Yes
Observations	13243	13243	7234	7225
Loglikelihood	-5086.791	-6861.441	-3225.832	-4146.436

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.8: Probit Marginal Effects for condom use (Perceived Risk)

	Urban		Rural	
Variables	CUFS	CULS	CUFS	CULS
Less Concerned	0.015	0.002	0.026*	0.043**
	(0.010)	(0.012)	(0.015)	(0.018)
More Concerned	0.019**	0.033***	0.020	0.044**
	(0.009)	(0.011)	(0.014)	(0.017)
Male	0.047***	0.060***	0.049***	0.065***
	(0.008)	(0.010)	(0.013)	(0.015)
Age	-0.008***	0.001	-0.015***	-0.000
	(0.003)	(0.003)	(0.004)	(0.004)
Age squared	0.000	-0.000	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.063	0.181	0.142**	-0.006
	(0.080)	(0.114)	(0.068)	(0.089)
Secondary	0.119	0.394***	0.224***	0.361***
	(0.074)	(0.105)	(0.070)	(0.085)
Higher	0.149*	0.321***	0.329***	0.425***
	(0.087)	(0.116)	(0.124)	(0.148)
Christianity	0.017*	0.013	0.017	0.037**
	(0.010)	(0.012)	(0.014)	(0.017)
Husband/Wife	-0.190***	-0.278***	-0.227***	-0.243***
	(0.010)	(0.013)	(0.017)	(0.022)
Cohabitation	-0.067***	-0.096***	-0.101***	-0.139***
	(0.010)	(0.012)	(0.014)	(0.017)
Other relationship type	-0.003	0.137***	0.037	0.133***
	(0.033)	(0.041)	(0.045)	(0.052)
Perceived Unfaithfulness (Yes)	-0.012	0.041***	-0.023	0.029*
	(0.009)	(0.012)	(0.014)	(0.017)
One Partner	0.008	-0.016	-0.001	-0.029
	(0.012)	(0.014)	(0.021)	(0.024)
Partner drunk at last sex	-	-0.100***	-	-0.087***
		(0.017)		(0.024)
PK	-0.011	0.068	0.066***	0.071**
	(0.033)	(0.047)	(0.025)	(0.032)
CRM	0.011	0.035*	0.031**	0.027*
	(0.014)	(0.020)	(0.012)	(0.016)
Primary*PK	0.009	-0.035	-0.027	0.024
	(0.039)	(0.055)	(0.034)	(0.044)
Secondary*PK	0.007	-0.109**	-0.019	-0.100**
	(0.035)	(0.050)	(0.033)	(0.041)
Higher*PK	0.000	-0.081	-0.033	-0.014
	(0.040)	(0.054)	(0.053)	(0.062)
Primary*CRM	-0.012	-0.033	-0.013	0.015
	(0.016)	(0.023)	(0.016)	(0.021)
Secondary*CRM	-0.008	-0.044**	-0.016	-0.019
	(0.015)	(0.022)	(0.016)	(0.020)
Higher*CRM	-0.009	-0.038	-0.016	-0.066**
	(0.017)	(0.024)	(0.030)	(0.034)
Wave 4 (2013)	0.067***	0.019*	0.070***	-0.009
	(0.009)	(0.011)	(0.013)	(0.016)
District FE	Yes	Yes	Yes	Yes
Observations	7703	7694	3985	3973
Loglikelihood	-2700.014	-3862.921	-1740.923	-2297.908
Prob>chi2	0.000	0.000	0.000	0.000

Notes: Robust standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type, no change for change in concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. CUFS: Condom use at first sex; CULS: Condom use at last sex ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Accounting for endogeneity in Rural and Urban Samples

Table A4.9: IV Probit Results for Condom use: Objective Risk

	Urban		Rural	
Variables	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
HIV Rate	-0.005	0.017**	-0.029	-0.008
	(0.008)	(0.007)	(0.019)	(0.014)
Male	0.221***	0.208***	0.229***	0.241***
	(0.044)	(0.018)	(0.032)	(0.034)
Age	-0.061***	0.001	-0.070***	-0.007
	(0.012)	(0.011)	(0.011)	(0.010)
Age squared	0.000*	-0.000	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.012	-0.075	0.369**	0.280**
	(0.266)	(0.200)	(0.150)	(0.118)
Secondary	0.524**	0.816***	1.170***	1.118***
	(0.269)	(0.164)	(0.238)	(0.223)
Higher	0.930***	0.684***	2.102***	1.487***
	(0.295)	(0.221)	(0.303)	(0.390)
Christianity	0.075*	0.007	-0.015	0.012
	(0.042)	(0.037)	(0.050)	(0.054)
Husband/Wife	-0.966***	-0.955***	-0.965***	-0.704***
	(0.037)	(0.061)	(0.062)	(0.069)
Cohabiting	-0.307***	-0.385***	-0.387***	-0.421***
	(0.031)	(0.049)	(0.053)	(0.036)
Other relationship type	0.171	0.304***	0.086	0.316**
	(0.115)	(0.091)	(0.116)	(0.143)
Perceived Unfaithfulness (Yes)	-0.023	0.155***	-0.024	0.125***
	(0.024)	(0.020)	(0.038)	(0.034)
One Partner	0.053	-0.041	0.013	-0.048
	(0.048)	(0.033)	(0.049)	(0.040)
PK	-0.055	0.034	0.305***	0.283***
	(0.109)	(0.074)	(0.065)	(0.055)
CRM	0.120***	0.119**	0.131***	0.085**
	(0.036)	(0.051)	(0.027)	(0.041)
Primary*PK	0.222**	0.243**	-0.020	-0.029
	(0.107)	(0.107)	(0.089)	(0.067)
Secondary*PK	0.161	-0.066	-0.124	-0.284***
	(0.120)	(0.084)	(0.126)	(0.077)
Higher*PK	0.113	-0.004	-0.407**	-0.058
	(0.149)	(0.123)	(0.175)	(0.135)
Primary*CRM	-0.069	-0.071	-0.021	0.026
	(0.048)	(0.055)	(0.038)	(0.039)
Secondary*CRM	-0.052	-0.094*	-0.052	-0.018
	(0.036)	(0.055)	(0.046)	(0.048)
Higher*CRM	-0.102**	-0.087	-0.092	-0.184***
	(0.046)	(0.064)	(0.058)	(0.066)
Partner drunk at last sex	-	-0.353***	-	-0.180***
		(0.035)		(0.055)
Observations	13259	13242	7234	7234
Exogeneity test for rural (Chi2, P>Chi2)	4.45	0.0349		
Exogeneity test in Urban(Chi2, P>Chi2)	2.42	0.100		

Notes: Each column presents results for separate regressions. Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.10: 2SLS results for the impact of perceived risk on condom use in rural areas

Variables	CUFS (1)	CUFS (2)	CULS (3)	CULS (4)
Less Concerned	-0.372**	-	-0.278*	-
	(0.145)		(0.164)	
More Concerned	-	0.630***	-	0.473
		(0.280)		(0.291)
Male	0.054***	0.053***	0.065***	0.064***
	(0.014)	(0.016)	(0.015)	(0.016)
Age	-0.008*	-0.010**	0.001	-0.001
	(0.004)	(0.005)	(0.004)	(0.005)
Age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.255***	0.182*	0.039	-0.016
	(0.095)	(0.102)	(0.098)	(0.098)
Secondary	0.420***	0.261**	0.415***	0.294***
	(0.080)	(0.104)	(0.084)	(0.106)
Higher	0.566***	0.252	0.496***	0.257
	(0.128)	(0.202)	(0.136)	(0.197)
Christianity	0.012	0.002	0.025	0.017
	(0.017)	(0.021)	(0.019)	(0.022)
Husband/Wife	-0.272***	-0.266***	-0.260***	-0.256***
	(0.024)	(0.027)	(0.025)	(0.027)
Cohabitation	-0.086***	-0.097***	-0.134***	-0.142***
	(0.016)	(0.018)	(0.019)	(0.019)
Other relationship type	0.045	0.056	0.104***	0.113***
	(0.037)	(0.041)	(0.035)	(0.037)
Perceived unfaithfulness (Yes)	-0.026*	-0.043**	0.023	0.013
	(0.015)	(0.021)	(0.017)	(0.022)
One Partner	0.019	0.009	-0.011	-0.018
	(0.020)	(0.022)	(0.022)	(0.023)
Partner drunk at last sex	-	-	-0.086***	-0.088***
			(0.026)	(0.027)
PK	0.090***	0.054	0.076**	0.050
	(0.032)	(0.040)	(0.033)	(0.040)
CRM	0.064***	0.043**	0.045**	0.029
	(0.020)	(0.019)	(0.021)	(0.019)
Primary*PK	-0.032	-0.019	0.021	0.031
	(0.046)	(0.052)	(0.048)	(0.050)
Secondary*PK	-0.060	-0.017	-0.110***	-0.078*
	(0.038)	(0.046)	(0.040)	(0.046)
Higher*PK	-0.088	-0.023	-0.040	0.008
	(0.056)	(0.073)	(0.060)	(0.073)
Primary*CRM	-0.041*	-0.035	0.003	0.007
	(0.025)	(0.026)	(0.026)	(0.025)
Secondary*CRM	-0.044**	-0.031	-0.029	-0.019
	(0.021)	(0.022)	(0.023)	(0.023)
Higher*CRM	-0.061*	-0.021	-0.083**	-0.051
	(0.034)	(0.040)	(0.035)	(0.036)
Observations	3985	3985	3981	3981
Exogeneity test (Chi2, P)	8.047	0.0046		
Instrument strength (F, P>F)	49.384	0.000		

Notes: Each column presents results for separate regressions. The dependent variable in columns 1 and 2 is condom use at first sex, while it is condom use at last sex in columns 3 and 4. Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. Instrumented: Less concerned and more concerned. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.11: 2SLS results for the impact of perceived risk on condom use in urban areas

Variables	CUFS (1)	CUFS (2)	CULS (3)	CULS (4)
Less Concerned	-0.459***	-	0.096	-
	(0.085)		(0.094)	
More Concerned	-	1.534**	-	-0.320
		(0.599)		(0.338)
Male	0.034***	0.038**	0.060***	0.059***
	(0.010)	(0.019)	(0.010)	(0.011)
Age	0.005	-0.000	0.004	0.005
	(0.003)	(0.007)	(0.003)	(0.004)
Age squared	-0.000***	-0.000*	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.013	-0.060	0.241*	0.254*
	(0.146)	(0.225)	(0.138)	(0.142)
Secondary	0.121	-0.060	0.443***	0.458***
	(0.132)	(0.199)	(0.121)	(0.123)
Higher	0.132	-0.107	0.405***	0.458***
	(0.142)	(0.261)	(0.136)	(0.157)
Christianity	0.009	-0.008	0.010	0.013
	(0.012)	(0.026)	(0.012)	(0.015)
Husband/Wife	-0.238***	-0.201***	-0.333***	-0.342***
	(0.015)	(0.033)	(0.015)	(0.019)
Cohabitation	-0.036***	-0.060***	-0.094***	-0.089***
	(0.011)	(0.022)	(0.012)	(0.013)
Other relationship type	0.005	0.007	0.073***	0.076***
	(0.023)	(0.056)	(0.019)	(0.022)
Perceived unfaithfulness (Yes)	0.002	-0.031	0.034***	0.040***
	(0.010)	(0.023)	(0.011)	(0.012)
One Partner	0.006	-0.025	-0.009	-0.003
	(0.011)	(0.028)	(0.012)	(0.015)
PK	-0.057	-0.112	0.084	0.093
	(0.066)	(0.093)	(0.056)	(0.059)
CRM	0.014	0.031	0.040	0.038
	(0.028)	(0.043)	(0.025)	(0.025)
Primary*PK	0.067	0.091	-0.048	-0.051
	(0.074)	(0.106)	(0.067)	(0.069)
Secondary*PK	0.046	0.025	-0.119**	-0.113**
	(0.067)	(0.080)	(0.058)	(0.058)
Higher*PK	0.037	0.050	-0.102	0.103
	(0.070)	(0.100)	(0.063)	(0.063)
Primary*CRM	-0.010	-0.026	-0.040	-0.039
	(0.031)	(0.049)	(0.028)	(0.028)
Secondary*CRM	-0.011	-0.020	-0.048*	-0.048*
	(0.028)	(0.045)	(0.026)	(0.026)
Higher*CRM	-0.016	0.019	-0.045*	-0.055*
	(0.030)	(0.051)	(0.028)	(0.030)
Partner drunk at last sex	-	-	-0.099***	-0.099***
			(0.018)	(0.020)
Observations	7719	7719	7694	7694
Exogeneity test (Chi2, P)	38.617	0.000		
Instrument Strength (F, P>F)	114.112	0.000		

Notes: Each column presents results for separate regressions. The dependent variable in columns 1 and 2 is condom use at first sex, while it is condom use at last sex in columns 3 and 4. Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. Instrumented: Less concerned and more concerned. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.12: Ordered Probit Marginal Effects (at means) of HIV on sexual partners among males

Variables	0 Partner	1 partner	>1 partner
HIV Rate	0.001	-0.000	-0.000
	(0.002)	(0.001)	(0.001)
Age	-0.053***	0.019***	0.033***
	(0.002)	(0.001)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.060**	0.022**	0.038**
	(0.025)	(0.009)	(0.016)
Secondary	-0.117***	0.043***	0.074***
	(0.030)	(0.011)	(0.019)
Higher	-0.120***	0.044***	0.076***
	(0.038)	(0.014)	(0.024)
Christianity	0.023***	-0.009***	-0.015***
	(0.007)	(0.002)	(0.004)
Cohabitation	-0.042***	0.015***	0.026***
	(0.009)	(0.004)	(0.006)
Never married	0.074***	-0.027***	-0.047***
	(0.015)	(0.005)	(0.009)
Other Marital status	0.150***	-0.055***	-0.095***
	(0.025)	(0.009)	(0.016)
Rural	0.010	-0.004	-0.006
	(0.013)	(0.005)	(0.008)
PK	-0.050***	0.018***	0.032***
	(0.012)	(0.004)	(0.008)
CRM	-0.010	0.004	0.007
	(0.008)	(0.003)	(0.005)
Primary*PK	0.016	-0.006	-0.010
	(0.017)	(0.006)	(0.011)
Secondary*PK	0.003	-0.001	-0.002
	(0.015)	(0.005)	(0.009)
Higher*PK	-0.005	0.002	0.003
	(0.018)	(0.007)	(0.011)
Primary*CRM	0.007	-0.002	-0.004
	(0.009)	(0.003)	(0.006)
Secondary*CRM	0.010	-0.004	-0.006
	(0.010)	(0.004)	(0.006)
Higher*CRM	0.003	-0.001	-0.002
	(0.010)	(0.004)	(0.006)
Wave 3 (2008)	-0.014	0.005	0.009
	(0.009)	(0.004)	(0.006)
Wave 4 (2013)	-0.086***	0.032***	0.054***
	(0.012)	(0.005)	(0.007)
District FE	Yes	Yes	Yes
Observations	13188	13188	13188

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, married for marital status. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.13: Ordered Probit Marginal Effects (at means) of HIV on sexual partners among Females

Variables	0 Partner	1 partner	>1 partner
HIV Rate	0.000	-0.000	-0.000
	(0.002)	(0.001)	(0.000)
Age	-0.054***	0.040***	0.013***
	(0.002)	(0.002)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.080*	0.060*	0.020*
	(0.043)	(0.033)	(0.011)
Secondary	-0.032	0.024	0.008
	(0.035)	(0.027)	(0.009)
Higher	-0.063	0.048	0.016
	(0.052)	(0.039)	(0.013)
Christianity	0.067***	-0.050***	-0.017***
	(0.010)	(0.007)	(0.002)
Cohabitation	-0.006	0.004	0.001
	(0.010)	(0.007)	(0.002)
Never married	0.212***	-0.160***	-0.053***
	(0.013)	(0.010)	(0.003)
Other Marital status	0.354***	-0.266***	-0.088***
	(0.022)	(0.017)	(0.005)
Rural	0.012	-0.009	-0.003
	(0.009)	(0.007)	(0.002)
PK	-0.022	0.017	0.006
	(0.019)	(0.015)	(0.005)
CRM	-0.011	0.008	0.003
	(0.009)	(0.007)	(0.002)
Primary*PK	0.006	-0.004	-0.001
	(0.023)	(0.017)	(0.006)
Secondary*PK	-0.033	0.025	0.008
	(0.021)	(0.016)	(0.005)
Higher*PK	-0.005	0.004	0.001
	(0.027)	(0.020)	(0.007)
Primary*CRM	0.017	-0.013	-0.004
	(0.011)	(0.008)	(0.003)
Secondary*CRM	0.013	-0.010	-0.003
	(0.011)	(0.008)	(0.003)
Higher*CRM	-0.003	0.002	0.001
	(0.011)	(0.008)	(0.003)
Wave 3 (2008)	-0.022**	0.017**	0.006**
	(0.010)	(0.007)	(0.002)
Wave 4 (2013)	-0.070***	0.053***	0.017***
	(0.011)	(0.009)	(0.003)
District FE	Yes	Yes	Yes
Observations	16413	16413	16413

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, married for marital status. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.14: Ordered Probit Marginal Effects (at means) of Perceived risk on sexual partners in male

Variables	0 Partner	1 partner	>1 partner
Less concerned	-0.016*	0.003	0.014*
	(0.010)	(0.002)	(0.008)
More concerned	-0.002	0.000	0.002
	(0.009)	(0.001)	(0.008)
Age	-0.043***	0.007***	0.037***
	(0.002)	(0.001)	(0.002)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.004	0.001	0.003
	(0.054)	(0.008)	(0.046)
Secondary	-0.023	0.004	0.019
	(0.050)	(0.008)	(0.043)
Higher	-0.052	0.008	0.044
	(0.065)	(0.010)	(0.045)
Christianity	0.014*	-0.002	-0.012
	(0.009)	(0.001)	(0.007)
Cohabitation	-0.038***	0.006***	0.032***
	(0.010)	(0.002)	(0.008)
Never married	0.035***	-0.005**	-0.029***
	(0.011)	(0.002)	(0.009)
Other Marital status	0.079***	-0.012***	-0.067***
	(0.028)	(0.005)	(0.024)
Rural	0.024**	-0.004**	-0.020**
	(0.011)	(0.012)	(0.009)
PK	-0.026	0.004	0.022
	(0.020)	(0.003)	(0.017)
CRM	0.003	-0.000	-0.003
	(0.010)	(0.001)	(0.008)
Primary*PK	-0.018	0.003	0.015
	(0.027)	(0.004)	(0.023)
Secondary*PK	-0.028	0.004	0.024
	(0.024)	(0.004)	(0.020)
Higher*PK	-0.014	0.002	0.011
	(0.028)	(0.004)	(0.024)
Primary*CRM	0.008	-0.001	-0.006
	(0.012)	(0.002)	(0.010)
Secondary*CRM	-0.002	0.000	0.002
	(0.012)	(0.002)	(0.010)
Higher*CRM	-0.009	0.001	0.008
	(0.014)	(0.002)	(0.012)
Wave 4 (2013)	-0.066***	0.010***	0.056***
	(0.009)	(0.002)	(0.007)
District FE	Yes	Yes	Yes
Observations	7116	7116	7116

Notes: Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type, no change for change in concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.15: Ordered Probit Marginal Effects (at means) of Perceived risk on sexual partners in females

Variables	0 Partner	1 partner	>1 partner
Less concerned	-0.012	0.008	0.004
	(0.010)	(0.007)	(0.003)
More concerned	-0.006	0.004	0.002
	(0.010)	(0.007)	(0.003)
Age	-0.048***	0.033***	0.015***
	(0.002)	(0.002)	(0.001)
Age squared	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Primary	-0.015	0.010	0.005
	(0.059)	(0.041)	(0.019)
Secondary	-0.028	0.019	0.009
	(0.055)	(0.038)	(0.017)
Higher	-0.046	0.031	0.014
	(0.085)	(0.058)	(0.027)
Christianity	0.059***	-0.040***	-0.018***
	(0.012)	(0.008)	(0.004)
Cohabitation	-0.014	0.010	0.004
	(0.010)	(0.007)	(0.003)
Never married	0.173***	-0.118***	-0.054***
	(0.011)	(0.009)	(0.003)
Other Marital status	0.288***	-0.198***	-0.091***
	(0.022)	(0.017)	(0.007)
Rural	0.010	-0.007	-0.003
	(0.011)	(0.008)	(0.003)
HIV Prevention Knowledge	-0.023	0.015	0.007
	(0.023)	(0.015)	(0.007)
Correcrejecmis	0.001	-0.001	-0.000
	(0.012)	(0.008)	(0.004)
Primary*HIV Knowledge	-0.011	0.007	0.003
	(0.030)	(0.021)	(0.010)
Secondary*HIV Knowledge	-0.030	0.021	0.009
	(0.026)	(0.018)	(0.008)
Higher*HIV Knowledge	0.002	-0.002	-0.001
	(0.037)	(0.026)	(0.012)
Primary*Correcrejecmis	0.005	-0.004	-0.002
	(0.015)	(0.010)	(0.005)
Secondary*Correcrejecmis	0.010	-0.007	-0.003
	(0.014)	(0.010)	(0.004)
Higher*Correcrejecmis	-0.015	0.010	0.005
	(0.019)	(0.013)	(0.006)
Wave 4 (2013)	-0.046***	0.031***	0.014***
	(0.009)	(0.006)	(0.003)
District FE	Yes	Yes	Yes
Observations	9140	9140	9140

Notes: Standard are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type, no change in change in concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.16: Probit Marginal Effects of HIV Rate on condom use by gender

	Males		Females	
Variables	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
HIV Rate	0.004*	0.006***	0.004*	0.000
	(0.002)	(0.002)	(0.002)	(0.003)
Age	-0.016***	-0.014***	-0.017***	0.007**
	(0.003)	(0.003)	(0.002)	(0.003)
Age squared	0.000**	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.045	0.052	0.144***	0.103*
	(0.039)	(0.061)	(0.054)	(0.060)
Secondary	0.204***	0.383***	0.248***	0.335***
	(0.053)	(0.064)	(0.055)	(0.064)
Higher	0.347***	0.267***	0.316***	0.442***
	(0.076)	(0.089)	(0.064)	(0.073)
Christianity	0.007	0.008	0.019*	0.018
	(0.010)	(0.009)	(0.011)	(0.017)
Husband/Wife	-0.208***	-0.255***	-0.237***	-0.270***
	(0.009)	(0.014)	(0.008)	(0.020)
Cohabiting	-0.078***	-0.130***	-0.091***	-0.117***
	(0.008)	(0.012)	(0.011)	(0.010)
Other relationship type	0.037	0.099***	-0.018	0.027
	(0.026)	(0.027)	(0.031)	(0.043)
Perceived Unfaithfulness (Yes)	-0.017	0.056***	-0.011*	0.031***
	(0.011)	(0.012)	(0.006)	(0.005)
Rural	-0.051***	-0.055***	-0.047***	-0.059***
	(0.005)	(0.012)	(0.008)	(0.008)
One Partner	0.027***	0.010	-0.015	-0.044***
	(0.008)	(0.008)	(0.017)	(0.017)
PK	0.025	0.056***	0.063***	0.077***
	(0.021)	(0.017)	(0.022)	(0.029)
CRM	0.019**	0.025	0.027***	0.029**
	(0.008)	(0.016)	(0.010)	(0.014)
Primary*PK	0.033	0.029	-0.034	-0.010
	(0.021)	(0.026)	(0.026)	(0.034)
Secondary*PK	-0.016	-0.089***	-0.029	-0.076**
	(0.027)	(0.018)	(0.030)	(0.032)
Higher*PK	-0.026	-0.040	-0.053**	-0.055
	(0.038)	(0.036)	(0.027)	(0.042)
Primary*CRM	-0.017*	-0.008	-0.011	-0.006
	(0.010)	(0.015)	(0.012)	(0.020)
Secondary*CRM	-0.012	-0.029*	-0.017	-0.015
	(0.008)	(0.016)	(0.013)	(0.017)
Higher*CRM	-0.041***	-0.022	-0.008	-0.048***
	(0.010)	(0.019)	(0.017)	(0.019)
Partner drunk at last sex	-	-0.093***	-	-0.092***
		(0.012)		(0.019)
Wave 3 (2008)	0.093***	0.070***	0.094***	0.060***
	(0.008)	(0.013)	(0.007)	(0.010)
Wave 4 (2013)	0.172***	0.054***	0.173***	0.072***
	(0.011)	(0.014)	(0.013)	(0.015)
District FE	Yes	Yes	Yes	Yes
Observations	9273	9265	11220	11210
Loglikelihood	-3641.604	-4679.703	-4663.476	-6325.804

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. CUFS: Condom use at first sex, CULS: Condom use at last sex. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.17: Probit Marginal Effects of Perceived risk on condom use by gender

	Males		Females	
Variables	CUFS	CULS	CUFS	CULS
Less Concerned	0.017	0.022	0.019*	0.011
	(0.012)	(0.014)	(0.011)	(0.014)
More Concerned	0.020*	0.028**	0.019*	0.045***
	(0.011)	(0.013)	(0.010)	(0.013)
Age	-0.010***	-0.011***	-0.010***	0.010***
	(0.003)	(0.004)	(0.003)	(0.004)
Age squared	0.000	0.000**	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.143**	0.058	0.114	0.126
	(0.063)	(0.088)	(0.071)	(0.101)
Secondary	0.247***	0.370***	0.188***	0.453***
	(0.063)	(0.083)	(0.064)	(0.089)
Higher	0.303***	0.263**	0.257***	0.488***
	(0.081)	(0.102)	(0.098)	(0.119)
Christianity	0.012	0.017	0.028**	0.033**
	(0.010)	(0.013)	(0.012)	(0.015)
Husband/Wife	-0.182***	-0.262***	-0.218***	-0.277***
	(0.014)	(0.017)	(0.012)	(0.015)
Cohabitation	-0.074***	-0.113***	-0.081***	-0.108***
	(0.012)	(0.014)	(0.011)	(0.013)
Other relationship type	0.022	0.142***	-0.008	0.072
	(0.030)	(0.038)	(0.053)	(0.057)
Perceived Unfaithfulness	-0.012	0.057***	-0.017	0.025**
	(0.013)	(0.016)	(0.010)	(0.013)
Rural	-0.055***	-0.066***	-0.047***	-0.066***
	(0.013)	(0.016)	(0.012)	(0.015)
One Partner	0.014	0.007	-0.018	-0.062***
	(0.013)	(0.015)	(0.018)	(0.021)
PK	0.044*	0.030	0.028	0.116***
	(0.024)	(0.033)	(0.026)	(0.038)
CRM	0.022*	0.027*	0.025**	0.035*
	(0.012)	(0.016)	(0.013)	(0.018)
Primary*PK	-0.020	0.014	-0.018	-0.034
	(0.031)	(0.044)	(0.035)	(0.049)
Secondary*PK	-0.041	-0.089**	-0.008	-0.142***
	(0.030)	(0.040)	(0.030)	(0.042)
Higher*PK	-0.051	-0.047	-0.032	-0.101*
	(0.036)	(0.045)	(0.042)	(0.053)
Primary*CRM	-0.024	-0.006	-0.008	-0.009
	(0.015)	(0.019)	(0.015)	(0.022)
Secondary*CRM	-0.022	-0.035**	-0.012	-0.031*
	(0.014)	(0.018)	(0.015)	(0.021)
Higher* CRM	-0.031*	-0.029	-0.007	-0.058**
	(0.018)	(0.022)	(0.020)	(0.026)
Partner drunk at last sex	-	-0.103***	-	-0.086***
		(0.019)		(0.020)
Wave 4 (2013)	0.062***	0.021*	0.073***	-0.001
	(0.011)	(0.013)	(0.010)	(0.012)
District FE	Yes	Yes	Yes	Yes
Observations	5245	5232	6459	6443
Loglikelihood	-1903.769	-2566.311	-2532.648	-3591.979
Prob>chi2	0.000	0.000	0.000	0.000

Notes: Each column lists results for separate regression. Robust standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type, no change for change in concern level. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Accounting for endogeneity in Males/Females sub-samples

Table A4.18: IV Probit Results for Condom use by Gender: Objective Risk

	Males		**Females**	
Variables	CUFS	CULS	CUFS	CULS
	(1)	(2)	(3)	(4)
HIV Rate	-0.008	0.005	-0.012	0.014
	(0.015)	(0.010)	(0.010)	(0.011)
Age	-0.059***	-0.041***	-0.070***	0.023***
	(0.011)	(0.008)	(0.010)	(0.009)
Age squared	0.000**	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.208	0.184	0.510**	0.294
	(0.146)	(0.215)	(0.217)	(0.181)
Secondary	1.059***	1.362***	1.111***	1.050***
	(0.219)	(0.227)	(0.218)	(0.200)
Higher	1.655***	0.969***	1.472***	1.419***
	(0.309)	(0.305)	(0.241)	(0.219)
Christianity	0.024	-0.001	0.074*	0.038
	(0.045)	(0.033)	(0.039)	(0.051)
Husband/Wife	-0.939***	-0.901***	-0.999***	-0.850***
	(0.041)	(0.059)	(0.041)	(0.068)
Cohabiting	-0.338***	-0.447***	-0.346***	-0.354***
	(0.037)	(0.045)	(0.038)	(0.035)
Other relationship type	0.225*	0.366***	-0.039	0.102
	(0.118)	(0.092)	(0.120)	(0.129)
Perceived Unfaithfulness (Yes)	-0.030	0.212***	-0.026	0.096***
	(0.045)	(0.042)	(0.027)	(0.019)
Rural	-0.203***	-0.203***	-0.183***	0.169***
	(0.038)	(0.038)	(0.028)	(0.029)
One Partner	0.096**	0.031	-0.072	-0.139***
	(0.039)	(0.026)	(0.071)	(0.049)
PK	0.150*	0.208***	0.272***	0.251***
	(0.086)	(0.061)	(0.086)	(0.084)
CRM	0.106***	0.094*	0.159***	0.100**
	(0.033)	(0.054)	(0.041)	(0.044)
Primary*PK	0.152*	0.103	-0.095	-0.023
	(0.083)	(0.090)	(0.101)	(0.104)
Secondary*PK	-0.089	-0.315***	-0.099	-0.235**
	(0.112)	(0.067)	(0.117)	(0.095)
Higher*PK	-0.141	-0.152	-0.219**	-0.183
	(0.161)	(0.119)	(0.099)	(0.126)
Primary*CRM	-0.065	-0.026	-0.047	-0.015
	(0.043)	(0.053)	(0.050)	(0.062)
Secondary*CRM	-0.052	-0.096*	-0.076	-0.043
	(0.037)	(0.055)	(0.056)	(0.053)
Higher*CRM	-0.175***	-0.076	-0.048	-0.148**
	(0.044)	(0.066)	(0.068)	(0.060)
Partner drunk at last sex	-	-0.336***	-	-0.239***
		(0.043)		(0.066)
Observations	9273	9265	11220	11210
Exogeneity test in Males(Chi2, Pvalue)	5.53	0.0187		
Exogeneity test in females(Chi2, Pval)	2.55	0.100		

Notes: Each column lists results for separate regression. Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.19: 2SLS results for the impact of perceived risk on condom use among males

Variables	CUFS (1)	CUFS (2)	CULS (3)	CULS (4)
Less Concerned	-0.385***	-	0.036	-
	(0.114)		(0.127)	
More Concerned	-	1.413*	-	-0.128
		(0.835)		(0.465)
Age	0.003	-0.007	-0.006	-0.005
	(0.004)	(0.009)	(0.004)	(0.005)
Age squared	-0.000***	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.326***	0.018	0.105	0.134
	(0.109)	(0.212)	(0.112)	(0.144)
Secondary	0.473***	0.072	0.407***	0.443***
	(0.091)	(0.255)	(0.094)	(0.155)
Higher	0.543***	-0.028	0.337***	0.390*
	(0.108)	(0.363)	(0.118)	(0.216)
Christianity	0.003	-0.013	0.011	0.012
	(0.013)	(0.030)	(0.014)	(0.017)
Husband/Wife	-0.233***	-0.184***	-0.317***	-0.322***
	(0.020)	(0.046)	(0.020)	(0.027)
Cohabitation	-0.059***	-0.058**	-0.114***	-0.114***
	(0.013)	(0.027)	(0.015)	(0.015)
Other relationship type	0.027	0.019	0.085***	0.086***
	(0.021)	(0.047)	(0.018)	(0.019)
Perceived unfaithfulness (Yes)	-0.003	-0.042	0.050***	0.053***
	(0.012)	(0.033)	(0.013)	(0.017)
Rural	-0.057***	-0.068***	-0.078***	-0.076***
	(0.012)	(0.023)	(0.013)	(0.014)
One Partner	0.013	-0.014	0.010	0.013
	(0.012)	(0.030)	(0.013)	(0.016)
PK	0.092**	-0.059	0.046	0.057
	(0.042)	(0.099)	(0.043)	(0.060)
CRM	0.043*	0.014	0.031	0.034
	(0.022)	(0.034)	(0.021)	(0.023)
Primary*PK	-0.064	0.059	0.002	-0.009
	(0.054)	(0.099)	(0.055)	(0.065)
Secondary*PK	-0.108**	-0.024	-0.093**	-0.100**
	(0.045)	(0.077)	(0.046)	(0.049)
Higher*PK	-0.117**	0.009	-0.064	-0.075
	(0.052)	(0.106)	(0.053)	(0.064)
Primary*CRM	-0.044*	-0.034	-0.005	-0.006
	(0.026)	(0.039)	(0.025)	(0.025)
Secondary*CRM	-0.037	0.002	-0.037*	-0.041
	(0.023)	(0.040)	(0.022)	(0.026)
Higher*CRM	-0.061**	0.013	-0.035	-0.043
	(0.026)	(0.054)	(0.026)	(0.034)
Partner drunk at last sex	-	-	-0.102***	-0.102***
			(0.020)	(0.021)
Observations	5245	5245	5232	5232
Exogeneity test (Chi2, P value)	14.233	0.0002		
Instrument strength (F, P value)	62.360	0.0000		

Notes: Each column lists results for separate regression. In the first two columns the dependent variable is condom use at first sex, while its condom use at last sex in the last columns. Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. Instrumented: *Less concerned* and *more concerned.* CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table A4.20: 2SLS results for the impact of perceived risk on condom use among females

Variables	CUFS (1)	CUFS (2)	CULS (3)	CULS (4)
Less Concerned	-0.455***	-	-0.066	-
	(0.095)		(0.105)	
More Concerned	-	1.011***	-	0.148
		(0.308)		(0.235)
Age	0.002	-0.000	0.012***	0.011***
	(0.004)	(0.005)	(0.004)	(0.004)
Age squared	-0.000***	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.154	0.186	0.131	0.136
	(0.107)	(0.138)	(0.101)	(0.100)
Secondary	0.389***	0.317***	0.499***	0.488***
	(0.087)	(0.112)	(0.081)	(0.084)
Higher	0.410***	0.199	0.531***	0.499***
	(0.110)	(0.178)	(0.112)	(0.129)
Christianity	0.021	-0.006	0.031*	0.027
	(0.015)	(0.024)	(0.016)	(0.019)
Husband/Wife	-0.258***	-0.247***	-0.308***	-0.306***
	(0.017)	(0.023)	(0.017)	(0.018)
Cohabitation	-0.050***	-0.078***	-0.104***	-0.108***
	(0.012)	(0.017)	(0.014)	(0.013)
Other relationship type	0.004	0.051	0.061	0.068
	(0.044)	(0.067)	(0.042)	(0.043)
Perceived unfaithfulness	-0.009	-0.035*	0.025***	0.021
	(0.011)	(0.018)	(0.012)	(0.013)
Rural	-0.052***	-0.056***	-0.062***	-0.063***
	(0.012)	(0.017)	(0.012)	(0.013)
One Partner	0.000	-0.016	-0.048***	-0.050***
	(0.016)	(0.024)	(0.017)	(0.017)
PK	0.021	0.001	0.118***	0.115***
	(0.041)	(0.052)	(0.037)	(0.038)
CRM	0.068***	0.063**	0.041**	0.040**
	(0.022)	(0.027)	(0.021)	(0.021)
Primary*PK	0.025	-0.015	-0.024	-0.030
	(0.055)	(0.070)	(0.051)	(0.050)
Secondary*PK	-0.005	-0.016	-0.144***	-0.145***
	(0.044)	(0.055)	(0.040)	(0.040)
Higher*PK	-0.027	-0.034	-0.112**	-0.113**
	(0.050)	(0.069)	(0.051)	(0.051)
Primary*CRM	-0.036	-0.041	-0.012	-0.013
	(0.026)	(0.034)	(0.025)	(0.025)
Secondary*CRM	-0.057**	-0.055*	-0.040*	-0.040*
	(0.023)	(0.029)	(0.023)	(0.023)
Higher*CRM	-0.052*	-0.009	-0.063**	-0.057**
	(0.027)	(0.037)	(0.027)	(0.027)
Partner drunk at last sex	-	-	-0.085***	-0.086***
			(0.021)	(0.021)
Observations	6459	6459	6443	6443
Exogeneity test (Chi2, Pvalue)	29.358	0.0000		
Instrument Strength (F, Pvalue)	105.312	0.0000		

Notes: Each column lists results for separate regression. In the first two columns the dependent variable is condom use at first sex, while its condom use at last sex in the last columns. Standard errors are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. Instrumented: Less concerned and more concerned. CUFS: Condom use at first sex, CULS: Condom use at last sex. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

One vs More than one sexual partner

Table A4.21: Second Stage Results for Condom use: One partner vs more than one partner

	One Partner		More than One partner	
Variables	IV Probit	2SLS	IV Probit	2SLS
	(1)	(2)	(3)	(4)
HIV Rate	-0.001	-0.001	0.009	0.005
	(0.010)	(0.003)	(0.014)	(0.004)
Male	0.268***	0.201***	0.133**	0.084**
	(0.025)	(0.008)	(0.055)	(0.015)
Age	-0.037***	-0.033***	-0.073***	-0.048***
	(0.007)	(0.002)	(0.021)	(0.005)
Age squared	0.000*	0.000***	0.001**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Primary	0.182	0.097	1.025***	0.686***
	(0.153)	(0.039)	(0.392)	(0.102)
Secondary	0.990***	0.756***	2.750***	2.058***
	(0.183)	(0.051)	(0.418)	(0.094)
Higher	1.171***	0.881***	2.704***	2.072***
	(0.228)	(0.066)	(0.507)	(0.103)
Christianity	0.035	0.025	-0.064	-0.045
	(0.030)	(0.009)	(0.073)	(0.018)
Husband/Wife	-1.032***	-0.889***	-1.347***	-1.170***
	(0.042)	(0.014)	(0.100)	(0.029)
Cohabiting	-0.424***	-0.361***	-0.600***	-0.489***
	(0.024)	(0.007)	(0.066)	(0.022)
Other relationship type	0.181*	0.111*	0.456***	0.233***
	(0.105)	(0.027)	(0.130)	(0.020)
Rural	-0.206***	-0.160***	-0.120**	-0.085**
	(0.025)	(0.007)	(0.052)	(0.013)
Perceived Unfaithfulness (Yes)	0.047**	0.036***	0.060	0.041
	(0.021)	(0.006)	(0.053)	(0.014)
PK	0.168**	0.100**	0.413**	0.314**
	(0.070)	(0.018)	(0.165)	(0.045)
CRM	0.072**	0.051**	0.482***	0.359***
	(0.031)	(0.009)	(0.118)	(0.031)
Primary*PK	0.072	0.078	-0.019	0.039
	(0.077)	(0.020)	(0.208)	(0.060)
Secondary*PK	-0.148*	-0.068	-0.625***	-0.445***
	(0.087)	(0.024)	(0.203)	(0.052)
Higher*PK	-0.173	-0.103	-0.401	-0.309**
	(0.108)	(0.031)	(0.256)	(0.061)
Primary*CRM	-0.014	-0.007	-0.358***	-0.264***
	(0.034)	(0.010)	(0.127)	(0.035)
Secondary*CRM	-0.035	-0.023	-0.454***	-0.339***
	(0.033)	(0.010)	(0.128)	(0.034)
Higher*CRM	-0.050	-0.033	-0.489***	-0.370***
	(0.045)	(0.014)	(0.165)	(0.040)
Observations	17634	17634	2854	2854

Notes: Standard errors clustered at district level, are in parenthesis. Omitted categories are none for education, boyfriend/girlfriend for relationship type. 2SLS results have been rescaled by 2.5 (Amemiya, 1981). PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

HIV prevalence and perceived risk

Table A4.22: Probit Marginal effects of those who increased level of concern

Variables	Marginal Effect (1)	Standard errors (2)
HIV Rate	0.004***	0.001
Male	0.001	0.009
Age	0.004	0.003
Age squared	-0.000	0.000
Primary	0.120	0.082
Secondary	0.183**	0.073
Higher	0.322***	0.090
Christianity	0.027**	0.011
Husband/Wife	-0.022*	0.013
Cohabiting	0.003	0.010
Other relationship type	-0.012	0.027
Rural	-0.023**	0.010
Perceived Unfaithfulness (Yes)	0.020**	0.010
PK	0.082**	0.033
CRM	-0.002	0.015
Primary*PK	-0.053	0.041
Secondary*PK	-0.041	0.036
Higher*PK	-0.062	0.042
Primary*CRM	0.012	0.017
Secondary*CRM	-0.002	0.016
Higher*CRM	-0.006	0.019
ARVMBS	0.001***	0.000
Observations	11731	
Loglikelihood	-7152.027	

Notes: Standard errors are clustered at district level. Omitted categories are none for education, boyfriend/girlfriend for relationship type. PK: HIV prevention knowledge, CRM: Correct rejection of misconceptions. ***Significant at the 1 per cent level, **Significant at the 5 per cent level, *Significant at the 10 per cent level.

Table B4.1: Summary of the empirical evidence

Author(s)/Country	Investigate:	Data	Est. Method	Main Results
Objective risk				
Lakdawalla et al. (2006): United States	Being in HAART and sexual behaviour	3 waves of panel from HIV Costs and Services Utilization Study	Treatment effect model. Access to insurance (Medicaid) used as an instrument for being on HAART	Receipt of HAART increases the number of sexual partners by about 4
Mechoulan (2007): United States	HIV testing, being in HAART and sexual behaviour	Non repeated cross section collected by Sop Aids Project agency	Linear Probability Model	Testers are more likely to engage in unprotected sex, while non-testers have fewer partners.
Keetile and Kgosidintsi (2018): Botswana	Being in HAART and sexual behaviour (inconsistent condom use, multiple sexual partners, intergenerational and transactional sex)	2013 Botswana AIDS Impact Survey	Logistic regression	There is no association between being in HAART and sexual behaviour.
Ahituv et al. (1996): United States	HIV/AIDS prevalence and use of condoms	Panel for young adults for the period 1984-1990	Logit model	An increase in HIV prevalence increases the likelihood of using condoms
Oster (2012): sub-Saharan Africa (14 countries)	HIV Prevalence and sexual behaviour (more than one partner, more than one partner and non-use of condoms with a secondary partner, number of non-marital partners)	Demographic and Health Surveys	Instrumental Variable method. Uses distance from African HIV origin (DRC) as an instrument for sexual behaviour	Doubling of HIV prevalence leads to a 20% decline in the likelihood of multiple sexual partnerships and a 30% decline in the likelihood of multiple partnerships with no condom.
Keetile (2014): Botswana	High-risk factors and sexual behaviour (inconsistent condom use and multiple partners) and implications for HIV/AIDS prevention efforts	2008 Botswana AIDS Impact Survey	Logistic regression	Having ever tested for HIV/AIDS is associated with an increased likelihood of inconsistent condom use and reduced incidence of multiple partnerships
Malema (2012): Botswana	Condom use across men and women	2004 Botswana AIDS Impact Survey	Logistic regression	Being employed is associated with reduced likelihood of condom use by both men and women (attributed to the observation that the employed had tested for HIV/AIDS)
Subjective indicator studies				
Hart and Williamson (2005): Scotland	HAART-induced perception and sexual behaviour (unprotected sex with casual partners and	Repeated cross sections (1996, 1999, 2002)	Logistic regression	Men less concerned about being infected with HIV as a result of ART are more likely to engage in an unsafe

	with more than one partner)			sexual act with casual partners and with more than one partner
Keetile and Letamo (2015): Botswana	Beliefs and attitudes about ART and condom use	2008 Botswana AIDS Impact Survey	Logistic regression	A positive association between the belief that ART cures AIDS and inconsistent condom use
Letamo et al. (2017): Botswana	ARV perception and sexual behaviour (condom use, multiple partners)	2013 Botswana AIDS Impact Survey	Logistic regression	Believing that ARV cures HIV/AIDS is associated with a reduced likelihood of unsafe sex and multiple partners.
Wilson et al. (2014): Kenya	Male circumcision and sexual behaviour (number of partners)	From a nested study in RCT, which randomised receipt of male circumcision	OLS, LPM, Fixed and random effect models	Circumcision-induced beliefs lead to a reduced number of partners and the likelihood (10-20%) of having multiple sexual partners and an increase in condom use
Godlonton et al. (2016): Malawi	Learning the risk reduction effect of male circumcision and impact on sexual behaviour (number of partners and use of condoms)	Experimental: Information on male circumcision randomly provided to participants at baseline. A follow-up survey a year later to source information on belief and sexual behaviour sourced	Binary probit regressions	Providing information on male circumcision induces the circumcised males to reduce the number of partners and increase the use of condoms. Circumcised men also updated their beliefs about the risk reduction effect of male circumcision.
Balekang and Dintwa (2016): Botswana	Male circumcision and risky sexual behaviour (number of sexual partners, sexual intercourse while intoxicated, condom use)	2008 Botswana AIDS Impact Survey	Logistic regression	Non-circumcised men with secondary education less likely to have more than one sexual partner. Primary education positively associated with an increased likelihood of more than one sexual partner
Agiya and Letamo (2011): Botswana	Male circumcision and condom use	2008 Botswana AIDS Impact Survey	Pearson Chi-square statistic Logistic regression	Male circumcision has no effect on condom use
Both risks				
Thornton (2008): Malawi	Learning HIV status and sexual behaviour	2004 wave of the panel data with the experimental component: Participants offered HIV	Instrumental Variable: Learning HIV status is instrumented with monetary incentive and	Those informed of their HIV positive status were more likely to purchase condoms and bought two more condoms than those who do not know their

		test and randomly assigned a monetary incentive to obtain information on their HIV status	distance to the HIV test result centre	status. No effect on those who tested HIV negative.
Delavande and Kohler (2012): Malawi	HIV testing, subjective expectations and sexual behaviour	2004 and 2006 waves of Panel, with experimental HIV testing design of random assignment to financial incentive to learn HIV status	Control function approach for attrition between 2004 and 2006	Learning HIV negative status in 2004 is associated with increased (and less accurate) subjective probability of being infected in 2006. Among the HIV+, learning one`s status has a lower perception about the likelihood of being infected while learning HIV- status has no effect Knowing a positive HIV status leads to a reduced number of partners, rates of pregnancy (among the non-married), and increased likelihood of condom use. No effect on those who tested negative for HIV.
Gong (2015): Tanzania and Kenya	HIV testing and sexual behaviour	1995-1998 panel data from HIV Voluntary Counselling and Testing Efficacy study, which randomly offered HIV testing and counselling to participants	Linear Probability Model	Those surprised by HIV positive test results increase their risky sexual behaviour while those surprised by negative test results were less likely to engage in risky sexual behaviour
Baird et al. (2014): Malawi	HIV testing and sexual behaviour	Experimental: Participants randomly assigned HIV testing and counselling	Linear Probability Model	Among those who discover to be HIV positive at baseline (n=73), the prevalence of HSV2 is 23% higher at follow up.
Paula et al. (2014): Malawi	Updated believes and sexual behaviour	Panel (2006 and 2008 waves of the Diffusion and Ideal Change Project	Semi parametric panel estimator	Men who believe to have a higher likelihood of HIV infection are less likely to report extra marital sexual relationship or

				multiple sex partners. Those with low belief are more likely to report risky behaviour
Francis (2008): United States	Having a relative living or dead from AIDS and sexual behaviour	1992 National Health and Social Survey	Linear Probability Model	Having a relative with or dead from AIDS makes some men shift from homosexual activity to heterosexual ones, and women from heterosexual to homosexual activities.
Tenkorang et al (2011): South Africa	Perceptions and sexual behaviour	Panel	Ordered logit model	Those who perceive themselves to be more susceptible to infection are less likely to report risky behaviour.
Dupas (2011): Kenya	Responding to HIV risk information	Experimental: Random provision of information on risk avoidance (In some schools' teachers trained on HIV while in others information on risk reduction was assigned to teenagers on the 8th grade.	Difference in difference, accounting for schools` unobservable effect.	Risk reduction information leads to reduced incidence of pregnancy, while risk avoidance has no effect.
Akwara et al. (2003): Kenya	Perceived likelihood of infection and sexual behaviour	1998 Demographic and Health survey	Logistic regression	Higher perceived risk is associated with less likelihood of risky behaviour

CHAPTER 5: CONCLUSIONS

This thesis has sought to address three questions: does the utilization of prenatal care have any effect on infant birth weight in Botswana? What is the role of birth order and household size on child nutritional indicators? What is the impact of objective and perceived HIV risk on sexual behaviour? These questions, addressed in the first, second and third essay respectively, are motivated by the observed deterioration in survey-based health indicators in Botswana and limited existing evidence identifying possible reasons. Data show an increasing proportion of low birth weight (i.e., less than 2.5kg) and malnourished children (stunted, underweight and wasted), as well as increases in the prevalence of HIV/AIDS. The pattern for birth weight placed Botswana among the worst performers in the WHO African region and the upper-middle-income category countries. The trend in rates of child malnutrition brought into question why the reduction in malnutrition rates achieved over the 1990s was reversed.

5.1 Summary of findings

In answering the first question, the first essay (chapter 2) accounts for the endogeneity of prenatal care utilization using the probability of care facility being sufficiently close and availability of care facilities as instruments and potential sample selection due to the non-random absence of infant`s birth weight. The main result is that adequate prenatal care utilization is positively associated with infant birth weight. Failure to account for endogeneity may understate the effect. We show evidence that on average for mothers that fully utilise prenatal care the birth weight increases by 0.67 kg in the full (national) sample, 0.73 kg in the urban sample and 0.64 kg in the rural sample. We do not find the presence of sample selection bias in either the rural or urban samples.

The second essay (chapter 3) answers the question on the effect of household size and childbirth order on child nutrition as measured by the three anthropometric indicators (height for age, weight for age and weight for height). A random effects model is estimated to explore the within and between household effect on child nutritional indicators. Results show that household size is negatively associated with child nutrition and high birth order children (i.e. later born) are more likely to fare worse than their lower birth order counterparts. From a variation decomposition analysis, results show higher variation in child nutrition across than within households. Specifically, for all the three indicators the between household explained

shares are in the range 5-8%, while for within household explained shares are less than 1%. Of the total variation, a large share is accounted for by the unexplained component.

The third essay (chapter 4) answers the question on the impact of objective risk (HIV prevalence rate) and perceived risk (concern about the likelihood of being infected by HIV due to antiretroviral therapy) on sexual behaviour in Botswana through the use of instrumental variables. Through the ordered probit model, we address a concern within the existing literature which excludes the sexually inactive. Overall, results suggest people are more responsive to perceived (subjective) risk than to objective risk. This result is robust to male/female and rural/urban subsamples. Those who updated their level of concern to being more concerned about the likelihood of HIV infection due to antiretroviral therapy are more likely to use condoms. However, those who became less concerned are less likely to abstain and use condoms, implying a possible externality associated with antiretroviral therapy. Further, results show a positive correlation between the unobservable factors affecting being sexually active and condom use, justifying the need to include the inactive as we have done.

5.2 Implications

Overall, results suggest that promotional/educational efforts should target the group that makes the least use of prenatal care to improve infant health outcomes. Combined with the provision of quality care as intended through the country`s essential health service package, this could also reduce maternal mortality. It would appear that care quality could be an issue given that utilization rates are high, but outcomes have been deteriorating. The policy objective to address inequality in care access should be coupled with efforts to address inequality in child health outcomes. While results suggest inequality in child health outcomes to be higher across than within households, depending on the distribution of children at a household level, the within variation should not be underestimated especially that a large share of the total variation in child nutrition is accounted for by the unexplained component. With regards to HIV/AIDS, results imply the need for a robust and continued emphasis that antiretroviral therapy does (and should) not substitute for protective behaviour. The above broad implications, along the sustainable development goal of “healthy lives for all”, could aid the realisation of the main health policy`s objective of “healthier Botswana”. Implications specific for each essay are outlined in the following.

The essential health service package has the objective to "ensure access to quality antenatal care, and quality care during and after delivery to mothers and their babies" (Republic of Botswana, 2010b:13). However, the finding from the first essay that adequate prenatal care utilization is positively associated with birth weight implies that efforts should go beyond *access* and promote adequate *use* of care among those that do not fully use it as this can improve infant health outcomes. Such efforts should be coupled with the need to reflect on the quality of care. Previous reports by the Ministry of Health and Wellness, focusing on another outcome of prenatal care use (maternal mortality) highlight that quality could be an issue. For instance, the review for the 2012-2014 period observes that of the 216 maternal deaths, sub-standard care indicators (delayed intervention, delayed referral, and inappropriate management) were present in 89% of the cases, representing a 10-percentage point increase from the 2007-11 review. This presents a threat to improved health. However, as we outlined in the introductory chapter, and as our results have shown, it should be noted that we do not suggest that deteriorating health outcomes is solely attributed to formal health care.

The absence of selection bias evident in our results implies that the assumption in Botswana (Okurut et al., 2013) that the absence of infant birth weight information is due to home births may not always hold. The differences across the sub samples also underscore the need to consider them separately, as this should further help address the policy issue of inequality in care access. Moreover, this also adds to those studies that attribute mixed evidence to the treatment of health outcome distribution: our results show that geographical locations could also matter both in terms of care use and health outcome.

From the second essay, the birth order effects may not be interpreted to mean that low birth order children are more preferred than high birth order children. It could mean that later born are disadvantaged by entering the 'household' late as they add to the household size. Nonetheless, as households with children receive a food ration every month, it could reflect the effect of other intrahousehold allocations that we could not explore due to data limitations (such as the extent to which children benefit from such rations in the presence of other members in the household). While results further show higher variation in child nutrition across than within households, they could be sensitive to the number of children in a certain household; therefore, the within variation should not be overlooked. The negative effect of household size on child nutritional indicators could suggest that there are other unobserved intrahousehold issues that could be playing a role. While this thesis does not evaluate the food ration programme, results

imply that its effectiveness could be threatened by such unobserved factors. This follows from our results that higher variances are unexplained by the observed characteristics. Therefore, an essential health service package objective "to enable each child to reach his/her maximum potential within the resources available and to enable as many children as possible to reach adulthood with their potential uncompromised by illness, disability, environmental hazard and unhealthy lifestyle" (Republic of Botswana, 2010a:13), may not be realised if child health outcomes are unequal.

The country, through the revised 2012 HIV policy, aims at preventing the spread of HIV infection and reducing its socio-economic impact, which was to be achieved through access to prevention information (Republic of Botswana, 2012). We presented evidence that the majority of the sample were knowledgeable about HIV/AIDS. However, the finding from the third essay that sexual behaviour is not responsive to HIV prevalence could be a cause for concern. If there are unobserved fatal behaviour, the effectiveness of public policy could be threatened by such non-responsiveness. As programmes such as antiretroviral therapy prolong the lives of the infected, there could be negative externalities. The results suggest this is especially the case for those who become less concerned about the likelihood of HIV as they reduce the likelihood of both protected sexual intercourse and abstinence. While we also establish that the effect of the HIV rate on sexual behaviour could be indirect by changing perception, if the effect of HIV on perception is outweighed by the effect of concern related to antiretroviral therapy on sexual behaviour, it might be the case that negative externalities could be evident. Hence efforts should address these unintended consequences.

5.3 Further Research

It is important to acknowledge that the surveys do not collect all the information that would ideally be useful for exploring these topics. Firstly, the Botswana Family Health Survey does not collect information on other measures of prenatal care, such as activities performed by health professionals during visits and time of care initiation. As a result, we were only able to provide evidence based on care visits. However, it may not be assumed that all women who fully utilize care receive the same extent or quality of care, and some will make more effective use of care and advice than others. What is provided may partially be determined by other health care supply factors over which they have no control. Future work would require more detailed surveys to capture these additional measures. Such surveys would help ascertain the nature and extent of inadequacies in the system (if any).

Another concern with both the BFHS and the BCWIS is the treatment of fathers. While the 2007/08 BFHS covered males for the first time, we could not establish whether the males were fathers to the infants. Hence, we were unable to assess their role on prenatal care use (Mwabu, 2008), a factor that would have provided a monitoring indicator against the objective stated in the essential health service package for Botswana (Republic of Botswana, 2010b). The BCWIS only has fathers` information where the child resides with the father. Partly because of low marriage rates in the country, the majority of fathers do not reside with their children. While we observe whether such absent fathers contribute to their children's' welfare even when not residing with them, it would have been of interest if their other characteristics were observed.

For children introduced to solid food, other indicators such as food diversity scores could also be considered to establish the link between monthly food rations and child nutrition. While it may not be easy to exclude other household members from benefiting from food ration, the aim would be to gauge the extent to which that could compromise child health.

Lastly, in the BAIS, we do not observe whether individuals know their status from the test conducted during the survey. The 2013 wave reports that it was initially planned for the results to be shared with those interested in knowing them, but this was not achieved because of the delays. As a result, we could not assess responsiveness to an individual`s HIV status. Perhaps this also calls for a shift in the way survey information is collected. Future work could consider experimental methods or longitudinal surveys as is the case in other sub-Saharan African countries. These methods should be able to facilitate controlling for self-selection into knowing one`s status. Another potential area to be further explored is the effect of antiretroviral therapy controlling for the possible selection into both HIV testing and enrolment in antiretroviral therapy.

One common trait for data used in this thesis is the fact that it is not panel. In the absence of such, some considerations are worth mentioning for subsequent data collection. For instance, in the case of AIDS impact surveys, a question relating to the knowledge of someone (living or dead) with HIV/ADS could be supplemented by with information on the relationship between these individuals. This could facilitate analysis of an indirect elasticity to the risk, by taking advantage of the fact that people do not choose their relatives (Francis, 2008). It is also important to ensure consistency in the questions asked to enable the consideration of all available surveys in program evaluation.

BIBLIOGRAPHY

Abrevaya, J., and Dahl, C. M. (2008). The effects of birth inputs on birth weight: evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26(4), 379-397.

Agiya, N., and Letamo, G. (2011). Impact of male circumcision on risk compensation through the impediment of condom use in Botswana. *African Health Sciences*, 11(4), 550-559.

Aguero, J. M., and Bharadwaj, P. (2014). Do the more educated know more about health? Evidence from schooling and HIV knowledge in Zimbabwe. *Economic Development and Cultural Change*, 62(3), 489-517.

Ahituv, A., Hotz, V. J., and Philipson, T. (1996). The responsiveness of the demand for condoms to the local prevalence of AIDS. *The Journal of Human Resources*, 31(4), 869-897.

Ajao, K. O., Ojofeitimi, E. O., Adebayo, A. A., Fatusi, A. O., and Afolabi, O. T. (2010). Influence of family size, household food security status, and child care practices on nutritional status of under-five children in Ile-Ife, Nigeria. *African Journal of Reproductive Health*, 14(4), 123-132.

Akwara, P. A., Madise, N. J., and Hinde, A. (2003). Perception of risk of HIV/AIDS and sexual behaviour in Kenya. *Journal of Biological Science*, 35, 385-411.

Alderman, H., and Behrman, J. R. (2006). Reducing the incidence of Low Birth Weight in Low-Income Countries has substantial economic benefits. *The World Bank Research Observer*, 21(1), 25-48.

Alexander, B. T., Dasinger, J. H., and Intapad, S. (2014). Effect of low birth weight on women's health. *Clinical Therapeutics*, 36(12), 1913-1923.

Alexander, G. R., and Korenbrot, C. C. (1995). The role of prenatal care in preventing low birth weight. *The future of children*, 5(1), 103-120.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, 19(4), 1483-1536.

Angrist, J. D., and Pischke, J-S. (2009). *Mostly Harmless Econometrics: An Empiricist`s Companion*. Princeton, New Jersey. Princeton University Press, 41 William Street.

Ansah, E. K., and Powell-Jackson, T. (2013). Can we trust measures of health care utilization from household surveys? *BMC Public Health*, 13,853. doi: 10.1186/1471-2458-13-853

Auld, M. C. (2006). Estimating behavioural response to the AIDS epidemic. *Contributions to economic analysis & policy*, 5(1).

Auld, M. C (2003). Choices, beliefs and infectious disease dynamics. *Journal of Health Economics*, 22, 361-377.

Awiti, J. O. (2014). A multilevel analysis of prenatal care and birth weight in Kenya, *Health Economics Review*, 4(33).

Baird, S., Gong, E., McIntosh, C., and Ozler, B. (2014). The heterogeneous effects of HIV testing. *Journal of Health Economics*, 37, 98-112.

Balekang, G. B., and Dintwa, K. F. (2016). A comparison of risky sexual behaviour between circumcised and uncircumcised men aged 30-34 years in Botswana. *African Health Science*, 16(1), 105-115.

Baratedi, W. M. (2013). *Lived experiences of HIV sero-discordant couples in Botswana*. PhD Thesis. University of South Africa.

Barcellos, S. H., Carvalho, L. S., and Lleras-Muney, A. (2014). Child gender and parental investment in India: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, 6(1), 157-189.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75 (299), 493-517.

Becker, G. S. (1960). 'An economic analysis of fertility' In a conference of the Universities-National Bureau Committee for Economic Research (Eds), *Demographic and Economic Change in Developed Countries*: 209-240. Princeton: Princeton University Press.

Becker, G. S., and Lewis, G. H. (1973). On the interaction between quantity and quality of children. *Journal of Political Economy*, 81(2), S279-S288.

Becker, G. S., and Tomes, N. (1976). Child endowments and the quantity and quality of children, *Journal of political economy*, 84(4), S143-S162.

Becker, M. H., and Joseph, J. G. (1988). AIDS and behavioural change to reduce risk: a review. *American Journal of Public Health*, 78, 394-410.

Behrman, J. R. (1997). 'Intrahousehold distribution and the family' In R. Rosenzweig and O. Stark (eds), *Handbook of Population and Family Economics*, Vol 1A, Amsterdam: North Holland.

Behrman, J. R., and Deolalikar, A. B. (1988). 'Health and Nutrition', In H. Chenery, and T.N. Srinivasan (eds), *The Handbook of Development Economics*, Elsevier Science Publishers, Vol I, 633-690.

Bell, A., and Jones, K. (2015). Explaining fixed effects: random effects modelling of time series cross sectional and panel data. *Political Science Research and Methods*, 3(1), 133-153.

Bingenheimer, J. B. (2010). Men`s multiple sexual partnership in 15 Sub-Saharan African countries: Sociodemographic patterns and implications. *Studies in Family Planning*, 41, 1-17.

Birungi, H., Nyarko, P., Armar-Klemesu, M., Arhinful., and Deganus, S. (2006). Acceptability and feasibility of introducing the WHO focused antennal care package in Ghana. Frontiers final report. Washington, DC: Population Council

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the cradle to the labour market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics,* 122(1), 409-439.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children`s education. *Quarterly Journal of Economics*, 669-700.

Blake, J. (1981). Family size and the quality of children. *Demography*, 18 (4), 421-442.

Bloom, D., and Canning, D. (2001). 'Cumulative Causality, Economic Growth, and the Demographic Transition' in N. Birdsall, C. Kelley, and S. Sinding. (eds), *Population Matters: Demographic Change, Economic Growth and Poverty in the Developing World*. Oxford University Press.

Bongaarts, J. (2001). Household size and composition in the developing world in the 1990s. *Population Studies*, 55(3), 263-279.

Bongaarts, J. (1982). The fertility-inhibiting effects of the intermediate fertility variables. *Studies in Family Planning*, 13(6/7), 179-189.

Bonita, R., Beaglehole, R., Kjellstrom, T. (2006). *Basic Epidemiology*. 2nd edition. World Health Organization.

Boozer, M. A., and Philipson, T. J. (2000). The impact of public testing for Human Immunodeficiency Virus. *The Journal of Human Resources*, 35(3), 419-446.

Botswana Institute for Development Policy Analysis (BIDPA). (2010). *Review of the National Strategy for Poverty Reduction (NSPR): Social Protection for Poverty and Vulnerability Reduction*. Unpublished consultancy report.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between instruments and endogenous explanatory variable is weak. *Journal of American Statistical Association*, 90(430), 443-450.

Burch, T. K. (1979). Household and family demography: A bibliography essay. *Population index*, 45(2), 173-195.

Bras, H., Kok, J., and Mandemakers, K. (2010). Sibship size and status attainment across context: Evidence from the Netherlands, 1840-1925. *Demographic Research*, 23(4), 73-104.

Brenϕe, A. A., and Molitor, R. (2018). Birth order and health of New-borns: What can we learn from the Danish Registry Data? *Journal of Population Economics*, 31, 363-395- DOI 10.1007/s00148-017-0660-1.

Brockerhoff, M., and Biddlecom, A. E. (1999). Migration, sexual behaviour and the risk of HIV in Kenya. *The International Migration Review*, 33(4), 833-856.

Cameron, A. C., and Trivedi, P. (2010). *Microeconometrics using Stata*. College Station, TX: Stata Press.

Carlson, K. (2015). Fear itself: The effects of distressing economic news on birth outcomes. *Journal of Health Economics*, 41, 117-132.

Carroli, G., Villar, J., Piaggio, G., Khan-Neelofur, D., Gulmezoglu, M., Mugford, M., Lumbiganon, P., Farnot, U., and Bersgjø, P. (2001). WHO systematic review of randomised controlled trials of routine antennal care. *The Lancet*, 357,1565-1570.

Celik, Y., and Younis, M. Z. (2007). Effects of antenatal care services on birth weight: Importance of model specification and empirical procedure used in estimating the marginal productivity of health inputs. *Journal of Medical System*, 31, 197-204.

Central Statistics Office (CSO). (2009a). *2007 Botswana Family Health Survey IV*. Gaborone, Botswana: CSO.

Central Statistics Office (CSO). (2009b). *Botswana Demographic Survey, 2006*. Gaborone, Botswana: Central Statistics Office.

Central Statistics Office (CSO). (2002). *2001 Botswana AIDS Impact Survey*. Gaborone, Botswana: Central Statistics Office.

Central Statistics Office (CSO). (2001). *Botswana Multiple Indicator Survey, 2000.* Gaborone, Botswana: Central Statistics Office

Central Statistics Office (CSO). (1999). *Botswana Family Health Survey III, 1996.* Gaborone, Botswana: Central Statistics Office

Central Statistics Office (CSO). (1989). *Botswana Family Health Survey II, 1988*. Gaborone, Botswana: Government Printer.

Central Statistics Office (CSO) and National AIDS Coordinating Agency (NACA). (2009). *2008 Botswana AIDS Impact Survey III, statistical report*. Gaborone, Botswana: Central Statistics Office.

Cerulli, G. (2014). Ivtreatreg: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection. *The Stata Journal*, 14(3), 453-480.

Charmarbagwala, R., Ranger, M., Waddington, H., and White, H. (2004). The determinants of child health and nutrition: A meta-analysis. The World Bank. Independent Evaluation Group Working Paper.

Chege, J. (2005). Feasibility of introducing a comprehensive integrated package of antenatal care services in rural public clinics in South Africa. Washington DC: Population council, Frontiers in reproduction Health.

Chiolero, A., Bovet, P., and Paccaud, F. (2005). Association between maternal smoking and low birth weight in Switzerland: the EDEN study. *Swiss Medical Weekly*, 135,525-530.

Clement, S., Candy, B., Sikorski, J., Wilson, J., and Smeeton, N. (1999). Does reducing the frequency of routine antenatal visits have long term effects? Follow up of participants in a randomized control trial. *British Journal of Obstetrics and Gynecology*, 106(4), 367-370.

Conway, K. S., and Deb, P. (2005). Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *Journal of Health Economics*, 24, 489-513.

Corno, L., and Paula, A. (2019). Risky sexual behaviours: Biological markers and self-reported data. *Economica*, 86, 229-261.

Corman, H., Dave, D., and Reichman, N. E. (2019). The effects of prenatal care on birth outcomes: reconciling a messy literature. *Oxford Research Encyclopedia Economics and Finance*

Crepaz, N., Hart, T. A., and Marks, G. (2004). Highly Active Antiretroviral Therapy and Sexual Risk Behaviour. A Meta-analytic review. *JAMA*, 292(2), 224-236.

Crown, W. H., Henk, H. J., and Vanness, D. J. (2011). Some cautions on the use of instrumental variables estimators in outcomes research: How bias in instrumental variables estimators is affected by instrument strength, instrument contamination and sample size. *Value in health*, 14, 1078-1084.

Culyer, A. J. (1971). The nature of the commodity 'health care' and its efficient allocation. *Oxford Economic Papers*, 23(2), 189-211.

Currie, J., and Reagan, P. B. (2003). Distance to hospital and children`s use of preventive care: Is being closer better and for whom? *Economic Inquiry*, 1, 378-391.

Dasgupta, K., and Solomon, K. T. (2018). Family size effects on childhood obesity: Evidence on the quantity-quality trade-off using the NLSY. *Economics and Human Biology*, 29, 42-55.

Davis, K. R., and Weller, S. C. (1999). The effectiveness of condoms in reducing heterosexual transmission of HIV. *Family Planning Perspectives*, 31(6), 272-279.

Daysal, N. M., Trandafir, M., and van Ewijk, R. (2015). Saving lives at births: The impact of home births on infant outcomes. *American Economic Journal: Applied Economics*, 7(3), 28-50.

Delavande, A., and Kohler, H-P. (2012). The impact of HIV testing on subjective expectations and risky behaviour in Malawi. *Demography*, 49(3), 1011-1036.

De Haan, M. (2010). Birth order, family size and educational attainment. *Economics of Education Review*, 29, 576-588.

De Haan, M., Plug, E., and Rosero, J. (2014). Birth order and Human Capital Development: Evidence from Ecuador. *The Journal of Human Resources*. 49(2), 359-392.

De Keyser, L., and Rossem, R. V. (2017). Birth order and height development in early childhood: empirical evidence for twenty-first-century Flanders. *The History of the Family*, 22(2-3), 236-257.

De Neve, J-W., Fink, G., Subramanian, S. V., Moyo, S., and Bor, J. (2015). Length of secondary schooling and risk of HIV infection in Botswana: Evidence from a natural experiment. *Lancet Global Health*, 3(8), e470-77.

De Walque, D. (2009). Does education affect HIV status? Evidence from five African countries. *The World Bank Economic Review*, 23(2), 209-233.

De Walque, D., Kazianga, H., and Over, M. (2012). Antiretroviral therapy perceived efficacy and risky sexual behaviour: Evidence from Mozambique. *Economic Development and Cultural Change*, 61(1), 97-126.

Deb, P., and Sosa-Rubi, S. G. (2005). Does onset or quality of prenatal care matter more for infant health? *HEDG Working paper* 05/11. University of York.

Dehejia, R., and Lleras-Muney, A. (2004). Booms, busts, and babies` health. *Quarterly Journal of Economics*, 119, 1091-1130.

Desai, S. (1995). When are children from large families disadvantaged? Evidence from cross-national analyses. *Population Studies*, 49(2), 195-210.

Deshmukh, J. S., Motghare, D. D., Zodpey, S. P., and Wadhva, S. K. (1998). Low birth weight and associated maternal Factors in an urban area. *Indian pediatrics*, 35, 33-36.

Dičkute, J., Padaiga, Ž., Grabauskas, V., Nadišauskienė, R. J., Basys, V., and Gaižauskienė, A. (2004). Maternal socio-economic factors and the risk of low birth weight in Lithuania. *Medicina* (Kaunas), 40(5), 475-482.

Dow, W. H., Philipson, T. J., and Sala-i-Martin, X. (1999). Longevity complementarities under competing risks. *American Economic Review*, 89(5), 1358-1371.

Dowswell, T., Carroli, G., Duley, L., Gates, S., Gulmezoglu, A. M., Khan-Neelofur, D., and Piaggio, G.G.P (2010). Alternative versus standard packages of antenatal care for low-risk pregnancy. *Cochrane Database systematic review*.

Dryden-Peterson, S., Jayeoba, O., Hughes, M. D, Jibril, H., Keapoletswe, K., Tlale, J., Modise, T. A., Asmelash, A., Moyo, S., Widenfelt, E. Van., Makhema, J., Shapiro, R.L., and Lockman, S. (2011). Highly Active Antiretroviral Therapy versus Zidovudine for Prevention of Mother-to Child Transmission in a programmatic setting. *Journal of Acquired Immune Deficiency Syndrome*. November 1, 58(3), 353–357.

Duncan, C., Jones, K and Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine*, 46(1), 97-117.

Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics*, 31(1), 1-34.

Dzakpasu, S., Powell-Jackson, T., and Campbell, O. M. R. (2014). Impact of user fees on maternal health service utilization and related outcomes: a systematic review. *Health policy and planning*, 29, 137-150.

Evans, W. N., and Lien, D. S. (2005). The benefits of prenatal care: evidence from the PAT bus strike. *Journal of Econometrics*, 125, 207–239.

Fako, T.T., Forcheh, N., and Ncube, E. (2004). Prospects of safe motherhood in Botswana: midwifery training and nurses` ability to complete the Botswana obstetric record. *Social Science and Medicine*, 58, 1109-1120.

Ferrer, R., and Klein, W. M. (2015). Risk perception and health behaviour. *Curr Opin Psychology*, 1(5), 85-89.

Finlay, J. (2007). The role of health in economic development. PGDA, Working paper No 21. http://www.hsph.havard.edu/pgda/working.htm

Fiscella, K. (1995). Does prenatal care improve birth outcomes? A critical review. *Obstetrics and Gynecology*, 85, 468-79.

Fortson, J. G. (2008). The gradient in Sub-Saharan Africa: Socio Economic status and HIV/AIDS. *Demography*, 45(2), 303-322.

Francis, A. M. (2008). The economics of sexuality: The effects of HIV/AIDS on homosexual behaviour in the United States. *Journal of Health Economics*, 27(3), 675-689.

Frick, K. D., and Lantz, P. M. (1996). Selection Bias in Prenatal Care Utilization: An Interdisciplinary Framework and Review of Literature. *Medical Care Research and Review*, 53(4), 371-396.

Friedman, W. H. (2018). Antiretroviral drug access and behaviour change. *Journal of Development Economics*, 135, 392-411.

Gajate-Garrido, G. (2013). The impact of adequate prenatal care on urban birth outcomes: an analysis in a developing country context. *Economic Development and Cultural Change*, 62(1), 95-130.

Geoffard, P-Y., and Philipson, T. (1996). Rational epidemics and their public control. *International Economic Review*, 37(3), 603-624.

Gerber, T. P., and Berman, D. (2008). Heterogenous condom use in contemporary Russia. *Studies in Family Planning*, 39(1), 1-17.

Gersovitz, M., and Hammer, J. S. (2003). Infectious diseases, public policy, and the marriage of economics and epidemiology. *The World Bank Research Observer*, 18(2), 129-157.

Glewwe, P. (1999). Why does Mother`s Schooling Raise Child Health in Developing Countries? Evidence from Morocco. *The Journal of Human Resources*, 34(1), 124-159.

Glick, P. J., and Sahn, D. E. (2008). Are Africans practicing safe sex? Evidence from Demographic and Health Surveys for Eight countries. *Economic Development and Cultural Change*, 56(2), 397-439.

Gobotswang, K. (1993). Determinants of the nutritional status of children in a rural African setting: the case of Chobe district, Botswana. *Food and Nutrition Bulletin*, 19(1), 42-45.

Godlonton, S., Munthali, A., and Thornton, R. (2016). Responding to risk: circumcision, information and HIV prevention. *The Review of Economics and Statistics*, 98(2), 333-349.

Godlonton, S., and Thornton, R. L. (2013). Learning from others` HIV testing: updating beliefs and responding to risk. *American Economic Review*: Papers & Proceedings, 103(3), 439-444.

Gong, E. (2015). HIV testing and risky sexual behaviour. *The Economic Journal*, 125, 32-60.

Grace, K., Davenport, F., Hanson, H., Funk, C., and Shukla, S. (2015). Linking climate change and health outcomes: Examining the relationship between temperature, precipitation and birth weight in Africa. *Global Environmental Change*, 35, 125-137.

Grace, K., Brown, M., and McNally, A. (2014). Examining the link between food prices and food insecurity: A multi-level analysis of maize price and birth weight in Kenya. *Food policy*, (46), 56-65.

Gray, P. B. (2004). HIV and Islam: Is HIV prevalence lower among Muslims. *Social Science & Medicine*, 58, 1751-1756.

Greene, W. M. (2008). *Econometric Analysis, sixth edition*. Upper Saddle River, New Jersey: Prentice Hall.

Gregson, S., Zhuwau, T., Anderson, R. M., and Chandiwana, S. K. (1998). Is there evidence for behavioural change in response to AIDS in rural Zimbabwe*? Social Science & Medicine*, 46(3), 321-330.

Gronau, R. (1977). Leisure, home production and work-the theory of allocation of time revisited. *Journal of Political Economy*, 85(6), 1099-1123.

Grossman, M. (1972). On the concept of Health Capital and the Demand for Health. *The Journal of Political Economy*, 80(2), 223-255.

Grossman, M., and Joyce, T. J. (1990). Unobservables, pregnancy resolution and birth weight production function in New York city. *Journal of Political Economy*, 98(5), 983-1007.

Guilkey, D. K., Popkin, B. M., Akin, J. S., and Wong, E. L. (1989). Prenatal care and pregnancy outcome in Cebu, Philippines. *Journal of Development Economics*, 30, 241–272.

Guillon, M., and Thuilliez, J. (2015). HIV and rational risky behaviour: a systematic review of published empirical literature (1990-2013). Working paper of the Sorbonne economics centre-ISSN: 1955-611X

Gujarati, D. N. (2004). *Basic Econometrics*. McGraw Hill.

Habibov, N. N., and Fan, L. (2011). Does prenatal health care improve child birthweight outcomes in Azerbaijan? Results of the national demographic and health survey. *Economics and Human Biology*, 9(1), 56-65.

Habibov, N., Fan, L., Campbell, D., and Cheung, C. Ho. (2017). Effect of prenatal care frequency, timing and quality on child birth weight in Tajikistan. *World Medical and Health Policy*, 9(1), 89-102.

Hamoudi, A., and Thomas, D. (2014). Endogenous coresidence and program incidence: South Africa`s Old Age pension. *Journal of Development Economics*, 109, 30-37.

Hart, G., and Williamson, L. (2005). Increase in HIV sexual risk behaviour in homosexual men in Scotland, 1996-2002: prevention failure? *Sexually Transmitted Infections*, 81(5), 367-372.

Hartman, R. S. (1991). A Monte Carlo analysis of alternative estimators in models involving selectivity. *Journal of Business and Economic Statistics*, 9(1), 41-49.

Hauck, K. (2018). The economics of infectious diseases. *Oxford Research Encyclopedia of Economics and Finance*. DOI: 10.1093/acrefore/9780190625979.013.251

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.

Hoddinott, J., and Mekasha, T. J. (2017). Social protection, household size and its determinants: Evidence from Ethiopia, Ethiopian Development Research Institute Working Paper 107.

Horton, S. (1988). Birth order and child nutritional status: evidence from the Philippines. *Economic Development and Cultural Change*, 36(2), 341-354.

Horton, S. (1986). Child nutrition and family size in the Philippines. *Journal of Development Economics*, 23, 161-176.

Howell, E. M., Holla, N., and Waidmann, T. (2016). Being the younger child in a large African Family: a study of birth order as a risk factor for poor health using demographic and health surveys for 18 countries. *BioMed Central*, 2(61).

Jayachandran, S., and Pande, R. (2017). Why are Indian children so short? The role of birth order and son preference. *American Economic Review*, 107(9), 2600-2629.

Jayachandran, S., and Kuziemko, I. (2011). Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. *Quarterly Journal of Economics*, 126(3), 1485-1538.

Jewell, R. T. (2007). Prenatal care and birthweight production: evidence from South America. *Applied Economics*, 39(4), 415-426.

Jewell, R. T., and Triunfo, P. (2006). The impact of prenatal care on birthweight: the case of Uruguay. *Health Economics*, 15, 1245-1250.

Joshi, H. S., Subba, S. H., Dabral, S. B., Dwivedi, S., Kumar, D., and Singh, S. (2005). Risk factors associated with low birth weight in newborns. *Indian Journal of Community Medicine*, 30(4).

Joyce, T. (1994). Self-selection, prenatal care and birth weight among Blacks, Whites and Hispanics in New York City. *The Journal of Human Resources*, 29(3), 762-794.

Kabagenyi, A., and Rutaremwa, G. (2013). The effect of household characteristics on child mortality in Uganda. *American Journal of Sociological Research*, 3(1), 1-5.

Kabubo-Mariara, J., Ndenge, G. K., and Mwabu, D. K. (2008). Determinants of children`s nutritional status in Kenya. Evidence from Demographic and Health Survey. *Journal of African Economies*, 18(3), 363-387.

Katisi, M., and Daniel, M. (2015). Safe male circumcision in Botswana: Tension between traditional practices and biomedical marketing. *Global public health*, 10(5-6), 739-756.

Kebede, B. (2005). Genetic endowments, parental and child health in rural Ethiopia. *Scottish Journal of Political Economy*, 52(2), 194-221.

Keetile, M. (2014). High risk behaviours among adult men and women in Botswana: Implications for HIV prevention efforts. *Journal of Social Aspects of HIV/AIDS*, 11(1), 158-166.

Keetile, M., and Kgosidintsi, G. (2018). Sexual behaviours of HIV positive adults receiving HAART in Botswana: a cross sectional study. *African Health Sciences*, 18(3), 503-511.

Khan, M. W., Arbab, M., Murad, M., Khan, M. B., and Abdullah, S. (2014). Study of factors affecting and causing low birth weight. *Journal of scientific research*, 6(2), 387-394.

Kramer, M. S. (1987). Determinants of low birth weight: methodological assessment and meta-analysis. *Bulletin of the World Health Organization*, 65(5), 663-737.

Kremer, M. (1996). Integrating behavioural choice into epidemiological models of AIDS. *Quarterly Journal of Economics*, 111, 559-573.

Lakdawalla, D., Sood, N., and Goldman, D. (2006). HIV breakthroughs and risky sexual behaviour. *Quarterly Journal of Economics*, 121(3), 1063-1102.

Lawson, D., and Mace, R. (2008). Sibling configuration and childhood growth in contemporary British families. *International Journal of Epidemiology*, 27, 1408-1421.

Lawson, D., and Appleton, S. (2007). Child health in Uganda: Policy determinants and measurement. *The European Journal of Development Research*, 19(2), 210-233.

Laxminarayan, R and Malani, A. (2011). “Economics of infectious diseases”. In S. Glied and P. Smith (Eds), *The Oxford Handbook of Health Economics*.1:189-205. Oxford: Oxford University Press.

Letamo, G., Keetile, M., and Navaneetham, K. (2017). The impact of HIV antiretroviral treatment perception on risky sexual behaviour in Botswana: a short report. *AIDS Care*, 29(12), 1589-1593.

Letamo, G., and Majelantle, R. G. (2001). Factors influencing low birth weight and prematurity in Botswana. *Journal of biological science*, 33(3), 391-403.

Letamo, G., and Mokgatlhe, L. (2013). Predictors of risky sexual behaviour among young people in the era of HIV/AIDS: Evidence from the 2008 Botswana AIDS Impact Survey III. *African Journal of Reproductive Health*, 17(3), 169-181.

Lewbel, A., Dong, Y., and Yang, T. T. (2012). Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics*, 45(3), 809-829.

Lewit, E. M. (1977). *Experience with pregnancy, the demand for prenatal care and the production surviving infants*, PhD Thesis, Graduate Centre: City University of New York.

Lin, S-J. (2004). Assessing the effect of prenatal care on birth outcome in Taiwan. *The Social Science Journal*, 41, 493-501.

Lindo, J. M. (2011). Parental Job Loss and Infant health. *Journal of Health Economics*, 30, 869-879.

Liu, G. G. (1998). Birth outcomes and the effectiveness of prenatal care. *Health Services Research*, 32(6), 805–823.

Løken, K. V., Mogstad, M., and Wiswall, M. (2012). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics*, 4(2), 1-35.

Lu, Y., and Treiman, D. J. (2008). The effect of Sibship size on educational attainment in China: Period variations. *American Sociological Review*, 73(5), 813-834.

Lucas, A. M., and Wilson, N. L. (2018). Schooling, wealth, risky sexual behaviour and HIV/AIDS in Sub-Saharan Africa. *The Journal of Development Studies*.

Ludbrook, A., and Cohen, D. (2003). ‘Economics of health and health improvement’, in A. Scott., A. Maynard., and R. Elliot(eds). *Advances in health economics.* John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ: England.

Lundberg, E., and Svaleryd, H. (2016). Birth order and child health. Mimeo, Linnaeus University.

Lundborg, P., Ralsmark, H., and Rooth, D-O. (2015). The more the healthier? Health and family size, Mimeo, Lund University.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

Mahgoub, S. E. O., Nnyepi, M., and Bandeke, T. (2006). Factors affecting prevalence of malnutrition among children under three years of age in Botswana. *African Journal of Food Agriculture Nutrition and Development*, 6(1).

Majelantle, R. (2014). "Infant and Child Mortality Levels and Trends in Botswana". In Statistics Botswana (2014a), *Population and Housing Census analytical report*. Gaborone, Botswana: Statistics Botswana.

Majoko, F., Munjanja, S.P., Nystrom, L., Mason, E., Lindmark, G. (2007). Randomised controlled trial of two antenatal care models in rural Zimbabwe. *General Obstetrics,*114,802-811.

Malema, B. W. (2012). Determinants of Condom use in Botswana: An Empirical Investigation of the role of gender. *Botswana Journal of Economics*, 10(14), 59-78.

Mathule, M. S. L., Kennedy, T., Gates, G., and Spicer, M. T. (2005). Predictors of Birthweight in Healthy Women Attending a Rural Antenatal Clinic. *African Journal of Food Agriculture Nutrition and Development*, 5(1).

Maughan-Brown, B., and Venkataramani, A. S. (2018). Accuracy and determinants of perceived HIV risk among young women in South Africa. *BMC Public Health*, 18(42).

McDuffie, R. S., Beck, A., Bischoff, K., Cross, J., and Orleans, M. (1996). Effects of frequency of prenatal care visits on perinatal outcome among low risk women: a randomized controlled trial. *Journal of American Medical Association*, 275(11), 847-851.

Mchenga, M., Burger, R., and von Fintel, D. (2019). Examining the impact of WHO`s Focused Antenatal Care policy on early access, underutilisation and quality of antenatal care services in Malawi: a retrospective study. *BMC Health Services Research*, 19:295

Mechoulan, S. (2007). Risky sexual behaviour, testing and HIV treatments. *Forum for Health Economics & Policy*, 10(2).

Michael, R. T. (2004). Sexual capital: an extension of Grossman`s concept of health capital. *Journal of health economics*, 23, 643-652.

Millimet, D. L., and Wang, L. (2011). Is the quantity-quality trade-off a trade-off for all, none or some? *Economic Development and Cultural Change*, 60(1), 155-195.

Mills, S., Leburu, V., El-Halabi, S., Mokganya, L., and Chowdhury, S. (2010). *Fertility Decline in Botswana 1980-2006*: A case study. World Bank.

Ministry of Finance and Development Planning (MFDP). (2017). *National Development Plan 11. April 2017 to March 2023*. MFDP. Gaborone, Botswana.

Ministry of Finance and Development Planning (MFDP). (Various). *Financial Statements, Tables and Estimates of the Consolidated and Development Fund Revenues*. Gaborone, Botswana: Government Printer.

Ministry of Finance and Development Planning (MFDP) and United Nations Population Fund (UNFPA). (2010). Revised national population policy. National Council on Population and Development, Gaborone: MFDP.

Ministry of Health (MOH). (2010). *Annual Tuberculosis and Leprosy report 2006-2008.* Gaborone, Botswana: Ministry of Health.

Ministry of Health (MOH). (2011a). *2011 Botswana second generation HIV/AIDS Antenatal surveillance technical report*. Gaborone, Botswana: Ministry of Health.

Ministry of Health (MOH). (2011b). *2010-2011 Botswana Tuberculosis and Leprosy report.* Gaborone, Botswana: Ministry of Health.

Ministry of Health (MOH). (2012). 2012 *Botswana Tuberculosis and Leprosy report.* Gaborone, Botswana: Ministry of Health

Ministry of Health (MOH) (various). *Tuberculosis annual reports.* Gaborone, Botswana: Ministry of Health.

Ministry of Health (MOH). (nd). *5-year maternal mortality report (2007-2011). Exploring the causes of maternal mortalit*y. Ministry of Health.

Ministry of Health and Wellness (MOHW). (2017). http://www.moh.gov.bw/. Accessed on 3rd February 2017.

Ministry of Health and Wellness (MOHW). (nd). *3-year maternal mortality report (2012-2014). Exploring causes of maternal mortality*. Ministry of Health and Wellness.

Minnis, A. M., Steiner, M. J, Gallo, M. F. Warner, L., Hobbs, M., van der Straten, A., Chipato, T., Macaluso, M., and Padian, N. S. (2009). Biomarker validation of reports of recent sexual activity: results of a randomized controlled study in Zimbabwe. *American Journal of Epidemiology*, 170(7), 918-924.

Mmopelwa, D., and Seleka, T. B. (2012). Public transfers and communal cattle marketing decisions: Empirical evidence from Botswana. *Journal of Agribusiness*, 30(1), 87-106.

Mogobe, K. D., Tshiamo, W., and Bowelo, M. (2007). Monitoring Maternity Mortality in Botswana. *Reproductive Health Matters*, 15(30), 163-171.

Mogstad, M., and Wiswall, M. (2016). Testing the quantity-quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, 7, 157-192.

Mokomane, Z. (2005). Cohabitation in Botswana: An alternative or prelude to marriage. *African Population Studies*, 20(1), 19-37.

Morgan, S. L., and Winship, C. (2007). *Counterfactuals and casual inference: Methods and principles for social research.* New York: Cambridge University Press.

Morwaagole, D. (2013). Re a nyalana project: Yes to Marriage, No to cohabitation. Presented at SADC gender protocol summit, Gaborone, Botswana.

Mosweunyane, T. (2007). Update on 2006/07 Malaria transmission season; paper presented at the annual malaria conference at Phakalane Golf Estate; Gaborone, Botswana.

Moultrie, T., Dorrington R., Hill, A., Hill, K., Timaeus, I., and Zaba, B. (2013). *Tools for Demographic Estimation. Paris:* International Union for the Scientific Study of Population demographicestimation.iussp.org.

Mundlak, Y. (1978). On the pooling of time series and cross-section data. *Econometrica*, 46(1), 69-85.

Munjanja, S.P., Lindmark, G., Nystrom, L. (1996). Randomised controlled trial of a reduced-visits programme of antenatal care in Harare, Zimbabwe. *The Lancet*, 348,364-369.

Murray, C. J. L., Gakidou, E.E., and Frenk, J. (1999). Health inequalities and social group differences: what should we measure*? Bulletin of the World Health Organization* 77 (2), 537–543

Mushkin, S. J. (1962). Health as an investment. *Journal of political economy*, 70(5),129-157.

Mussa, R. (2015). Intrahousehold and interhousehold child nutrition inequality in Malawi. *South African Journal of Economics*, 83(1), 140-153.

Mwabu, G. (2008). The production of child health in Kenya: A structural Model of Birth Weight. *Journal of African Economies*, 18(2), 212-260.

National AIDS Coordinating Agency (NACA). (2015). Progress report of the national response to the 2011 declaration of commitments on HIV and AIDS.

National AIDS Coordinating Agency (NACA). (2003). *Botswana Second Generation HIV/AIDS Surveillance. National AIDS Coordinating Agency (NACA) in collaboration with other development partners*, Botswana.

National AIDS Coordinating Agency (NACA) and Central Statistics Office (CSO). (2005). *Botswana AIDS Impact Survey II, 2004*. Gaborone, Botswana: Central Statistics Office.

Nnyepi, M. S. (2007). Household factors are strong indicators of children`s nutritional status in children with access to primary health care in the greater Gaborone area. *Scientific Research and Essay*, 2(2), 055-061.

Okurut, F. N., Kepaletswe, F., and Mmopelwa, D. (2013). Determinants of Birth Weight in Botswana. *International Journal of Academic Research in Economics and Management Sciences*, 2 (52), 107-117.

Oster, E. (2012). HIV and sexual behaviour change: Why not Africa? *Journal of Health Economics*, 31, 35-49.

Oster, E. (2005). Sexually transmitted infections, sexual behaviour and the HIV/AIDS epidemic. *Quarterly Journal of Economics*, 467-515.

Palen, L., Smith, E., Caldwell, L., Flisher, A., Wegner, L., and Vergnani, T. (2008). Inconsistent reports of sexual intercourse among South African high school students. *Journal of Adolescent Health*, 42(3), 221-227.

Paula, A., Shapira, G., and Todd, P. E. (2014). How beliefs about HIV status affect risky behaviours: evidence from Malawi. *Journal of Applied Econometrics*, 29, 944-964.

Philipson, T. J., and Posner, R. A. (1995). A theoretical and empirical investigation of the effects of public health subsidies for STD testing. *Quarterly Journal of Economics*, 100(2), 445-474.

Philipson, T. J., and Posner, R. A. (1993). *Private choices and public health: The AIDS epidemic in an economic perspective*. Cambridge: Harvard University Press.

Picard, N., and Wolff, F-C. (2010). Measuring educational inequalities: a method and application to Albania. *Journal of Population Economics*, 23, 1117-1151. DOI 10.1007/s00148-008-020-z.

Presidential Task Team (2016). *Vision 2036: Achieving prosperity for all*. Lentswe La Lesedi (PTY) Ltd. Gaborone: Botswana.

Price, J. (2008). Parent-child quality time. Does birth order matter? *The Journal of Human resources*, 240-265.

Qian, D., Pong, R.W., Yin, A., Nagarajan, K.V., and Meng, Q. (2009). Determinants of health care demand in poor, rural China: the case of Gansu Province. *Health Policy and Planning*, 24, 324-334.

Rahman, M. (2016). Association between order of birth and chronic malnutrition of children: a study of nationally representative Bangladeshi sample. *Cadernos de saude pulica*, 32(2). doi:10.1590/0102-311X00011215 PMID:26958818.

Randall, S., and Coast, E. (2015). Poverty in African Households: The Limits of Survey and Census Representations. *The Journal of Development Studies*, 51(2), 162-177.

Regional Hunger and Vulnerability Programme (RHVP). (2011). Social protection in Botswana-a model for Africa? Frontiers of social protection brief, 9, 1-20.

Reichman, N. E., Corman, H., Noonan, K., and Dave, D. (2009). Infant health production functions: What a difference the data make. *Health Economics*, 18, 761-782.

Republic of Botswana. (2014). *Botswana Habitat III report*. Ministry of Lands and Housing. Gaborone, Botswana.

Republic of Botswana. (2012). *The revised Botswana national policy on HIV and AIDS.*

Republic of Botswana. (2011). *National Health Policy*. Gaborone, Botswana: Ministry of Health.

Republic of Botswana (2010a). *Integrated Health Service Plan: A strategy for changing the health sector for a healthy Botswana 2010-2020.* Gaborone, Botswana: Ministry of Health

Republic of Botswana (2010b). The Essential Health Service Package for Botswana. Gaborone, Botswana: Ministry of Health.

Republic of Botswana (2010c). *Revised National Youth Policy*. Ministry of youth, sports and culture.

Republic of Botswana. (2009). *Review of the Botswana Anti-Retroviral treatment programme.* Unpublished report.

Republic of Botswana and UN (2015). *Botswana Millennium Development Goal status report: sustaining progress to 2015 and beyond.*

Rosenzweig, M. R., and Schultz, T. P. (1988). The stability of household production technology: a replication. *The Journal of Human Resources*, 23(4), 535-549.

Rosenzweig, M. R., and Schultz, T. P. (1983). Estimating a household production function: heterogeneity, the demand for health input, and their effects on birth weight. *Journal of Political Economy*, 91, 723-746.

Rosenzweig, M. R., and Schultz, T. P. (1982). 'The behaviour of mothers as inputs to child health: The determinant of birth weight, gestation and the rate of fetal growth', in V. R. Fuchs, (ed), *Economic Aspects of Health*: The University of Chicago Press: 53-92.

Rous, J. J., Jewell, R. T., and Brown. R. W. (2003). The effect of prenatal care on birthweight: a full-information maximum likelihood approach. *Health Economics,* 13, 251–264.

Sabone, M., Magowe, M., Busang, L., Moalosi, J., Binagwa, B., and Mwambona, J. (2013). Impediments for the uptake of Botswana government`s male circumcision initiative for HIV prevention. *The Scientific World Journal*. http://dx.doi.org/10.1155/2013/387508

Sahn, D. E., and Stifel, D. C. (2002). Parental preferences for nutrition of boys and girls: Evidence from Africa. *Journal of Development Studies*, 39(1), 21-45.

Schaan, M. M., Taylor, M., and Marlink, R. (2014). Reproductive behaviour among women on antiretroviral therapy in Botswana: mismatched pregnancy plans and contraceptive use. *African Journal of AIDS Research*, 13(3), 305-311.

Schwartz, A. L., and Sommers, B. D. (2014). Moving for Medicaid? Recent eligibility expansions did not induce migration from other states. *Health Affairs*, 33(1),88-94.

Seleka, T. B., and Lekobane, K. R. (2017). Public transfers and participation decisions in Botswana`s subsistence economy. *Review of Development Economics*, 21(4), 1380-1400.

Seleka, T. B., Makepe, P. M., Kebakile, P., Batsetswe, L., Mmopelwa, D., Mbaiwa, K., and Jackson, J. (2008). *The feasibility of mandatory fortification of cereals in Botswana.* Lentswe La Lesedi (PTY) Ltd. Gaborone: Botswana.

Shavit, Y., and Pierce, J. L. (1991). Sibship size and educational attainment in nuclear and extended families: Arabs and Jews in Israel. *American Sociological Review*, 56(3), 321-330.

Sikorski, J., Wilson, J., Clement, S., Das, S., and Smeeton, N. (1996). A randomized controlled trial comparing two schedules of antenatal visits: the antenatal project. *British Medical Journal*, 312, 546-553.

Silles, M. A. (2012). Maternal Smoking, birth weight and infant health. *Research Memorandum* 89. University of Hull.

Simkhada, B., van Teijlingen, E. R., Porter, M., and Simkhada, P. (2007). Factors affecting the utilization of antenatal care in developing countries: a systematic review of literature. *Journal of Advanced Nursing*, 61(3), 244-260.

Smith, J. P. (2009). The impact of childhood health on adult labour market outcomes. *The Review of Economics and Statistics*, 91(3), 478-489.

Staiger, D., and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557-586.

StataCorp. (2017). Stata Release 15. Statistical Software, College Station TX: StataCorp LP.

Statistics Botswana. (2018). *Botswana Demographic Survey Report 2017*. Gaborone, Botswana: Statistics Botswana.

Statistics Botswana. (2017). *Health Statistics, stats brief 2007-2015*. Gaborone, Botswana: Statistics Botswana.

Statistics Botswana. (2014a). *2011 population and housing census analytical report*, Gaborone, Botswana: Statistics Botswana.

Statistics Botswana (2014b). *Vital Statistics Report 2011*. Gaborone, Botswana: Statistics Botswana.

Statistics Botswana. (2012). *Health Statistics 2009*, Gaborone, Botswana: Statistics Botswana.

Statistics Botswana and NACA. (2016). *Botswana AIDS Impact survey IV, statistical report*. Gaborone, Botswana: Statistics Botswana.

Statistics Botswana and UNICEF (2008). *Indicators of child wellbeing in Botswana between 2001 and 2008*. Gaborone, Botswana: Statistics Botswana.

Sterck, O. (2013). HIV/AIDS and fatalism: should prevention campaigns disclose the transmission rate of HIV? *Journal of African Economies*, 23(1), 53-104.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20(4), 518-529.

Straus, J., and Thomas, D. (1995). 'Human Resources: Empirical Modelling of Household and Family Decisions', in J. Behrman and T.N. Srinivasan(eds), *Handbook of Development Economics*, Vol. 3, Amsterdam: North Holland Press.

Strauss, J., and Thomas, D. (1998). Health, nutrition and economic development. *Journal of Economic Literature*, 36(2), 766-817.

Strauss, J., and Thomas, D. (2008). 'Health over life course', in T. P. Schultz and J. Strauss (eds), *Handbook of Development Economics*, Vol 4, Amsterdam: North Holland Press.

Tenkorang, E. Y., Maticka-Tyndale, E., and Rajulton, F. (2011). A multi-level analysis of risk perception, poverty and sexual risk-taking among young people in Cape Town, South Africa. *Health & Place*, 17(2), 525-535.

Tennekoon, V., and Roseman, R. (2014). Behold a virgin is with HIV! Misreporting sexual behaviour among the infected adolescent. *Health Economics*, 2, 345-358

Tharakan, T. C., and Suchindran, C. M. (1999). Determinants of malnutrition in Botswana: An intervention model for Botswana. *Nutrition Research*, 19(6), 843-860.

Thompson, B. (2009). Impact of the financial and economic crisis on nutrition: policy and programme responses. Nutrition and consumer protection division. FAO.

Thornton, R. L. (2008). The demand for, and impact of learning HIV status. *American Economic Review*, 98(5), 1829-1863.

Tobah, Y. S. B., LeBlanc, A., Branda, M. E., Inselman, J.W., Morris, M. A.,Ridgeway, J. L., Finnie, D. M., Theiler, R., Torbenson, V. E, Brodrick, E.M. de Mooij, M. M., Gostout, B., Famuyide, A. (2019). Randomized comparison of a reduced-visit prenatal care model enhanced with remote monitoring. *American Journal of Obstetrics & Gynecology*, 221:638.e1-9

Tough, S., Svenson, L., and Schopflocher, D. (1999). *Maternal Risk Factors in Relationship to Birth outcome*. Alberta Health and Wellness, Alberta, Canada.

United Nations (UN). (2008). *Principles and recommendations for population and housing censuses. Revision 2*. Statistical Paper Series M No 67/Rev 2. United Nations Department of Economic and Social Affairs, Statistical Division. Publication Sales No. E.07.XVII.8. New York: United Nations.

United Nations Development Programme (UNDP). (2000). *Botswana Human Development Report: towards an AIDS-free generation*. Gaborone, Botswana: Petadco Printing House.

University of California, San Francisco (UCSF). (2016). Eliminating malaria in Botswana: country briefing. www.shrinkingthemalariamap.org/eliminantion-progress/botswana

Vella, F. (1998). Estimating models with sample selection bias: a survey. *The Journal of Human Resource*s, 33(1), 127-169.

Victora, C.G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., and Sachdev, H. S. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *Lancet*, 371, 340-357.

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Belizán, J. M., Farnot, U., Al-Mazrou, Y., Carroli, G., Pinol, A., Donner, A., Langer, A., Nigenda, G., Mugford, M., Fox-Rushby, J., Hutton, G., Bergsjϕ, P., Bakketeig, L., and Berendes, H. (2001). WHO antennal care randomised trial for the evaluation of a new model of routine antenatal care. *Lancet*, 357, 1551-1564.

Vogel, J. P, Habib, N. A., Souza, J. P, Gulmezoglu, A. M., Dowswell, T., Carroli, G., Baaqeel, H. S., Lumbiganon, P., Piaggio, G. and Oladapo, O. T. (2013). Antenatal care packages with reduced visits and perinatal mortality: A secondary analysis of the WHO antennal care trial. *Reproductive Health*, 10, 19.

Wagstaff, A. (1986). The demand for health: theory and applications. *Journal of Epidemiology and community health*, 40, 1-11.

Warner, G. L. (1995). Prenatal care demand and birthweight production of black mothers. *The American Economic Review*, 85(2), 132–137.

Warner, G. (1998). Birthweight productivity of prenatal care. *Southern Economic Journal*, 65(1), 42-63.

Waters, E. A., McQueen, A., and Cameron, L. D. (2014). *Perceived risk and health risk communication*: The Routledge Handbook of Language and Health Communication. https://www.routledgehandbooks.com/legal/doi/10.4324/9781315856971.ch3

Wilson, N. L., Xiong, W., and Mattson, C. L. (2014). Is sex like driving? HIV prevention and risk compensation. *Journal of Development Economics*, 106,78-91.

Wojtczak, A. (2005). 'Health policies versus public policies', in S. William., A. Gunn., P.B Mansourian., A. M. Davies., A. Piel., and B. M. Sayers (eds). *Understanding the global dimensions of health*: The International Association for Humanitarian Medicine Brock Chisolm: USA.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2), 420-445.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. USA: Massachusetts Institute of Technology.

Wooldridge, J. M. (2007). What`s new in econometrics? Control functions and related methods.

World Bank. (2015). *Botswana Poverty Assessment*. The World Bank Group.

World Bank. (2001). *World Development Report 2000/2001: Attacking Poverty: World Development Report*, New York: Oxford University Press. © World Bank. https://openknowledge.worldbank.org/handle/10986/11856 License: CC BY 3.0 IGO.

World Health Organization (WHO). (2019): 2019 *World malaria report*. Geneva: World Health Organization.

World Health Organization (WHO). (2017). Tetanus vaccines: WHO position paper. *Weekly epidemiological record*, 92(6), 53-72.

World Health Organization (WHO). (2016). *WHO recommendations on antennal care for a positive pregnancy experience*. Geneva: World Health Organization.

World Health Organization (WHO). (2015a). *Health in 2015: From Millennium Development Goals to Sustainable Development Goals*. World Health Organization.

World Health Organization (WHO). (2015b). 2015 *World malaria report*. World Health Organization.

World Health Organization (WHO). (2011a). *World Health Statistics 2011*, Geneva: World Health Organization.

World Health Organization (WHO). (2011b). WHO Anthro for personal computers: Software for assessing growth and development of the world`s children. Available at (http://www.who.int/childgrowth/software/en/).

World Health Organization (WHO). (2006). Basic Documents, forty-fifth edition, Supplement. World Health Organization.

World Health Organization (WHO). (2004). ICD-10: International statistical classification of diseases and related problems: tenth revision, 2nd ed. World Health Organization. Available at https://apps.who.int/iris/handle/10665/42980

World Health Organization (WHO) (2002). WHO antenatal care randomized trial: manual for the implementation of the new model.

Yassin, K. M. (2000). Indices and sociodemographic determinants of childhood mortality in rural upper Egypt. *Social Science and Medicine*, 51, 185-197.

Zhong, H. (2017). The Effect of sibling size on children`s health and education: Is there a quantity-quality trade off? *The Journal of Developmental studies*, 53(8), 1194-1206.