UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

THE EFFECT OF PARENTAL EDUCATION ON CHILD AND ADULT HEALTH IN ZAMBIA: A REGRESSION DISCONTINUITY ANALYSIS

BY

LINCOLN DAKA

2365730

Supervisor: Professor Frederik Booysen

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy,

in the Faculty of Commerce, Law and Management,

School of Economics and Finance.

21st August 2024

DECLARATION

I, Lincoln Daka do hereby declare that this thesis which is submitted for the degree of Doctor of Philosophy to the University of the Witwatersrand, Johannesburg is my own work, except otherwise indicated and acknowledged. It has not been submitted before for any degree or examination in any other university.

Signature of candidate: 

Date: 21st August 2024

Abstract

This thesis expands upon and enhances existing research in the field of health economics. The thesis consists of three separate yet interrelated chapters that examine the effect of education on key demographic variables: child health, fertility and HIV/AIDS in Zambia, three key factors affecting the progress of development in Africa. The endogeneity problem is present in all of the three empirical papers examined. To circumvent this endogeneity problem and establish a credible causal effect, we explore the impact of Zambia's 2002 Universal Free Primary Education (UFPE) policy which created an exogenous source of variation in education as a quasi – experiment. The three substantial empirical studies, employ the same econometric methodology, a Regression Discontinuity Design (RDD), whose appealing feature is local randomisation. This characteristic has distinguished the method from other evaluation methods in terms of estimating unbiased treatment effects. Another advantage of the fuzzy Regression Discontinuity design is that it can account for the endogeneity of the treatment variable. The utilisation of the fuzzy Regression Discontinuity design is a valuable contribution in all of the research. Furthermore, every chapter makes a unique contribution within its respective sector.

We outline Zambia's Universal Free Primary Education (UFPE) Policy and also present the Regression Discontinuity Design methodology framework. We find significant causal impacts of maternal education on child health measured by height-for-age, weight-for-height and Weight-for-age. The findings also indicate that maternal education is associated with a reduction in the prevalence of stunting and underweight and no effect wasting contrary to other research. We present evidence of the several mechanisms by which maternal education impacts child health.

The results of our study indicate that a greater level of maternal education exerts a beneficial influence on child health through the postponement of marriage, the reduction in total fertility, and the delay in the age of first childbirth and sexual debut. Additionally, we have discovered indications of positive assortative mating. Furthermore, education empowers moms by facilitating their access to information via television and newspapers, equipping them with knowledge about the ovulation time, and helping them to make well-informed decisions regarding contraceptive techniques.

Conventional wisdom posits that decreased fertility may indicate the presence of "superior quality" children and increased rates of survival for both mother and child. Can education serve as a catalyst for decreasing fertility rates in developing nations? We find that female education reduces the number of children ever born. We present evidence of the reduction in total fertility as a result of female education. We also show that female schooling reduces the preferred number of children and increases the age at first birth. We find that female schooling affects fertility through age at first sex and marriage, literacy, assortative mating and the knowledge effect. There is no evidence to suggest that female schooling has a major impact labour market participation. We present evidence of the heterogeneous impacts of a mother's education based on "poor versus wealthy" criterion, whether rural/urban status, region and religion.

We also present evidence of the effect of female education on the HIV seroprevalence status, number of sexual partners and knowledge of HIV transmission mechanisms. We show that female education lowers HIV seroprevalence status, decreases the number of sexual partners and increases HIV knowledge. Our research suggests that educated women are more likely to have a deep and detailed understanding of HIV. Lastly, we present evidence of the heterogeneous effects of female education by household status on HIV related outcomes.

DEDICATION

To my parents, wife and siblings.

ACKNOWLEDGEMENTS

I recognise the divine intervention in granting me the fortitude, tenacity, and sound physical condition during the composition of this thesis. The task of writing a PhD thesis is arduous and demanding. My deepest gratitude goes to my supervisor, Prof. Frederik Booysen, for his support and guidance throughout the research. I am eternally grateful to have him as my supervisor for the eruditeness, dedication, drive, endurance, and readiness to support me throughout the entire thesis writing process. I have enormously benefited from the countless hours of his critiques and discussions.

I am immensely grateful to my friends, both in the graduate programme and throughout the globe, who provided unwavering support during my endeavour. I am sincerely grateful to the colleagues back home in Zambia from the Copperbelt University, Department of economics for their comments and constructive feedback on this research.

Finally, I would want to express my gratitude to my wife, Zita Lifelo, for being by my side during both the exhilarating and challenging times. I appreciate your unwavering faith in me.

Table of Contents

Abstract iii

DEDICATION v

ACKNOWLEDGEMENTS vi

CHAPTER 1: Introduction 1

1.1. Background to the study 1

1.2. Problem statement 5

1.3. Objectives 8

1.4. Main Findings 8

1.5. Contribution 10

1.6. Organisation of the study 12

CHAPTER 2: Methodology and Econometric Strategy 13

2.1. Zambia's Universal Free Primary Education (UFPE) Policy 13

2.2. Methodology: Regression Discontinuity Design 17

2.2.1. Principles of RD Design Using an Example from Higher Education 18

2.2.2. Formal Presentation 22

2.2.3. The Sharp RD Design 22

2.2.4. The Fuzzy RD Design 28

2.2.5. Special Cases 32

2.2.6. Comparing RDD and 2SLS 34

CHAPTER 3: Maternal education and child health: Quasi-experimental evidence from Zambia 36

3.1. Introduction 36

3.2. Data and Empirical Strategy 38

3.2.1. Data Sources 38

3.2.2. Empirical Strategy 39

3.3. Results 42

3.3.1. Data and Descriptive Statistics 42

3.3.2. Bandwidth Selection 43

3.3.3. Graphical representation 44

3.3.4. First Stage Results – Effects of the UFPE policy on Education 46

3.3.5. Second-Stage Results 48

3.4. Discussion 52

3.4.1. Maternal Education improves child health ... 52
3.4.2. Potential Mechanisms ... 53
3.4.3. Heterogeneity Effects: Poor versus Wealthy subsample results ... 57
3.5. Validity and Robustness Checks ... 59
3.5.1. Internal Validity ... 59
3.5.2. Unconfoundedness ... 59
3.5.3. External Validity ... 61
3.6. Robustness Checks ... 61
3.7. Conclusion ... 65
CHAPTER 4: Free primary education, female schooling and fertility: A regression discontinuity design ... 68
4.1. Introduction ... 68
4.2. Data and Empirical Strategy ... 71
4.2.1. Data Sources ... 71
4.2.2. Empirical Strategy ... 72
4.3. Results ... 76
4.3.1. Data and Descriptive Statistics ... 76
4.3.2. Bandwidth Selection ... 78
4.3.3. First Stage Discontinuity in educational attainment ... 78
4.3.4. First Stage Results – Effects of the UFPE policy on years of schooling ... 81
4.3.5. Effects of Education on fertility outcomes ... 82
4.4. Discussion ... 87
4.4.1. The Effects of Education on Fertility Outcomes ... 88
4.4.2. Heterogeneous effects: Poor versus Wealthy subsample results ... 89
4.4.3. Potential pathway mechanisms for the full sample ... 90
4.4.4. Heterogeneous effects of education on potential mechanisms ... 93
4.5. Robustness and validity ... 95
4.5.1. Internal Validity ... 95
4.5.2. External Validity ... 98
4.6. Robustness Checks ... 98
4.7. Conclusion ... 101
CHAPTER 5: The effect of female education on HIV seroprevalence status and knowledge in Zambia: Universal Primary Education as a natural experiment ... 103

5.1. Introduction 103
5.2. Data and Empirical Strategy 106
5.2.1. Data Sources 106
5.2.2. Empirical Strategy 106
5.3. Results 108
5.3.1. Data and Descriptive Statistics 108
5.3.2. Bandwidth Selection 110
5.3.3. First Stage Discontinuity in educational attainment 111
5.3.4. First Stage Results – Effects of the UFPE policy on years of schooling 112
5.3.5. Second Stage Results: Education and HIV related outcomes 114
5.3.6. Education and HIV Knowledge 116
5.4. Discussion 118
5.4.1. Second Stage (2SLS) Results 119
5.4.2. Heterogeneous Effects 122
5.4.3. Potential Transmission Mechanisms 123
5.5. Identification, internal and external validity 126
5.5.1. Internal Validity 126
5.5.2. External Validity 128
5.6. Robustness Checks 129
5.7. Conclusion 131
CHAPTER 6: Conclusion 133
References 138

List of tables

Table 2.1: Compliance type by treatment and the cutoff point 30
Table 2.2: Compliance type by treatment and the cutoff point given monotonicity 31
Table 3.1: Description of key variables 39
Table 3.2: Descriptive Statistics 42
Table 3.3: First Stage – Effect of the UFPE Policy on Maternal Education 47
Table 3.4: Second Stage Results: IV Estimates 50
Table 3.5: Potential Mechanisms – Fertility related outcomes 56
Table 3.6: Pretreatment covariates 60

Table 3.7: First stage estimates with different bandwidths.. 62
Table 3.8: Main Results with different functional forms... 63
Table 3.9: Manipulation Test of running variable 64
Table 4.1: Description of key variables 72
Table 4.2: Summary Statistics 77
Table 4.3: First Stage Estimates – Effect of the UFPE Policy on Maternal Education................ 82
Table 4.4: Second Stage (2SLS), Reduced form and OLS Estimations 85
Table 4.5: Potential Mechanisms – Fertility related outcomes.. 94
Table 4.6: Pretreatment covariates.. 97
Table 4.7: Manipulation Test of running variable 97
Table 4.8: First stage estimates with different bandwidths.. 99
Table 4.9: Main Results with different functional forms.. 100
Table 5.1: Summary Statistics 109
Table 5.2: First Stage Estimates -– Effect of the UFPE Policy on Maternal Education............ 114
Table 5.3: Years of Schooling, HIV seroprevalence status (HIV positive = 1) and number of sexual partners - Full sample 116
Table 5.4: Years of Schooling and Knowledge of HIV... 117
Table 5.5: Years of Schooling, HIV seroprevalence status subsample (HIV positive = 1) and Number of sexual partners Subsample 122
Table 5.6: Transmission Mechanisms.. 124
Table 5.7: Pretreatment covariates... 127
Table 5.8: Manipulation Test of running variable 128
Table 5.9: First stage estimates with different bandwidths.. 129
Table 5.10: Main Results with different functional forms... 130
Table A1: Heterogeneous Effects of maternal education on Rural/Urban. Region and Religion…………………………………………………………………………………153
Table A2: First stage results for Men 154
Table A3: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2018 155
Table A4: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2013-14.............................. 156
Table A5: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2007 157

List of Figures

Figure 2.1: Primary Gross Enrolment Ratio in Zambia, 1970 – 2017... 14
Figure 2.2: Discontinuity in math exam score at the cutoff score in the diagnostic placement test (the sharp RD design) 20
Figure 2.3: Discontinuity in math exam score at the cutoff score in the diagnostic placement test (the fuzzy RD design)... 20
Figure 2.4: Discontinuity in the probability of treatment at the cutoff score in the diagnostic placement test (the sharp and the fuzzy RD designs) 21

Figure 3.1: Women’s Schooling by Birth Year Cohort 45
Figure 3.2: Height for age z-score by birth year cohort 45
Figure 3.3: Weight for age z-score by birth year cohort 46
Figure 3.4: Weight for height z-score by birth year cohort 46
Figure 3.5: Manipulation Testing Plot 64
Figure 4.1: Women’s Schooling by Birth Year Cohort 79
Figure 4.2: Total children ever born by birth year cohort 80
Figure 4.3: Preferred number of children by birth year cohort 80
Figure 4.4: Manipulation Testing Plot 98
Figure 5.1: Women’s Schooling by Birth Year Cohort 111
Figure 5.2: HIV Seroprevalence status by birth year cohort 112
Figure 5.3: Manipulation Testing Plot 128
Figure A1: Men’s Schooling by Birth Year Cohort 154

CHAPTER 1: Introduction

1.1. Background to the study

The importance of education, particularly for women, has been recognised as the primary factor in generating beneficial impacts on economic and social results. Education has significant impacts on economic growth and income inequality (Mankiw, Romer, & Weil, 1992), as well as on worker productivity and earnings in the labour market (Card, 1999). Additionally, it has spillover effects on the productivity of other workers (Acemoglu & Angrist, 2000; Iranzo & Peri, 2009). Education affects non-market outcomes like individual's health, fertility, children's long-term economic outcomes, and child quality among other outcomes (Grossman, 2006). Given the widespread impact that education has on the economy and individual well-being, many developing countries have invested in school construction programs, conditional cash transfers, and compulsory schooling laws to give their population greater access and push to achieve more education (Burde & Linden, 2013; Cui, Liu, & Zhao, 2019; Schultz, 2004).

Education has a crucial role in promoting economic development and reducing poverty. There is undeniable scientific evidence that education is strongly correlated with economic growth and development (Schultz, 1960). Furthermore, additional empirical research has shown that education provides economic and social advantages to individuals. More precisely, education enhances individual productivity and income, decreases civil strife, improves personal health outcomes, and lowers fertility rates. Further, education and child health status have a robust positive correlation (Le and Nguyen, 2020; Keats 2018; Hahn et al., 2018; Strauss and Thomas, 1991).

Education can help reduce the risk of HIV transmission by providing more information on HIV and related prevention techniques (Aguero and Bharadwaj, 2014; De Walque, 2009). Education can reduce the likelihood of HIV transmission by improving cognitive skills needed for making informed choices and promoting better financial security (Psacharopoulos, 1994; Schultz, 1960; Willis, 1973). Education can help decrease the likelihood of HIV transmission by decreasing women's involvement in transactional sex and improving their ability to form relationships with partners who have a lower risk of transmitting the virus (Baird et al., 2012; Behrman and Rosenzweig, 2002; Rosenberg et al., 2015). Consequently, numerous African nations have implemented educational programmes post-independence with the aim of providing accessible education to all children, regardless of their gender, age, or region. These policies also strive to support the attainment of the Millennium Development Goals (MDGs) and expedite progress towards the Education for All (EFA) agenda.

The accessibility of education in Africa has significantly increased in the last twenty years as a result of policy execution. The Gross Enrollment Rate (GER) for elementary education now stands at 98 percent, as reported by the World Bank Open Data. Furthermore, the mean duration of schooling, particularly among females aged 15 and older, has significantly risen throughout Africa, albeit with disparities. For instance, the mean educational achievement of women in North Africa rose by 84 percent, from 3.5 to 6.5, whereas in Sub-Saharan Africa it improved by 48 percent, from 3.1 to 4.7. Given the rise in educational achievement and the substantial empirical data highlighting the significance of education, numerous social analysts and international organisations have fervently endorsed education, particularly maternal education, as a strategic approach to addressing global problems such as poor child health, high fertility rates and HIV related outcomes.

The majority of countries in Africa are classified as developing nations, and as is typical for such nations, they encounter numerous obstacles. Among these obstacles are poor child health issues, increasing fertility rates, and the HIV/AIDS pandemic. This thesis aims to enhance knowledge on these interconnected topics by analysing three specific concerns: (1) the relationship between maternal education and child health outcomes, (2) the relationship between female education and fertility and (3) the relation between female education and HIV related outcomes.

The first chapter investigates the effect of parental (maternal) education on child health. Poor early-life health, specifically malnutrition during infancy, can have long-lasting consequences on an individual's life. These consequences include lower educational attainment, higher vulnerability to chronic diseases, and a decrease in both productivity and wages. Consequently, it is crucial to examine the factors that influence the well-being of children and identify potential measures to enhance it. Multiple studies indicate that the level of education a mother has is a significant factor in determining the health of her child (Grossman, 1972; Thomas et al., 1991; Schultz, 2004; Grossman, 2006; Chou et al., 2010; Keats, 2018; Le and Nguyen, 2020). Education might potentially impact the health of children, either directly or indirectly.

Education has a direct impact on the mother's ability to collect, assimilate, and analyse knowledge, leading to improved parenting skills. Mothers with higher levels of education may possess a more comprehensive comprehension of medical procedures and, thus, exhibit greater awareness of the benefits associated with medical interventions (Dinser et al., 2014). Mothers who have had more education may experience an indirect consequence, such as earning higher incomes. Pursuing a higher education can provide access to more lucrative employment opportunities that demand advanced skills and expertise. The additional income generated from these occupations can be allocated towards enhancing one's well-being and planning for unforeseen health crises. Conversely, multiple studies have demonstrated that a mother's employment may adversely impact the health of her child (Le and Nguyen, 2020). Assortative mating is another possible factor that could account for the positive impact of a mother's education on the health of her child. Multiple studies indicate a significant correlation between women's education and the educational achievements of their husbands (Behrman and Rosenzweig, 2002; Hahn et al., 2018; Le and Nguyen, 2020).

We also examine various mechanisms through which maternal education impacts child health. We have identified four main groups of mechanisms that contribute to these positive intergenerational impacts. These mechanisms include: (i) the mother's fertility behaviour, which is measured by the number of children she has and her age at first birth, (ii) assortative matching, which is measured by the husband's education level, (iii) access to information and health knowledge, which is measured by factors such as whether the mother watches television, reads newspapers, knows about the ovulation cycle, and uses contraception, and (iv) labour market outcomes, which are measured by various indicators of the mother's participation in the workforce and her earnings.

Specifically, mothers with greater levels of education are more likely to decrease the number of children they have, delay the age at which they have their first child, and marry spouses who are also highly educated. Education additionally motivates women to seek information through television and newspapers, to gain understanding of the ovulation cycle, to utilise contraceptive methods, and to actively participate in and increase their earnings from market work. Finally, we detect heterogeneity in the impacts of mother's education on child health by household health status and type of place of residence

The second chapter of this thesis examines the relationship between women's education and fertility outcomes. The child quantity-quality trade-off hypothesis proposed by Gary Becker posits that smaller family sizes contribute to improved outcomes for children. This is because parents with fewer children have a greater amount of resources at their disposal to allocate to each individual child. Hence, comprehending household choices regarding the number of children is important for formulating strategies that efficiently diminish family size and enhance children's welfare. Prior empirical investigations of developing nations that explored this correlation have encountered challenges in addressing the issue of omitted variables, such as a woman's inert ability, preferences, and community background. Education has a dual effect on fertility: it decreases the overall fertility rate and also causes a delay in childbearing due to the higher opportunity cost associated with having children for women (Becker and Lewis, 1973). Neoclassical fertility models suggest that higher levels of female education result in a trade-off between the number of children and their quality, as proposed by Becker (1960) and Becker and Lewis (1973).

Various theoretical frameworks have been proposed to elucidate the influence of female education on fertility. In general, education can delay fertility through the following mechanisms: First, education has the potential to expand an individual's knowledge. Consequently, an increase in female education is likely to improve women's capacity to utilise contraception with effectiveness and efficiency (Rosenzweig and Schultz, 1989). Second, the education of women amplifies the opportunity cost of getting married and having children at a young age. Child care requires a greater amount of time and effort compared to other activities. As the value of a woman's time increases, she is likely to prefer spending less time on child care activities and desire fewer children in order to improve her involvement in the labour force (Schultz, 1988). Third, women who have had a formal education are more inclined to actively participate in household decision-making processes, particularly when it comes to determining the optimal number of children for their family. The augmentation of women's years of education enhances their ability to negotiate and make autonomous decisions, increasing their "bargaining power" without the need to contact their husbands. Finally, ensuring that women remain enrolled in educational institutions decreases the likelihood of their involvement in actions that may result in unintended pregnancies. The process described by Black et al. (2008) is referred to as the "incarceration effect". Ultimately, the marriage market functions in a manner that significantly increases the likelihood of an educated woman being paired with an educated man. The phenomenon of couples with similar levels of education being paired together is known as assortative matching in the field of fertility research. This concept has been discussed by various scholars including Breierova and Duflo (2004), Lavy and Zablotzky (2015), and Tequame and Tirivayi (2015).

The third chapter of this thesis examines the relationship between female education and HIV related outcomes. These outcomes include HIV seroprevalence status, number of sexual partners and HIV knowledge. Education can reduce the risk of HIV transmission by providing more opportunities to learn about HIV and how to avoid it (Aguero and Bharadwaj, 2014; De Walque, 2009). Education can reduce the risk of HIV transmission by improving cognitive abilities needed to make informed choices and promoting better socioeconomic stability (Psacharopoulos, 1994; Schultz, 1960; Willis, 1973).

We also investigate the mechanism by which female education may influence one's HIV seroprevalence status and other HIV related outcomes, employing a consistent methodology to establish causal relationships. We begin by examining the impact of education on various indicators pertaining to sexual behaviour during adolescence and adulthood, namely age at first sexual debut and age at first marriage as these are associated with the risk of contracting HIV. Another vital mechanism is literacy. Literacy holds significance as a rudimentary indicator of abilities or cognitive advancement that occurs within educational institutions.

This study uses data from three rounds of the Zambia Demographic and Health Surveys (ZDHS), which are nationally representative surveys of reproductive-aged women conducted in Zambia in 2007, 2013–2014, and 2018. In addition these three rounds, of the ZDHS also contain Aids Indicator Surveys (AIS) as sub surveys where vital information relating to HIV status and knowledge is obtained.

1.2. Problem statement

The correlations between female education, reduced fertility rates, and enhanced child health and human capital are widely recognised based on empirical evidence from studies conducted by Cochrane (1979), Strauss and Thomas (1991), and Grossman (2006). Consequently, the emphasis on promoting women's education in developing nations is justified as it can effectively contribute to achieving various social goals. The World Bank has stated that educating females is the most effective investment for accomplishing development goals. However, there is currently no conclusive data to support the notion that an increase in female education directly leads to a decrease in fertility and an improvement in child investments and outcomes. Although there is an increasing amount of experimental and quasi-experimental research indicating that education has a postponing effect on fertility, the evidence is inconclusive over whether fertility rates decrease overall. Empirical evidence demonstrating a cause relationship between education and a decrease in adolescent pregnancies has been discovered (Breierova and Duflo, 2004; Black et al., 2008; Ferre, 2009; Duflo et al., 2015; Lavy and Zablotsky, 2015; Adu-Boahen and Yamauchi, 2017; Ozier, 2018).

Nevertheless, within the limited number of research that investigate the potential impact of education on general fertility rates, the findings are inconclusive. Although certain studies (Osili and Long, 2008; Lavy and Zablotsky, 2015) suggest that education is associated with a decrease in the number of children born, other research (Breierova and Duflo, 2004) indicates that highly educated women who postpone childbearing in their teenage years eventually have a similar number of children as less educated women. Although there is substantial empirical evidence supporting the relationship between education and fertility or the timing of marriage, accurately estimating the causal influence is challenging because educational attainment is endogenous in both the equations for fertility and the timing of marriage. For example unobserved characteristics including ability and preference, can have an impact on both educational achievement and fertility outcomes.

Regarding child health, in 2018, stunting among children under the age of 5 in Africa was 30 percent, which is significantly higher than the global average of 22 percent, and three times higher than the rate in Latin American and Caribbean countries (according to World Bank Open Data). There is a limited number of comprehensive studies that thoroughly investigate the correlation between maternal education and child health outcomes, and the findings from these studies are equally inconclusive. Breierova and Duflo (2004), Chou et al. (2010), and Currie and Moretti (2003) discover that parental education has beneficial impacts on child health in Indonesia, Taiwan, and the U.S., correspondingly. Le and Nguyen (2020) find that maternal education has positive impacts on child health in 68 developing countries. In contrast, McCrary and Royer (2011) and Lindeboom et al. (2009) observe minimal or negligible impacts on child health in the United States and United Kingdom, respectively. Although the connection between parental education and child health is unquestionable, the mechanisms via which this connection operates remain unknown, posing challenges in establishing a causal relationship. The primary concern pertains to the methodology used for estimating child health production functions, which establish a connection between "health outputs" and endogenous inputs (Aslam and Kingdon, 2012).

The prevalence of HIV/AIDS among young women in Sub-Saharan Africa (SSA) is a substantial health risk. HIV continues to be a substantial worldwide health issue as in 2021, 650 000 people were lost to AIDS and 1.5 million people newly acquired HIV (UNAIDS 2023). In 2022, almost 1.3 million people were infected, as reported by UNAIDS (2023). Multiple studies have extensively recorded the enduring ramifications of HIV/AIDS, including a significant decline in life expectancy, the demise of adults in their reproductive years, and the vulnerability of orphaned children. Additionally, the morbidity and absenteeism of workers have led to the displacement of businesses, while the death of economically productive adults has left their families in dire financial circumstances. These are just a few examples of the numerous implications associated with HIV/AIDS.

Considering the formidable task of administering antiretroviral therapy to a growing number of HIV-infected individuals, it is crucial for global initiatives to give priority to preventative techniques in tackling the HIV epidemic. Therefore, the adoption of formal education, with a specific focus on educating females, has been widely acknowledged as a preventive action to reduce the spread of HIV in society. It is widely believed that education can provide individuals with the knowledge and skills needed to reduce the chance of transmission by practicing preventive measures, such as using condoms and taking other precautions. However, there is a lack of conclusive evidence to support this claim (Vandemoortele J. and Delamonica E., 2000; Aguero and Bharadwaj, 2014; De Neve et al., 2015; Lindskog and Durevall, 2021).

The challenge lies in establishing a causality between female schooling and the likelihood of contracting HIV, given the significant correlation between educational achievement and factors such as socioeconomic status, psychological characteristics, and unobservable variables like cognitive ability and family preferences. These factors present difficulties in observational studies because fo their intricate nature, making it challenging to achieve complete control over them. Additionally, they can also impact the likelihood of contracting HIV.

Zambia implemented the universal free primary education (UFPE) compulsory universal primary education policy in 2002. This policy has resulted in a rise in the enrollment of students in primary schools and an improvement in the rate at which students are staying in school. The significant policy alteration can function as a natural experiment for examining the impact of female schooling on child health, fertility and HIV related outcomes. This surge in primary school enrollments creates a natural experiment, allowing the use of an instrumental variable (IV) framework in a fuzzy regression discontinuity design (RDD) to establish credible causal relationships between female education and the three specific outcomes, utilising this policy shift.

1.3. Objectives

This study utilises the implementation of Zambia's Universal Free Primary Education (UFPE) programme in 2002, which abolished school fees for primary school, to establish credible causal relationships and provide new insights on the effects of women's education on child health, fertility and HIV related outcomes. The aim of this study therefore, is to examine the effect of parental education on child and adult health in Zambia by applying a regression discontinuity design analysis. Specifically, the thesis objectives are as follows:

I. To examine the effect of the UFPE educational policy on women's education
II. To examine the effect of maternal education on child health outcomes
III. To examine the effect of female education on fertility outcomes
IV. To investigate the effect female schooling on HIV knowledge, risk and seroprevalence status.

1.4. Main Findings

The results of the first stage analysis show that for women who were assigned to treatment or benefited from the 2002 UFPE policy (14 years and younger), the average effect of the policy was 0.68 more years of education compared to their counterparts in the control cohort who were not assigned to treatment (15 years and older). Consequently, there is a noticeable (jump in years of schooling) at the discontinuity with an increase in educational attainment among individuals born in 1988. The impact is substantial from an economic standpoint and has a high level of statistical significance. We also uncover significant heterogeneities in the effect of the UFPE educational policy on female education based on household wealth and rural/urban status.

In chapter 3 we estimate the effect of maternal education on child health outcomes. In order to address the issue of endogeneity, we employ an instrumental variable (IV) approach to estimate the relationship. This approach relies on the exogenous variation in education that occurs due to the respondent's age at the time when the 2002 education policy was implemented in Zambia. We utilise this external alteration in education to tackle the possible endogeneity of schooling. We find that female schooling increases children's height-for-age z-scores. Our results equally show that the education of the mother increases the weight-for-height z-score. For the weight-for-age z-scores, the IV estimates show that controlling for age, children of women with extra years schooling have more weight. The results equally show that an additional year of schooling by the mother decreases incidences of stunting and underweight. We find no effect on wasting. For heterogeneous effects, the effect of an additional year of schooling on height-for-age z-scores is higher for the wealthy subgroup compared to the poor counterparts. As potential mechanisms through which maternal education affects child health, we find significant effects of maternal education on age at first birth and marriage, the number of children ever born, the education of the partner, literacy and variables that proxy information access and health knowledge. We find no effect of maternal education on labour market outcomes in all forms.

In chapter 4, we analyse the effects of female education on fertility. The paper uses the UFPE education policy as an exogenous source of variation in education to estimate the causal effect of additional education on fertility outcomes. We implore the implementation of Zambia's UFPE education policy reform using a quasi-experimental methodology, a method that has not been used before to assess possible effects. A fuzzy regression discontinuity design (RDD) is employed to identify exogenous variation in female schooling. Subsequently, instrumental variable (IV) estimations are performed to investigate the influence of increased primary education on fertility. We use three measures of fertility outcomes: total number of children ever born per woman, the ideal number of children and the age at first birth. We find that female schooling results in a reduction in the total number of children born. We also find that schooling reduces the ideal (desired) number of children. An additional year of schooling also increases the age at first birth. In general, we found that increased female education lowers the average number of children (actual and ideal), and it increases the age at first birth. As part of potential mechanisms, our results show that show that years of female schooling delays the age at first marriage by about. The results also show that women who get more schooling are likely to postpone their age at first sexual intercourse. We find evidence of positive assortative mating. Women with more education are more likely to marry males with more education. There are also heterogeneous effects of additional education on fertility outcomes.

In chapter 5, we examine the effect of female schooling on HIV seroprevalence status, number of sexual partners and HIV knowledge. To mitigate the endogeneity of schooling and establish a credible causal relationship, we employ an instrumental variables methodology that leverages age-specific exposure to an educational reform implemented in Zambia. We use this exogenous source of variation in educational attainment, created by this age – specific nature of the reform thereby estimating the causal impact of female schooling on HIV seroprevalence status, HIV knowledge and number of sexual partners. The UFPE reform had a significant protective effect on those women exposed to the reform. We find that female schooling has a protective effect against HIV seroprevalence status and lowers HIV seroprevalence status (HIV risk). Our estimates also show that female schooling reduces the number of sexual partners. Finally, our results equally show that there exists a statistically significant association between educational attainment and the level of HIV knowledge.

1.5. Contribution

This work adds to the body of knowledge regarding the relationship between good health and the number of years of formal education completed. The discussion of the relationship between good health and formal education has until the last decade been correlational. Recent evidence however, points to a causal relationship. Unravelling multiple econometric issues is generally difficult. Fortunately, a recent policy reform in Zambia's school system allows for the application of reliable methods to investigate a potential causal effect utilising a regression discontinuity framework. The UFPE policy abolished school fees for primary education and this increased the years of schooling of women born in 1988 or later than women born before 1988. This discontinuity in female education is the basis of the regression discontinuity design used in this study. The regression discontinuity design employed in this thesis has only been utilised in a limited number of prior research that investigated compulsory schooling regulations in Africa.

Chapter 3 contributes to an emerging debate about whether more parental schooling causes better child health. The chapter contributes to the literature between maternal education and child health. This chapter advances our knowledge of the relationship between maternal education and child health in a number of ways. Firstly, we examine Zambia, one of the developing countries with a fast expanding population that has a poorer healthcare system and relatively low levels of education. The prevailing body of research has primarily concentrated on developed nations. Second, the study adds to literature by providing evidence on whether the impact of the Universal Free primary education (UFPE) policy is the same across different subgroups. In some situations, there is general agreement that a link exists, but there is ongoing discussion regarding causality. We identify credible causal estimates subject to various robustness specifications. Third, we complement previous studies by using a mediation framework to examine mother-children relationships and the mechanism behind the improvement of children's well-being using rich household survey data. Lastly, this chapter also contributes to the literature by providing empirical evidence of the relationship between maternal education and children's well-being when access to education for women is low.

Chapter 4 contributes to the relationship between years of formal schooling and fertility outcomes in a developing country. This chapter provides three primary contributions to the existing body of literature. First, it presents new findings regarding the causal effect of female education on fertility in Zambia. This contributes to the existing research on the effects of education on fertility in a developing country with high birth rates. Although there is an increasing amount of research investigating the cause effect of education on fertility outcomes, only a few studies have specifically examined lower-income sub-Saharan African countries using reliable estimating methods. Second, a significant portion of the research conducted in this area has primarily concentrated on the effect of female schooling on adolescent females and their experiences with pregnancy. We contribute to the literature by finding empirical support for a reduction in total fertility as a result of increased education. Third, this research enhances the existing body of knowledge on the demographic impacts of education by elucidating the possible pathways by which female education affects fertility. Finally, our results support the perspective put forth by Black et al. (2008) regarding the "incarceration effect," which suggests that educational institutions serve as a deterrent for women to partake in activities that may result in unintended pregnancies.

Chapter 5 contributes to the few papers that have examined the relationship between years of formal schooling and HIV seroprevalence status and knowledge. The endeavour of establishing a causal association between female education and HIV seroprevalence status is intricate and fraught with difficulties. This chapter contributes to the literature in three main ways. Firstly, numerous studies have discovered a negative correlation between education and HIV infection in sub-Saharan Africa. Nevertheless, they fail to offer insights into the correlations. Only a limited number of research utilise methodologies that seek to establish a causal effect of education on HIV infection risk, with diverse outcomes (Psaki et al., 2019). Our work therefore builds on the few papers that establish causality between HIV seroprevalence status and years of formal schooling in Sub-Saharan Africa by providing new evidence from Zambia which has a national HIV prevalence of about 11 percent among adults (ZHDS, 2018). Secondly, we contribute to the literature by analysing the mechanisms that contribute to the protective effect of female education against HIV infection. Lastly, as a contribution, we examine heterogeneous effects of education by household wealth on HIV seroprevalence status and knowledge. From a policy perspective, knowing the underlying mechanisms can aid with focused programme design that amplifies the advantages of women's education.

1.6. Organisation of the study

The remaining part of this thesis is organised as follows: Chapter 2 presents the methodology and econometric strategy. Chapter 3 examines the effect of maternal education on child health. Chapter 4 examines the effects of female education on fertility outcomes. Chapter 5 investigates the effects of education on HIV seroprevalence status, number of sexual partners and knowledge. Chapter 6 provides a conclusion and examines the potential consequences for future research and policy discussions.

CHAPTER 2: Methodology and Econometric Strategy

2.1. Zambia's Universal Free Primary Education (UFPE) Policy

In order to draw causal inference regarding the effect of maternal education on child and adult health outcomes and female education on fertility and HIV risk using the RDD framework, there needs to be a discontinuity (jump) or exogenous source of variation in educational attainment which we can then attribute to the possible health outcomes given the endogenous nature of the education variable. Here, we briefly outline the Universal Free Primary Education (UFPE) Policy introduced in Zambia in 2002, as a source of this exogenous variation in women's education.

Since Zambia gained independence from the United Kingdom in 1964, the government was obligated to adopt the UFPE Policy, which was designed to address inaccessibility, unaffordability, injustice, and high school drop-out rates. The government of the Republic of Zambia introduced the UFPE in 2002 through the Ministry of General Education (MoGE) in response to falling education funding and low gross enrollment rates.

Zambia's UFPE Policy, which was announced by the Ministry of Education and immediately implemented in January 2002, was meant to make education more accessible, affordable, quality, relevant, and equitable for everyone. It also aimed to improve the socioeconomic well-being of all citizens and to achieve a high standard of living for everyone, as well as to make education more affordable and accessible (Bentaquet-Kattan, 2006).

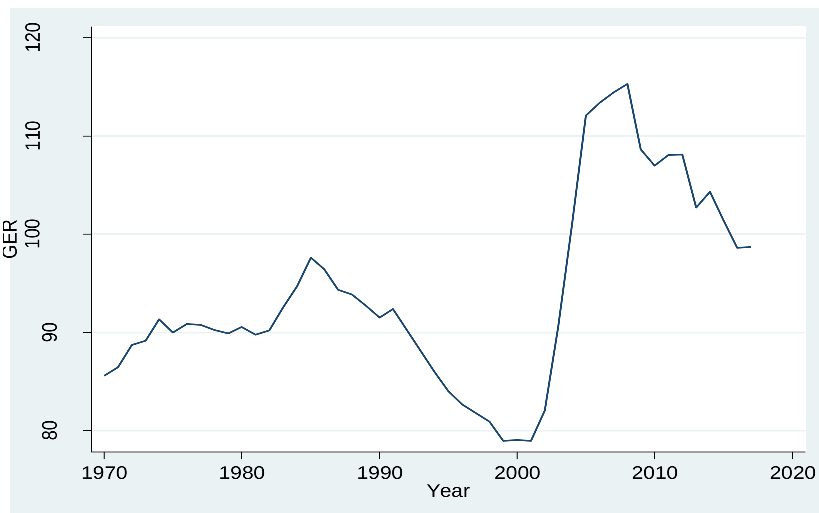
Prior to the aforementioned initiative, the country had witnessed development in the primary school subsector between 1972 and 1990, although the gross enrolment rate (GER)[1] was only around 59 percent (World Bank, 2015). Moreover, the system favored a tiny percentage of pupils who were deemed to be intellectually competent at the expense of the majority, so fostering an imperfect attitude of competitiveness at the expense of cooperation (Masaiti *et al.*, 2016). Similarly, high debt servicing costs directly impacted government budgetary allocations for health,

[1] The primary school gross enrollment rate is defined as total enrollment in primary school divided by the primary age school population. This figure can exceed 100 % if children over primary school age are still in primary school.

education, and agriculture in the 1990s and early 2000s, and many social safety nets were destroyed (World Bank, 2018).

The findings, together with supporting literature, attested to the government's UFPE Policy’s success in increasing public primary school net enrolment rates from 60% in the 2000s to over 120% in 2014, supported by significant infrastructure development, constant teacher recruitment, and sufficient school supplies to meet demand (see Figure 2.1).

Figure 2.1: Primary Gross Enrolment Ratio in Zambia, 1970 – 2017



Source: World Bank Data

The policy sought to eliminate unregulated tuition fees, prevalent then and to reduce parental expenditure on textbooks and other direct schooling costs. The government abolished tuition fees for all public primary schools. Prior to the abolishment of school fees in Zambia, parents were responsible for between 50% and 75% of overall primary education expenses (Bentaquet-kattan, 2006). Additionally, before fees were eliminated in Zambia, households spent 33% of their discretionary income on education (Boyle et al. 2002).

The implementation of this education programme resulted in a large increase in total primary school enrollment, which increased from approximately 1.7 million pupils in 2002 to nearly 3.1

million in 2014. (World Bank, 2015). The UFPE's major purpose was to ensure that all Zambians, male or female, were given the same chance to enroll in and complete basic education.

The establishment of UFPE policy in 2002 helped all children in basic school, regardless of grade level. Pupils who left school before 2002 owing to lack of school fees were allowed to return. A sizable proportion of new entrants were really pupils who had previously left school as a result of failure to pay school fees, particularly in rural areas (Mwansa et. al. 2004). As a result of the UFPE policy implementation, the percentage change in the gross enrolment ratio (GER) initially rose by 5% from 78 in 2001 prior to the policy implementation to about 82 after the UFPE policy implementation in 2002 (Bentaquet-Kattan, 2006). The GER then rose steadily to about 115 in 2008 before dropping (See Figure 2.1). The major reasons being advance for the drop in GER include a consistent decline in financing the education sector as a percentage of the country's budget steadily from 20.2 percent in 2015 to 12 percent in 2020 (Masaiti et al., 2016), posing a threat to the country's continued advancement and aspiration to provide quality UFPE as envisaged, to all qualified citizens.

The Zambian education system has a 7-5-4 structure, namely 7 years at primary school, 2 and 3 years at junior and higher secondary school respectively, and 4 years at university for undergraduate degrees. The academic year begins in January and ends in December, and is divided into three terms. Primary education lasts through grades 1 to 7. The official primary school entrance age is 7 and pupils are expected to complete primary education by the age of 13. However, due to late entry into school, repetition or temporary dropout, a sizeable proportion of primary school pupils fall outside the official age range for primary schooling of ages 7 to 13, especially in rural areas. Further, according to ZDHS (2007), the grades pupils attend are quite similar to the theoretical grades for their age; however, it should be noted that older pupils in the age range 9-13 were also present in other lower grades such as primary grades 1 (19.5%) and 2 (55.4%). Equally, findings of the Zambia educational statistics bulletin (2017), indicate that between 2006 and 2017, only 47% of pupils started their grade 1 at exactly the age of 7 while about 8% started under the age of 7. The rest (about 45%) started over the age of 7 years old.

As a result, we choose the average age of admission into primary grade 1 to be 8 and identify a discontinuity for children born in 1988. This means, individuals born in 1988 (14 years old at

UFPE implementation in 2002) or later are classified as part of the treatment group, where as those born in 1987 (15 years old at UFPE implementation in 2002) or before are classified as part of the control group.

The identification strategy is that being 14 years or younger provided more access to school than being a few years older did, hence limiting one's prospects to benefit from the change.

We take advantage of the age-specific character of the UFPE reform to assess the causal influence of maternal education on child health outcomes and other relevant pathways. We instrument for education using the age of the parent in 2002 to examine whether this plausibly exogenous source of education is associated with changes in child and adult health outcomes.

Although the UFPE was a political decision made outside of the Ministry of Education, the Ministry was expected to implement the policy immediately, and therefore had little time to adequately plan for the transformation. Guidelines mandating the elimination of fees for grades 1 to 7 were prepared and circulated to schools. In addition, the government developed bursary schemes for clothing, instructional materials, and boarding costs for vulnerable groups (Mwansa et. al. 2004).

Prior to the UFFPE, the education system was characterized by low teacher motivation, inadequate training, shortage of textbooks and instructional materials, inadequate school infrastructure, and inequitable access to education, both between urban and rural areas and for girls. Zambia experienced declining enrollments, with the primary school gross enrollment decreasing from 95 percent in 1985 to 89 percent in 1996. Access to education was lowest for children in remote and rural areas. The primary school drop-out rate was about 31 percent. Parents paid a variety of fees including tuition, exam fees, Parent Teacher Association (PTA) dues, uniforms, and boarding fees.

The UFPE did seem to have an impact on enrollment of disadvantaged groups. Mwansa et al. (2004) found that a significant number of new enrollments, especially in rural areas, were students who had previously left school due to the inability to pay school fees. Enrollment among orphans and children with special needs also rose. However, they also found evidence of a persistent gender gap, particularly in rural areas.

At the time that UFPE was introduced there was a cost-sharing framework already in place under the Basic Education Sub-Sector Investment Program (BESSIP). Communities were required to support the construction and rehabilitation of schools. Under the UFPE, communities stopped contributing to infrastructure improvements because they interpreted the policy as meaning that the government would pay for all education costs (Mwansa et. al. 2004).

In order to replace the loss of school fees, the government increased the amount of grants paid to schools. However, due to rushed implementation of the policy, the government was not able to carry out a school census, or needs-assessments, and gave each school, big or large, urban or rural, the same amount of grant money. There is some evidence that the grants proved to be inadequate and that quality of education is suffering. Increased enrollments and teacher shortages have forced most schools to implement multiple shifts. While official data show that the student to classroom ratio improved slightly, one study of 352 basic schools found that 31.5 percent of schools conducted class under trees, 14.2 percent held multiple sessions, and some schools used tents (3.9 percent) (Mwansa et al. 2004). This suggests that official aggregate data obscures large inequities between schools and regions. The same study found that the student to teacher ratio increased from 46:1 to 52:1 between 2001 and 2003, partially due to a teacher strike, as well as hiring freezes that had left about 9,000 trained teachers out of work. In some rural areas, student teacher ratios are as much as 100:1 (Mwansa et at. 2004).

Teacher shortage has serious implications for the success of the UFPE. Some students receive only two hours of education a day, learning only a minimal amount of curriculum. The government responded to the problem of inadequate learning materials by decentralizing the procurement process, which is expected to reduce the cost of transporting materials, as well as free up funds for more materials.

2.2. Methodology: Regression Discontinuity Design

The history of RD design dates back to 1960. Prior to the late 1990s, it was not a customary instrument in mainstream Economics, but it enjoyed popularity in Psychology and other related fields. Key references in the Economics literature since the late 1990s include van der Klauuw (2002), Hahn et al. (2001), Imbens and Lemieux (2008), Imbens and Wooldridge (2009), and Lee and Lemieux (2010). In recent times, the technique has been more accessible to applied economists through the inclusion of textbook chapters in Angrist and Pischke (2009) and Wooldridge (2010). This chapter primarily relies on the works of Wooldridge (2010), Hahn et al. (2001), van der Klauuw (2002), Angrist and Pischke (2009), and Lee and Lemieux (2010). Specifically, we utilise Wooldridge (2010) to develop notation.

The following part introduces the basics of RD design through the use of an example. The subsequent section provides a formal presentation of RD design.

2.2.1. Principles of RD Design Using an Example from Higher Education

RD design typically leverages discontinuities in policy assignment, where the treatment is determined by a cutoff point. For example, there is an age cutoff at which some individuals become eligible for a retirement pension, or an income cutoff at which some households become eligible for a cash transfer. The determination of these cutoffs is based on arbitrary institutional frameworks. In order to take advantage of discontinuities, it is presumed that the units just on either side of the cutoffs are nearly identical in terms of both observable and unobservable characteristics that affect the outcomes of interest. The two groups are different in terms of the status of the respective treatments. One group is treated, and the other group is untreated. In this case, difference in outcomes of these two groups can be attributed to the different treatment statuses.

There are two types of RD design: the sharp RD design and the fuzzy RD design. In the former, the treatment assignment follows a deterministic rule. In the latter, the probability of being treated is discontinuous at a known cutoff point. Graphical representations of two types of Regression Discontinuity designs are provided below, using an example from the field of higher education. The example is derived from Pietro's work (2012). The statistics used as examples are purely theoretical and do not include any actual data. They generate artificial data.

In the given scenario, a university is aware that on an annual basis, a portion of the students who are now enrolled in an undergraduate programme possess inadequate readiness for their particular courses. Consequently, the university enrols these students in supplementary courses, known as remedial courses, to adequately prepare them for the undergraduate programme. In order to identify these students, the university administers a diagnostic placement test to all enrolled students at the beginning of their undergraduate career. Students who have scores below a specified cutoff score on the diagnostic placement test are considered eligible. The other students do not meet the requirements.

An individual may have an inclination to observe the impact of remedial courses on a student's academic performance in any course within the undergraduate programme, like mathematics. A correlation can be established between the math exam score and the diagnostic placement test score.

Figure 2.2 depicts a hypothetical scenario where both variables are scaled on a range of 0 to 100. The relationship between the variables is illustrated using simulated data points.

It is assumed that there is a positive and nonlinear relationship between the variables. However, the predicted line representing math exam scores, which is generated through simulation, shows a discontinuity at the cutoff score of 50 in the diagnostic placement test. The presence of a discontinuity arises from the disparate treatment statuses assigned to students on either side of the cutoff score. In accordance with the illustration provided by Pietro (2012), students who fall below the cutoff score are assigned to remedial courses (treated), while those who above the cutoff score are not assigned to any additional courses (not treated).

It is assumed that, in the absence of the remedial courses, there would not be a discontinuity in math exam score at the diagnostic placement test score, 50. If this assumption holds, the discontinuity shown in

Figure 2.2 implies the impact of the remedial courses on math exam score, and the size of discontinuity is an ATE, which is positive here.

Figure 2.2: Discontinuity in math exam score at the cutoff score in the diagnostic placement test (the sharp RD design)

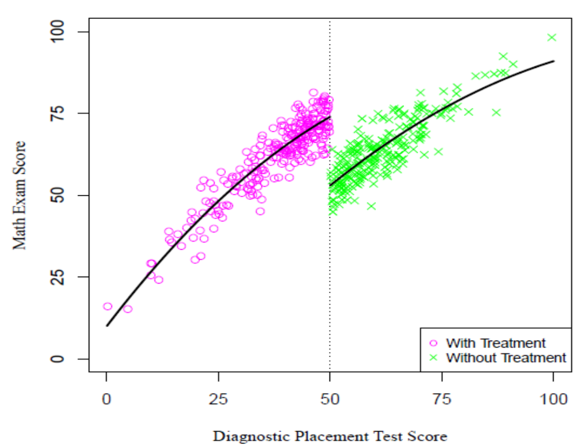


Figure 2.3: Discontinuity in math exam score at the cutoff score in the diagnostic placement test (the fuzzy RD design)

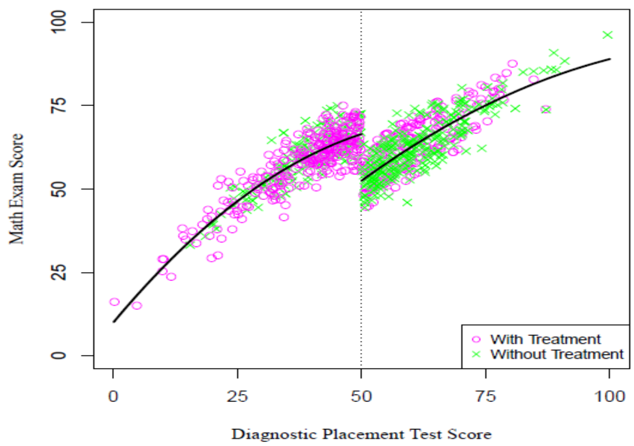


Figure 2.4: Discontinuity in the probability of treatment at the cutoff score in the diagnostic placement test (the sharp and the fuzzy RD designs)

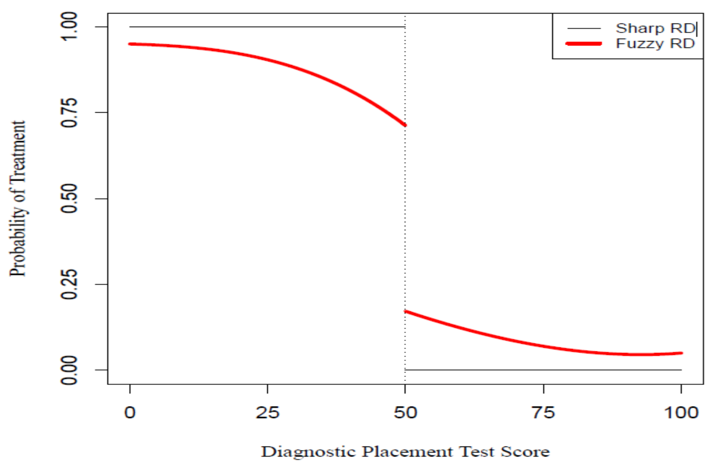


Figure 2.2 illustrates the sharp RD design, in which the diagnostic placement test score of 50 is the only factor used to determine whether students should be enrolled in remedial courses. Figure 2.3 depicts a fuzzy RD design, which is an example figure without actual data. The design assumes that, apart from the diagnostic placement test score of 50, there are several observed and unobserved factors that influence student enrollment. In the fuzzy RD, some students below the diagnostic placement test score, 50, cannot be enrolled in the remedial courses, but some students above it can be enrolled. Math exam score is still discontinuous at the diagnostic placement test score, 50, because we assume that the proportion who are treated is higher below the cutoff score in the diagnostic placement test than that above the cutoff score. This is a vital assumption in the fuzzy RD design.

Based on Pietro's (2012) example,

Figure 2.4 illustrates the correlation between the probability of treatment and the diagnostic placement test score in both the sharp RD design and the fuzzy RD design cases. It is important to note that this figure is only illustrative and does not represent actual data. In the sharp RD design case, the probability of treatment is a step function of diagnostic placement test score, and it changes abruptly from 0 to 1 at the cutoff score in the diagnostic placement test, 50.

In the fuzzy RD design case, the probability of treatment is a downward sloping curve, but discontinuous at the cutoff score in the diagnostic placement test, 50. In this case, a causal effect is estimated by dividing the size of discontinuity in math exam score shown in Figure 2.3 by the size of discontinuity in the probability of treatment shown in

Figure 2.4.

2.2.2. Formal Presentation

In the context of RD design with reference to our example, the variable referred to as the assignment variable (or forcing variable) is the diagnostic placement test score, whereas the variable referred to as the outcome variable is the math exam score. The sharp RD design and the fuzzy RD design will now be presented using the usual notation for these designs and other important variables.

2.2.3. The Sharp RD Design

Let y denotes the potential outcome (math score exam), and y_1 and y_0 denote the counterfactual outcomes with and without the treatment (the remedial courses). For a randomly drawn citizen i these are denoted by y_i, y_{i1} and y_{i0}, respectively. The treatment status is denoted by a binary variable, w_i, that contains 1 if student i is treated by the remedial courses and 0 otherwise. In the sharp RD design case, the treatment is determined as:

$$w_i = 1[x_i < c], \qquad (2.1)$$

where x_i denotes the assignment variable (age), and c denotes the cutoff score in it (in the example, $c = 50$). Along with the assignment variable, x_i, and treatment status, w_i, we observe

$y_i = (1-w_i)y_{i0} + w_i y_{i1}$. Let $\tau_i = y_{i1} - y_{i0}$ denotes the treatment effect of the remedial courses on math exam score of student i. Then

$$y_i = (1-w_i)y_{i0} + w_i y_{i1} \tag{2.2}$$

$$\begin{aligned} &= y_{i0} + (y_{i1} - y_{i0})w_i \\ &= y_{i0} + \tau_i w_i. \end{aligned} \tag{2.3}$$

In other words,

$$y_i = \begin{cases} y_{i1} = y_{i0} + \tau_i & if w_i = 1, \\ y_{i0} & if w_i = 0. \end{cases} \tag{2.4}$$

The counterfactual conditional expected outcomes are defined as $\mu_g(x_i) = E(y_{ig} \mid x_i), g = 0,1$. As in any RD design, we can believe that, in the absence of the remedial courses, students near the cutoff are similar in observed and unobserved characteristics that affect math exam score. This belief motivates the assumption:

Assumption (A1): $\mu_0(x_i)$ *is continuous at* $x_i = c$.

That is, there is no discontinuity in the conditional expected outcome without the treatment at the cutoff point, c. This assumption is an identification condition (also called the local randomisation condition) of a causal effect in a RD design. If the treatment effect for each student is heterogeneous, one more assumption needs to be satisfied to identify a causal effect. The assumption is (Hahn et al., 2001):

Assumption (A2): $E(\tau_i \mid x_i)$ *is continuous at* $x_i = c$.

Equation (2.4) implies, $E(y_{i1} \mid x_i) = E(y_{i0} \mid x_i) + E(\tau_i \mid x_i)$. Therefore, Assumptions (A1) and (A2) lead to the following:

Assumption (A3): $\mu_1(x_i)$ *is continuous at* $x_i = c$.

The ignorability assumption, also called the unconfoundedness assumption, is that w_i is independent of τ_i conditional on x_i near the cutoff. This assumption needs to be satisfied when the treatment is determined only by observed characteristics. In the sharp RD design case, w_i is deterministic at the cutoff, and therefore this assumption is automatically satisfied.

To estimate a causal effect, we cannot apply linear regression methods, as in the case of the sharp RD design, the conditional probability of treatment, defined as $Pr\left(w_i = 1 | x_i\right)$, does not vary between 0 and 1, i.e. overlapping does not hold. Here, $Pr\left(w_i = 1 | x_i\right) = 1$ for all $x_i < c$ and $Pr\left(w_i = 1 | x_i\right) = 0$ for all $x_i \geq c$.

We could estimate a causal effect by extrapolating $\mu_1\left(x_i\right)$ at $x_i \geq c$ and $\mu_0\left(x_i\right)$ at $x_i < c$. But neither parametric nor non-parametric settings for extrapolation would give convincing results. Furthermore, estimating a causal effect in this manner necessitates making substantial assumptions about the functional form, unless a constant treatment effect is anticipated.

We could estimate a causal effect by extrapolating $\mu_1\left(x_i\right)$ at $x_i \geq c$ and $\mu_0\left(x_i\right)$ at $x_i < c$. But neither parametric nor non-parametric settings for extrapolation would give convincing results. Moreover, a causal effect to be estimated in this way requires strong functional form assumptions unless a constant treatment effect is assumed.

However, a RD design focuses on the cutoff point in estimating a causal effect, defined as:

$$\tau_c \equiv E(\tau_i | x_i = c) = E(y_{i1} - y_{i0} | x_i = c) = \mu_1\left(c\right) - \mu_0\left(c\right), \qquad (2.5)$$

And this considers mainly local students around the cutoff. The external validity of the estimated causal effect is however ignored. However, there is an advantage to focus on τ_c, as its identification relies only on Assumptions (A1), (A2) and (A3).

Following Hahn et al. (2001), let $\Delta > 0$ be an arbitrarily small number. Using equation (2.3), the mean difference in observed outcomes for citizens below and above the cutoff is given by:

$$E[y_i | x_i = c - \Delta] - E[y_i | x_i = c + \Delta]$$

$$= E[y_{i0} + \tau_i w_i | x_i = c - \Delta] - E[y_{i0} + \tau_i w_i | x_i = c + \Delta]$$
$$= E[y_{i0} + \tau_i | x_i = c - \Delta] - E[y_{i0} | x_i = c + \Delta]$$

$$= E[y_{i0} \mid x_i = c - \Delta] + E[\tau_i \mid x_i = c - \Delta] - E[y_{i0} \mid x_i = c + \Delta]. \quad (2.6)$$

Given that Assumptions (A1), (A2) and (A3) are satisfied, we can take limits from the right and left hand sides of the cutoff, obtaining the following:

$$m^-(c) \equiv \lim_{\Delta \to 0} E[y_i \mid x_i = c - \Delta] = \mu_1(c), \quad (2.7)$$

$$m^+(c) \equiv \lim_{\Delta \to 0} E[y_i \mid x_i = c + \Delta] = \mu_0(c), \quad (2.8)$$

$$\lim_{\Delta \to 0} E[y_{i0} \mid x_i = c - \Delta] \equiv \lim_{\Delta \to 0} E[y_{i0} \mid x_i = c + \Delta], \quad (2.9)$$

$$\lim_{\Delta \to 0} E[\tau_i \mid x_i = c - \Delta] \equiv \tau_c. \quad (2.10)$$

Equations (2.6)-(2.10) imply, $\tau_c = m^-(c) - m^+(c)$, which is the size of discontinuity in math exam score at the cutoff score in the diagnostic placement test shown in

Figure 2.2. Although a sharp RD design gives a high weight to the local individuals near the cutoff point (Lee and Lemieux, 2010), it ultimately estimates an ATE considering all treated and untreated individuals. Here, τ_c is a parameter of an ATE in the sharp RD design. On the other hand, under heterogeneous treatment effects a fuzzy RD design estimates a Local Average Treatment Effect (LATE) for a subgroup of treated and untreated individuals, called compliers (we explain the LATE in greater detail in the following subsection).

For estimation purposes, consider the regression model based on equation (2.3):

$$y_i = \mu_0(x_i) + \tau_c w_i + e_i, \quad (2.11)$$

where $\mu_0(x_i) + e_i = y_{i0}$, and e_i is the error term, which captures the effects of observed and unobserved factors. In this equation, the treatment effect is assumed to be common at τ_c for each student. In fact, the regressor of interest, w_i, is a deterministic function of x_i and is correlated with x_i. τ_c is captured by distinguishing a discontinuous function, $1[x_i < c]$, from the smooth function, $\mu_0(x_i)$.

In equation (2.11), $\mu_0(x_i)$ can be estimated using both parametric and non-parametric regression methods. Among parametric methods, OLS is widely used in empirical studies. We can write a parametric equation in the following form (Angrist and Pischke, 2009):

$$y_i = \alpha + \beta x_i + \tau_c w_i + e_i, \qquad (2.12)$$

where $\mu_0(x_i) = \alpha + \beta x_i$ is assumed to have the same shape in both sides of the cutoff point. If one wants to allow for different trend functions for $E[y_{i0} \mid x_i]$ and $E[y_{i1} \mid x_i]$, the following models can be considered:

$$E[y_{i0} \mid x_i] = \alpha + \beta_0(x_i - c) \qquad (2.13)$$

$$E[y_{i1} \mid x_i] = \alpha + \beta_1(x_i - c) + \tau_c. \qquad (2.14)$$

Here, the normalization by subtracting c from x_i ensures that, in a regression model with interaction terms, the coefficient of w_i will be an ATE. The regression model can be derived by substituting (2.13) and (2.14) in the conditional expectation of equation (2.2):

$$\begin{aligned} &E[y_i \mid x_i] = (1 - w_i)E[y_{i0} \mid x_i] + w_i E[y_{i1} \mid x_i] \\ \Rightarrow \quad &y_i = \alpha + \beta_0(x_i - c) + \delta w_i(x_i - c) + w_i \tau_c + e_i, \end{aligned} \qquad (2.15)$$

where $\delta = (\beta_1 - \beta_0)$. If $\delta = 0$, equation (2.12) is a special case of equation (2.15). In general, empirical studies reveal that both equations produce almost similar estimates of an ATE (Angrist and Pischke, 2009).

However, we can use quadratic, cubic or in general, p order polynomials in $(x_i - c)$ to add flexibility in the functional form. It should be noted that in case of (2.15), the polynomials also need to be interacted with w_i. One needs to choose carefully the order of polynomial, p. One practice in choosing p is to plot a natural cubic spline of y_i against x_i (van der Klauuw, 2002). Indeed, any nonparametric plots may give an idea about p. If data are not noisy especially near the cutoff, scatter plots are also useful. In the case of noisy data, plotting means of y_i estimated in bins on x_i against x_i is suggested by Lee and Lemieux (2010).

Graphical presentations are informal methods in choosing *p*. Formal methods include Akaike information criterion (AIC) (Black et al.,2008), bin test (Lee and Lemieux, 2010), and generalized cross-validation procedure suggested in the literature on non-parametric series estimator (van der Klauuw, 2002). We discuss here the first two methods, as they are widely used in empirical studies. The AIC is as follows:

$$AIC = Nln(\sigma^{\hat{}2}) + 2(p+1),$$

where *N* is the number of observations, $\sigma^{\hat{}2}$ is the mean squared error of the regression and *p* is the order of polynomial taken in the regression. The polynomial order *p* will be chosen with minimum AIC. However, AIC is not an attractive method for its low power. Rather, bin test can capture nonlinearity of $\mu_0(x_i)$ well. For this test, one requires to start with the following (linear) model,

$$y_i = \alpha + \beta_0(x_i - c) + \delta w_i(x_i - c) + w_i\tau_c + \sum_{k=2}^{k-1}\varphi_k B_k + e$$

where B_k indicates *k*-th bin dummy and *K* is the number of bins. Two bin dummies are dropped due to collinearity with constant (α) and treatment dummy (w_i). Here, the null hypothesis is $\varphi_2 = \varphi_3 = \cdot\cdot\cdot = \varphi_{K-1} = 0$, which is tested to choose *p*. It is required to incorporate a higher order of polynomial into the model until the null hypothesis is not rejected. Bin test is better than AIC because it also gives an idea whether discontinuity exists at other points than the cutoff point, which can be a way to check the validity of discontinuity at the cutoff.

Among non-parametric methods, the kernel based models were initially used to estimate an ATE. However, the treatment effect estimation using these models are asymptotically biased like many other nonparametric estimators. Hahn et al. (2001) have suggested local linear regression, τ_c in the local linear regression is found by solving:

$$\min_{\alpha,\tau c,\beta,\delta} \sum_{i=1}^{N} [c-h \le x_i \le c+h](y_i - \alpha - \beta_0(x_i - c) - \tau_c w_i - \delta w_i(x_i - c)), \quad (2.16)$$

where h is a suitable bandwidth under the rectangular kernel (Imbens and Lemieux, 2008). Although other kernels (such as triangular, Epanechnikov etc.) could be used, the rectangular kernel is an optimal one in the local linear regression at the boundary point (Fan and Gijbels, 1996). Moreover, the rectangular kernel is easy to apply because there is nothing to add in the local linear regression but to choose only the bandwidth, h, around the cutoff. In choosing h, it is required to maintain both precision and bias of the local linear regression because there is a tradeoff between them. With small h the estimator becomes less biased at the cost of less precision and vice versa.

Imbens and Kalyanaraman (2009) suggested the following optimal bandwidth ($\hat{h}_{opt}$) after minimizing the asymptotic mean squared error of the average treatment effect:

$$\hat{h}_{opt} = c_k \left(\frac{2.\sigma^{\wedge 2}(c) / {}^{\wedge}l(c)}{\left(f_-^{\wedge 2}(c) - f_+^{\wedge 2}(c) \right)^2} \right)^{\frac{1}{5}} .N^{\frac{1}{5}},$$

where c_k is an index estimated by kernel, $\sigma^{\wedge 2}(c)$ is the conditional variance of y_i at c, ${}^{\wedge}l(c)$ is the marginal distribution of x_i at c, N is the sample size, and $f_-^{\wedge 2}(c)$ and $f_+^{\wedge 2}(c)$ are the second derivatives of $\mu_0(x_i)$ at the left and right limits of c. Here, $\sigma^{\wedge 2}(c)$ and ${}^{\wedge}l(c)$ are reasonably assumed to be continuous, and therefore right and left limits of them are the same. Before this method, cross-validation or *ad-hoc* methods were often applied to choose the optimal bandwidth (see Ludwig and Miller (2007) for example). However, Imbens and Kalyanaraman (2009) suggested not to use them as they consider distribution of x_i in its whole range, resulting in a biased estimate of an ATE.

2.2.4. The Fuzzy RD Design

For this RD type, the conditional probability of treatment, $Pr\left(w_i = 1 | x_i\right)$, does not change from 0 to 1 at the cutoff like the sharp RD design, because besides the cutoff point, observed and unobserved factors influence the selection for treatment. However, the crucial assumption here is that $Pr\left(w_i = 1 | x_i\right) \equiv F\left(x_i\right)$ is discontinuous at c, and the size of discontinuity remains between 0 and 1. This assumption is not an identification condition of τ_c. It ensures the existence of the fuzzy RD design, because without discontinuity in F (x_i) at c, no RD design exists.

To identify τ_c in this case, we need to apply Assumptions (A1)-(A3) as in the sharp RD design case. Under heterogeneous treatment effects, we need additional assumption(s) to be satisfied. One additional assumption is the ignorability (unconfoundedness) assumption. If this assumption is satisfied along with Assumptions (A1)-(A3), τ_c is identified (Hahn et al., 2001), van der Klauuw, 2002). The assumption is as follows:

Assumption (A4): w_i *is independent of* τ_i *conditional on* x_i *near* c : $w_i \perp \tau_i | x_i$.

Unlike the sharp RD design case, for some arbitrary small number $\Delta > 0$, equation (2.3) implies that the mean difference in observed outcomes for citizens below and above the cutoff is:

$$E[y_i | x_i = c - \Delta] - E[y_i | x_i = c + \Delta]$$

$$= E[y_{i0} | x_i = c - \Delta] + E[w_i.\tau_i | x_i = c - \Delta] - E[y_{i0} | x_i = c + \Delta] - E[w_i.\tau_i | x_i = c + \Delta]. \quad (2.17)$$

Assumption (A4) implies:

$$E[w_i.\tau_i | x_i = c \pm \Delta] = E[w_i | x_i = c \pm \Delta].E[\tau_i | x_i = c \pm \Delta]. \quad (2.18)$$

Given Assumptions (A1), (A2) and (A3), we can take limits from the right and left hand sides of the cutoff, obtaining:

$$m^-\left(c\right) \equiv limE[y_i | x_i = c - \Delta], \quad (2.19)$$

$$m^{+}\left(c\right) \equiv limE[y_i \mid x_i = c + \Delta], \tag{2.20}$$

$$F^{-}\left(c\right) \equiv limE[w_i \mid x_i = c - \Delta], \tag{2.21}$$

$$F^{+}\left(c\right) \equiv limE[w_i \mid x_i = c + \Delta], \tag{2.22}$$

$$\lim_{\Delta \to 0} E[y_{i0} \mid x_i = c - \Delta] \equiv \lim_{\Delta \to 0} E[y_{i0} \mid x_i = c + \Delta], \tag{2.23}$$

$$\lim_{\Delta \to 0} E[\tau_i \mid x_i = c - \Delta] \equiv \lim_{\Delta \to 0} E[\tau_i \mid x_i = c + \Delta] \equiv \tau_c. \tag{2.24}$$

Using equations 2.17 – 2.24,

$$\tau_c = \frac{[m^{-}\left(c\right) - m^{+}\left(c\right)]}{[F^{-}\left(c\right) - F^{+}\left(c\right)]},$$

provided that $F^{-}(c) \neq F^{+}(c)$. Here, $m^{-}(c) - m^{+}(c)$ is the size of discontinuity in math exam score at the cutoff score in the diagnostic placement test shown in Figure 2.3, and $F^{-}(c) - F^{+}(c)$ is the size of discontinuity in the probability of treatment at the cutoff score in the diagnostic placement test shown in

Figure 2.4. In the sharp RD design case, $F^{-}(c) - F^{+}(c)$ is equal to 1, and thus, the sharp RD design is a special case of the fuzzy RD design. It should be noted that if Assumption (A4) holds, a fuzzy RD design estimates an ATE considering all treated and untreated individuals. In this case, above τ_c is the parameter of an ATE in a fuzzy RD design. Assumption (A4) implies that anticipated gains from treatment do not induce individuals to be selected into treatment.

Here we can drop this assumption, and in place of it we can use another assumptions to identify τ_c. In the consequence of dropping this assumption, under heterogeneous treatment effects we eventually identify the LATE for compliers at the cutoff point (Hahn et al., 2001). As in the real world heterogeneous treatment effects are available (while constant/same treatment effect is not logical), and Assumption (A4) is unrealistic, we can say that we always identify the LATE, not ATE in the fuzzy RD design. To identify the LATE, instead of Assumption (A4) two assumptions need to be satisfied as follows (Hahn et al., 2001):

Assumption (A5): $\left(\tau_i,\ w_i\left(x_i\right)\right)$ *is jointly independent of* x_i *near* c, and

Assumption (A6): *There exists* $\epsilon > 0$ *such that* $w_i\left(c-\Delta\right) \geq w_i\left(c+\Delta\right)$ *for all* $0<\Delta<\epsilon$.

Assumption (A5) is similar to exclusion restriction assumption, which is widely used in IV regressions (it should be noted that the fuzzy RD design is an IV regression analysis). Assumption (A6) is called monotonicity or no defiers assumption. In a real world, this assumption has a logic. To understand this assumption, we need to understand the compliance behaviour of different units (Imbens and Lemieux, 2008).

Table 2.1: Compliance type by treatment and the cutoff point

w_i	$x_i \geq c$	$x_i < c$
0	complier/never-taker	never-taker/defier
1	always-taker/defier	Complier/always-taker

From the example in the previous section, we know that students above the cutoff point are not eligible for treatment. However, in the fuzzy RD design some students above the cutoff point receive treatment because of some observed and unobserved factors that affect treatment. Similarly, although students below the cutoff point are eligible, but some of them do not receive any treatment.

Table 2.1 shows that, above the cutoff point (indicated by $x_i \geq c$), treated students (indicated by w_i = 1) include always-takers and defiers, and untreated students (indicated by w_i = 0) include compliers and never-takers. On the other hand, below the cutoff point, treated students are a collection of compliers and always-takers, and untreated students are consists of never-takers and defiers. It should be noted that a student cannot fall into more than one compliance group. A student can be either complier, or never-taker, or always-taker, or defier. Table 2.2 shows the compliance type under monotonicity or no defiers assumption. This table is similar to Table 2.1, but defiers are dropped.

Table 2.2: Compliance type by treatment and the cutoff point given monotonicity

w_i	$x_i \geq c$	$x_i < c$
0	complier/never-taker	never-taker
1	always-taker	complier/always-taker

Following Imbens and Angrist (1994), Assumptions (A5) and (A6) imply,

$$\begin{aligned} & E[w_i.\tau_i \mid x_i = c-\Delta] - E[w_i.\tau_i \mid x_i = c+\Delta] \\ & = \{E[w_i \mid x_i = c-\Delta] - E[w_i \mid x_i = c+\Delta]\} \cdot \\ & E[\tau_i \mid w_i(c-\Delta) - w_i(c+\Delta) = 1]. \end{aligned} \qquad (2.25)$$

Using equations (2.17), (2.19)-(2.23), (2.25), we identify the LATE,

$$\lim_{\Delta \to 0} E[\tau_i \mid w_i(c-\Delta) - w_i(c+\Delta) = 1] = \frac{[m^-(c) - m^+(c)]}{[F^-(c) - F^+(c)]}$$

where for $\Delta > 0$ sufficiently small, the conditioning part $\{w_i(c-\Delta) - w_i(c+\Delta) = 1\}$ corresponds to compliers (a subgroup) for whom treatment changes discontinuously at the cutoff point, c.

The procedure for choosing the optimal bandwidth ($\hat{h}_{opt}$) is the same as in the sharp RD design. Local polynomial regressions can also be considered to estimate the LATE parameters.

2.2.5. Special Cases

This subsection briefly presents the pattern of RD design in three special cases, if we do not have continuous assignment variable, if there exists multiple cutoffs or, if we are to deal with panel data estimators. This subsection is based on Lee and Lemieux (2010).

Discrete Assignment Variable and Specification Errors

If the assignment variable (x_i) is discrete, e.g. 'age in year', regression line of y_i is extrapolated at one side (depending on the treatment rule) of the cutoff point. For example, cutoff age is 65 for retirement. Now, if the retirement (w_i) is determined by $w_i = 1[x_i \geq 65]$, then regression line of y_i needs extrapolation from 64 to 65. Because to estimate the treatment effect (both ATE and LATE), one needs expected outcomes of retirees and non-retirees at 65, but at 65 no observation exists for non-retirees.

Lee and Card (2008) concerned about the extrapolation. If a regression model is not run with a correct specific form in terms of the polynomial order in the assignment variable, then the extrapolation may lead to a biased estimate of the treatment effect (both ATE and LATE). Using the following goodness-of-fit test, Lee and Card (2008) suggested a natural way of testing the specification of a regression model:

$$G \equiv \frac{(ESS_R - ESS_{UR})}{(J-K), ESS_{UR} / (N-J)}$$

where ESS_R is the restricted error sum of squares from a model with a polynomial order of the assignment variable, and ESS_{UR} is the unrestricted error sum of squares from the polynomial model including a full set of dummy variables for J values of the assignment variable. Under normality and homoscedasticity of the error term, the statistic is F distributed as $F(J-K, N-J)$ where K is the polynomial order in the models and N is the number of observations. If the F value is bigger than the critical value, it indicates that the polynomial model is too restrictive, and one needs to increase the polynomial order.

Even after selecting the specific model using the test, there can be random specification errors due to group structure in the assignment variable. This does not bias RD estimates (both ATE and

LATE parameters), but may bias standard errors if they are estimated in the conventional way. Running a model on cell means or clustering the standard errors are the ways to remove bias in standard errors (Lee and Card, 2008). Both the goodness of fit test and clustered standard errors should be applied in all types of models in the sharp and the fuzzy RD designs, when the assignment variable is discrete.

More than One Cutoff Point

In this case, different amounts of treatment are determined exploiting different cutoff points. In general, RD design is fuzzy in the case of multiple cutoffs.

Let K be a number of cutoffs and c_k denotes the k^{th} cutoff, where $k = 1,2, \ldots, K$. At every cutoff point, there exists a dummy instrument $z_{ki} = 1[x_i < c_k]$. In this case, $Pr(w_i = 1 | x_i)$ is discontinuous at every cutoff point. Here, the structural form equation of y_i is the same as (2.11). However, the reduced form equation for w_i is:

$$w_i = g_0(x_i) + \pi_1 z_{1i} + \pi_2 z_{2i} + \cdots + \pi_K z_{Ki} + \xi_{1i}.$$

Substituting the above reduced form equation into the structural equation of y_i, we have the following reduced form equation of y_i:

$$y_i = k(x_i) + \tau_c \pi_1 z_{1i} + \tau_c \pi_2 z_{2i} + \cdots + \tau_c \pi_K z_{Ki} + \xi_{2i},$$

where τ_c can be estimated, at every cutoff point. The estimated value of τ_c can be different at different cutoff points. To get the LATE, one should consider the following IV equation:

$$y_i = \mu_0(x_i) + \tau_c E(w_i | x_i) + e_i,$$

where $E(w_i | x_i)$ comes from the reduced form equation of w_i. Note that, the external validity increases if the number of cutoff points increases.

Panel Data and Fixed Effects

In the case of panel data, the structural model of y_i is as follows:

$$y_i = \mu_0\left(x_{it}\right) \ + \tau_c w_{it} + a_i + e_{it},$$

where t is time, a_i is the time invariant unobserved effect, and e_{it} is the idiosyncratic error term. In the traditional case, a_i is correlated with at least any one of the covariates used in the model. Therefore, fixed effects model is run to control its effect. However, in the RD designs, this type of fixed effects model is unnecessary for identifying the treatment effect. Because in the sharp RD design case, w_{it} is unrelated to any other factors except $\left(x_{it}\right)$, and in the fuzzy RD design case, instrumental variable z_{it} is used to control for the relationship between w_{it} and any unobserved factors. An alternative way is to run simply the pooled OLS model, taking care of within-individual correlation of the errors over time, using clustered standard errors (Lee and Lemieux, 2010). Time dummies should be used like any other covariates in the model.

2.2.6. Comparing RDD and 2SLS

When properly implemented and analysed, the RDD yields an unbiased estimate of the local treatment effect. The RDD can be almost as good as a randomised experiment in measuring a treatment effect. RDD, as a quasi-experiment, does not require ex-ante randomisation and circumvents ethical issues of random assignment. Well-executed RDD studies can generate treatment effect estimates similar to estimates from randomised studies. The major benefit of using non-parametric methods in an RDD is that they provide estimates based on data closer to the cut-off, which is intuitively appealing. This reduces some bias that can result from using data farther away from the cutoff to estimate the discontinuity at the cutoff as in a 2SLS design.

Some advantages of regression discontinuity design are:

- It provides a robust method to establish causal relationships in situations where random assignment is impractical or unethical. It allows researchers to estimate randomized control trial conditions, which makes it valuable in policy evaluations and real-life situations where controlled experiments are challenging.
- The method benefits from naturally occurring cutoffs and is cost-effective and practical. It enables efficient use of available data. Additionally, it saves time and resources.

- RDD aims to ensure that observed differences in outcomes are likely due to the treatment or intervention by comparing units just above and just below the threshold. This minimizes selection bias.
- This design is versatile and applicable across various domains. It serves as a robust instrument for evaluating the effects of policies and programs in practical settings where conducting controlled experiments is challenging.

The disadvantages of regression discontinuity design are:

- RDD relies on stringent assumptions. Violating these assumptions compromises the validity of the estimated treatment effects and results in biased conclusions.
- The selection of the threshold can impact the estimated treatment effect. Additionally, manipulating the assignment variable near the cutoff by entities might distort the findings.
- Findings from this design might not always apply to different settings. The dependence on specific thresholds restricts the broad application of the findings.
- It demands a large sample on either side of the threshold for accurate estimations. If data near the cutoff is scanty, it may compromise the accuracy of treatment effect calculations and increase statistical noise in the results.

CHAPTER 3: Maternal education and child health: Quasi-experimental evidence from Zambia

3.1. Introduction

Child health is a critical measure of a developing country's quality of life. Given the critical nature of a healthy childhood, child mortality reduction has been designated as one of the Sustainable

Development Goals. In 2022, globally, 148.1 million children under the age of 5 years of age were stunted, (those too short for their age) 45.4 million wasted, (those too thin for their height) and 36.9 million were overweight (UNICEF, World Health Organisation, World Bank, 2023). In Zambia, the prevalence rate of stunting is 35 per cent. This is very high, according to the World Health Organization. A further 9 percent of children have a low birth weight of less than 2.5 kg. (Zambia Demographic Health Survey, 2018, (ZDHS)).

Child malnutrition can have long-term repercussions on an individual's life cycle. They may include cognitive impairment, increased susceptibility to chronic diseases, decreased educational attainment, and reduced productivity (Alderman and Heady, 2017). In light of the adverse individual and societal ramifications, numerous developing nations are prioritising the enhancement of child health outcomes, with maternal education being regarded as a crucial pathway.

Maternal education may directly or indirectly impact a child's health. According to Grossman (1972, 2006), educated mothers can "produce" higher child health using a predetermined combination of health inputs (productive efficiency) and can additionally, deploy health resources more efficiently than less educated mothers (allocative efficiency). Child health can be viewed as a consumer good, with parents placing a premium on it in addition to other commodities and services. Child health is produced by a vector of inputs such as parental education, healthcare use, childcare time, and dietary intake, according to Grossman (1972a,b). The health production function is shaped by a child's traits, family background, genetic makeup, community characteristics, and the health environment.

There exists a paucity of studies that employ rigorous methodologies to investigate the correlation between maternal education and child health outcomes. Furthermore, the findings from these studies are inconclusive (see Currie & Moretti, 2003; Lindeboom et al., 2009; Chou, Liu, Grossman, & Joyce, 2010; McCrary & Royer, 2011; Keats, 2018; Le and Nguyen, 2020; Wild and Stadelmann, 2020, among others).

Given that the link between maternal education and child health is undeniable, the processes by which this relationship works are unknown, making a causal link difficult to establish. The issue is mostly methodological, and it involves estimating child health production functions linking health outputs to endogenous inputs (Aslam and Kingdon, 2012). Given that increased maternal education is predicted to improve child health outcomes, it is endogenous if unobserved maternal qualities (tastes, values, and preferences) are associated with both maternal education and child health (Aslam and Kingdon, 2012). Grossman (2006, 2015) asserts that establishing a causal relationship between maternal education and child health is complicated by endogeneity. Specifically, the existence of third omitted variables that affect both maternal education and child quality, in particular, genetics and family endowments, makes the determination of causal effects difficult.

In 2002, the government of the republic of Zambia introduced the Universal Free Primary Education Policy (UFPE) in response to falling education funding and low gross enrollment rates. This education policy entailed removal of tuition and other related user fees, expanding primary education access to all Zambians (Bentaquet-Kattan, 2006). We use this policy reform, an exogenous source of variation in education as a quasi – experiment to investigate the effect of expanded primary schooling of the mother on child health outcomes and other pathways through which education improves child health outcomes.

By capitalizing on the disparities in educational attainment caused by this exogenous educational policy shift, we use a fuzzy RDD and instrumental variable (IV) framework to establish causality between maternal education and child health outcomes. We first explore the effects of maternal education on child health outcomes and then the potential mechanisms through which maternal education affects child health.

The results suggest that the education policy implemented by UFPE had a favorable effect on the educational attainment of females, regardless of their socioeconomic background. As a result of the UFPE policy, there is a discontinuous jump in 1988 for the treatment cohort (aged 14 years or younger) compared to their counterparts in the control cohort (15 years and older) at UFPE implementation in 2002. Using this exogenous source of variation in education as an instrument, we find significant causal impacts of maternal education on child health measured by three widely utilised indices namely height-for-age, weight-for-height and Weight-for-age. The findings also indicate that each additional year of schooling is associated with a reduction in the prevalence of stunting and underweight, with no effect on wasting.

3.2. Data and Empirical Strategy

3.2.1. Data Sources

This study uses data from three rounds of the Zambia Demographic and Health Surveys (ZDHS), which are nationally representative surveys of reproductive-aged women conducted in Zambia in 2007, 2013/14, and 2018. The ZDHS collects detailed health information for women of reproductive ages 15–49 and their children, including Socioeconomic and demographic variables, health care utilization, child mortality, and nutritional outcomes, etc.

Our investigation focuses on maternal education and child health outcomes. The nutritional health status of children is determined by their height-for-age z-score (HAZ) and weight-for-height z-score (WHZ) and weight-for-age z-score (WAZ). We also generate three dichotomous variables for stunting, wasting and underweight. The key variable is maternal education, defined as the total number of completed years of education. We also control for other characteristics that may affect child health outcomes such as socioeconomic status, place of residence, survey and region fixed effects and health utilization.

Table 3.1 shows the description of key variables.

Table 3.1: Description of key variables

Variable	Description
Mothers' education	Years of schooling
Child health (1)	Height-for-age z score (continuous)

	Weight-for-age z score(continuous)
	Weight-for-height z score(continuous)
Child health (2)	Height-for-age z score (binary: 1=Stunted and 0 otherwise)
	Weight-for-age z score (binary: 1=Underweight and 0 otherwise)
	Weight-for-height z score (binary: 1=Wasting and 0 otherwise)[2]

Definitions: Stunting is defined as height-for-age z-score less than – 2. Wasting is defined as defined as weight-for-height z-score less than – 2. Underweight is defined as defined as weight-for-age z-score less than – 2.

3.2.2. Empirical Strategy

The link between child health outcomes and maternal education is stated using the following naïve parametric regression model:

$$Y_i = \alpha_0 + \beta_1 Educ_i + \theta X_i + \varepsilon_i \qquad (1)$$

where Y_i is the child health outcome of $child_i$; $Educ_i$ refers to the educational level of the mother. The vector X_i is a combination of control variables. The parameter α_0 is an intercept term and ε_i is an idiosyncratic error term. Our concern is the coefficient β_1, which measures the effect of maternal education on child health outcomes. The coefficient from Equation (1) does not have a causal meaning because of the potential confounding caused by the unobserved maternal level variables that influence child health outcomes, and this will lead to skewed results. For example, less educated mothers are more likely to have unwell children as a result of inadequate investments in their own health. Further, the OLS estimates could exaggerate the impact of education because education serves as a partial substitute for omitted characteristics leading to biased estimates upward.

Due to a dearth of randomized trials examining formal education, a number of studies have attempted to address this endogeneity differently. Breierova & Duflo, (2004) using a massive

[2] Stunting is defined as height-for-age z-score less than – 2. Wasting is defined as defined as weight-for-height z-score less than – 2. Underweight is defined as defined as weight-for-age z-score less than – 2.

school construction program that took place in Indonesia between 1973 and 1978 show that both male and female education significantly reduce child mortality. Linderboom et al., (2009) and McCrary and Royer (2011) find little evidence of a direct causal impact of maternal education on child health using an exogenous variation in parental education induced by a schooling reform in the United Kingdom and United States respectively. Le and Nguyen, (2020) examine heterogeneity in mothers' educational attainment between sisters and discovers that maternal education is positively connected with child health in 68 developing nations.

Numerous efforts have been conducted in developing nations to demonstrate a causality between maternal education and child health. Within an IV regression framework, most research relies on exogenous sources of variation in educational attainment caused through access to particular education policies (school construction, abolition of basic school fees, or mandatory education laws). Grepin and Bharadwaj (2015) link female education to secondary school expansion in Zimbabwe and find a significant decrease in child mortality which they attribute to highly educated mothers. Keats (2018) uses the abolishment of primary school fees in Uganda to demonstrate that greater maternal education leads to positive changes in child health and a decrease in child malnutrition. Gunes (2015) identifies better health outcomes for children born to more educated mothers in Turkey, however Dinser et al., (2014) locate only a smattering of evidence regarding child mortality.

To mitigate the endogeneity problem, we estimate Eq. (1) using an instrumental variable (IV) framework relying on the exogenous change in education as a result of the respondent's age at the time of implementation of the 2002 education policy. We use this exogenous variation in education to address the potential endogeneity of education. We use the fact that the respondent's age at the time of the policy's enactment altered the likelihood of exposure and hence benefiting from it. Those aged 14 and under had a better opportunity of benefiting from the UFPE strategy.

We utilize the individual's birth year rather than the year they first entered primary school as the forcing variable to create treatment and control groups since the timing of enrollment into primary school is influenced by numerous circumstances, including unseen variables that affect both school tenure and the results (Adu-Boahen and Yamauchi, 2018).

Due to Zambia's high rate of grade repetition, it is improbable that treatment assignment will be a deterministic function of it. Late school admission, long absences, early marriages, and job market participation are frequent among non-compliers. The fuzzy RDD is used to measure the impact of education in the presence of such noncompliance as opposed to a sharp RDD.

The estimand in a fuzzy RDD is similar to an IV estimate and is equivalent to the 2SLS estimator in an IV framework: The estimand (τ_{FRD}) is defined by the "ratio of the reduced form estimate of the discontinuity in the outcome variable Y_i to the reduced form estimate of the discontinuity in the treatment variable (education) both determined by the assignment or forcing variable X_i (year of birth) and evaluated at the cut-off c" (Hahn et al, 2001; Imbens and Lemieux 2008; Lee and Lemieux 2010) as follows:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} \mathrm{E}[Y \mid x] - \lim_{x \uparrow c} \mathrm{E}[Y \mid x]}{\lim_{x \downarrow c} \mathrm{E}[education \mid x] - \lim_{x \uparrow c} \mathrm{E}[education \mid x]}$$

As a result, estimation is divided into two stages: the first stage and the second stage. (1) a first-stage, in which the treatment, $education_i$ is regressed on the identifying instrument Z_i, whereby $education_i$ is as defined and Z_i is the dichotomous indicator $Z_i = 1[Year\ of\ Birth \geq c(1988)]$. In the second stage (2), we regress the dependent variable Y_i on the predicted value of education, $(\widehat{education}_i,)$ estimated in the first stage as follows:

$$Education_i = \alpha_0 + \alpha_1 Z_i + f(X_i - c) + \alpha_k \Theta_k + e_i \qquad (1)$$

$$Y_i = \beta_0 + \beta_{1RD} education_i + g(X_i - c) + \alpha_k \Theta_k + u_i \qquad (2)$$

The functions $f(.)$ and $g(.)$ are polynomials. The order of the polynomial and bandwidth have to be the same for both first stage and second stage regressions for calculation of standard errors as in a 2SLS framework. In our equations, $\Theta\ k$ are control variables and e_i, u_i are error terms. β_{1RD} is the causal effect of education on child health outcomes and reflects the "local average treatment effect (LATE) on compliers" (Angrist et al, 1996; Lee and Lemieux 2010; Wild and Stadelmann 2020).

3.3. Results

3.3.1. Data and Descriptive Statistics

The descriptive statistics are shown in Table 3.2 for the control and treatment cohorts. **Error! Reference source not found.** shows the figures for the full sample as well as the poor and wealthy sample which are a result of splitting the household wealth into two sub – samples, the "poor" and "wealthy" subsamples.[3]

***Table 3.2*: Descriptive Statistics**

	Full Sample		**Poor**		**Wealthy**	
	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)
Panel A: Characteristics of the mothers						
Number of observations	8248	6909	5502	4966	2743	2173
Years of education	5.87	6.73	4.39	5.65	8.81	9.13
	[3.89]	[3.49]	[3.08]	[3.04]	[3.69]	[3.17]
Age at survey	28.27	24.39	28.39	24.71	28.02	24.12
	[4.2]	[2.91]	[4.21]	[3.08]	[4.22]	[3.26]
Age at first birth	18.74	18.24	18.21	17.85	19.69	18.89
	[3.02]	[2.47]	[2.65]	[2.2]	[3.45]	[2.85]
Residence (Rural = 1)	0.63	0.64	0.85	0.84	0.18	0.17
Wealth Index combined (1-5)	2.8 [1.28]	2.67 [1.37]	1.97 [0.81]	1.91 [0.81]	4.45 [0.49]	4.44 [0.5]
Worked in last 12 months (Yes =1)	0.62	0.55	0.65	0.58	0.56	0.44
Child Health Outcomes						
Panel B: Children Aged 3 – 59 Months						
Number of observations	6999	5866	2294	4266	2295	1772
	-1.56	-1.55	-1.31	1.65	-1.31	-1.31

[3] Respondents from families scoring at or below the median wealth score are considered to be "poor," and respondents with a wealth level above the median are considered to be "wealthy."

Height-for-age z-score (HAZ)	[1.61]	[1.52]	[1.59]	[1.52]	[1.59]	[1.57]
Weight-for-age z-score (WAZ)	-0.85	-0.86	-0.67	-0.92	-0.67	-0.69
	[1.11]	[1.1]	[1.1]	[1.1]	[1.1]	[1.1]
Weight-for-height	0.06	-0.86	0.08	0.02	0.08	0.06
z-score (WHZ)	[1.78]	[1.11]	[1.33]	[1.25]	[1.33]	[1.29]
Stunting = 1	0.39	0.38	0.26	0.35	0.26	0.26
Underweight =1	0.14	0.14	0.09	0.13	0.09	0.09
Wasting	0.05	0.05	0.05	0.04	0.05	0.04

Notes: The study employs data from the 2007, 2013/2014, and 2018 Zambia DHS Surveys. The analysis utilizes a bandwidth size of 6 birth year cohorts on both sides of the cut-off. The survey participants are limited to individuals who are 19 years of age or older. The subsamples denoted as 'poor' and 'wealthy' comprise of female individuals whose wealth levels are either below or above the median, respectively. The statistical analysis incorporates the DHS sample weights and takes into consideration the pooling effect across survey rounds.

3.3.2. Bandwidth Selection

Following the recommendations by Abadie et al. (2017) and Nunn (2010) standard errors are clustered at the level of survey clusters. This approach to clustering yields standard errors that are robust against arbitrary patterns of within-cluster variation and covariation (Cameron and Miller, 2015). The standard errors are clustered at the survey cluster level since this is the level at which units in the sample were selected and there are clusters in the population that are not represented in the sample. Clustering at survey clusters is useful because while respondents are identical across cohorts, their characteristics may differ by location; in the context of this study, social and economic status can vary much within and between clusters. A section of the literature also clusters at the level of running variables. The point estimates remain robust to clustering at the level of the running variable. The strength of the instrument, however, becomes sensitive to alternative bandwidths in this approach.

Recent academic literature on the regression discontinuity design supports the preference for employing local linear regression estimation instead of global polynomial estimation. We perform an estimation of a local linear regression model on both sides of the threshold, allowing for variation in the slopes on either side of the threshold (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). By employing the 2SLS method, we estimate the Local Average

Treatment Effect at the Cutoff (LATE) specifically for individuals who comply with the treatment assignment.

In order to determine the most suitable bandwidth, we employ the mean squared error (MSE-optimal) bandwidth method, utilizing the local polynomial bandwidth selection method proposed by Cattaneo et al. (2014; 2017). The proposed methodology yields an optimal bandwidth of 5.74 years on both sides of the cutoff. Hence, all estimations in the first and second stage regressions were performed with a bandwidth of 6 years on both sides of the cutoff.

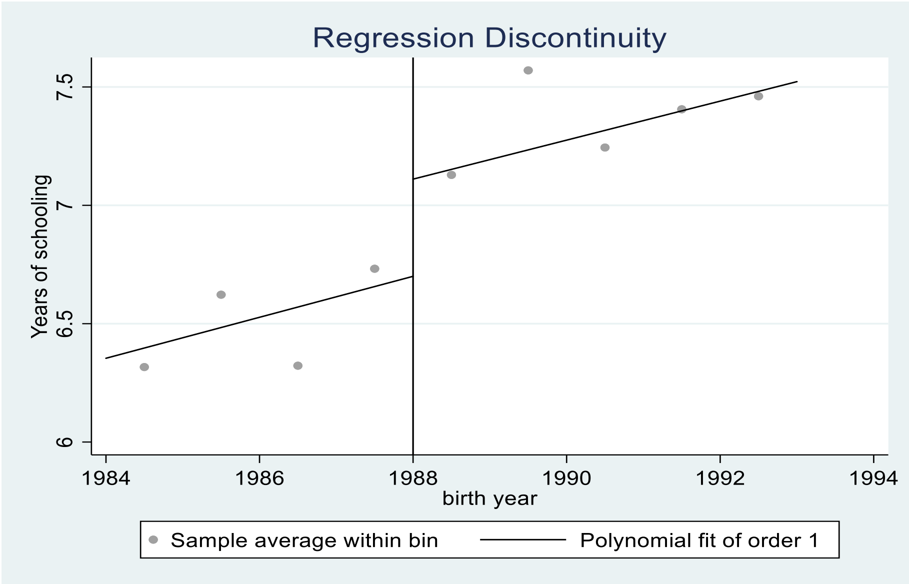
3.3.3. Graphical representation

The impact of the UFPE policy on women educational attainment is captured graphically in

Figure ***3.1*** showing the relationship between years of schooling for women and year of birth. The perpendicular line in the middle represents the 1988 birth year cutoff. To the right (left) of the cutoff lies the treatment (control) groups. The points in the graph reflect the average highest years of education by birth cohort year. The solid line depicts the values that were fitted using local linear regression. The figure shows noticeable discontinuities in the women's education. The surge in the number of years spent in schooling depicts the exogenous surge in primary year schooling and is the main motivation behind this study. In contrast to the years of formal schooling by birth year cohort for the mothers, Figure 3.2 and

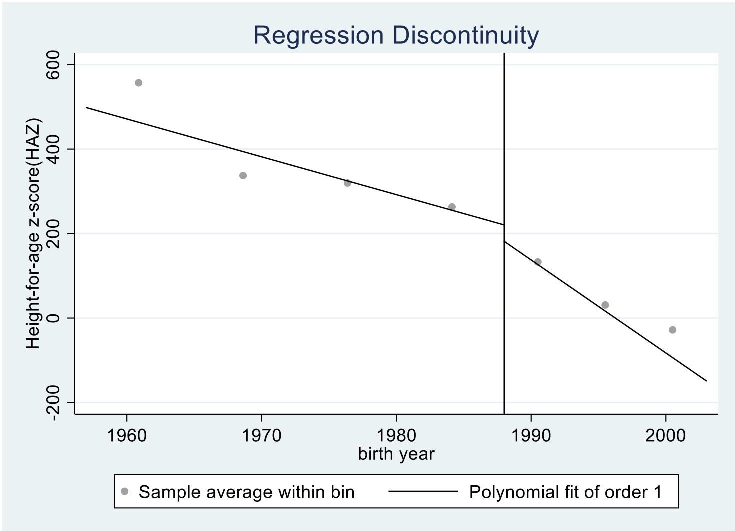
Figure ***3.3*** show that there are no discernible discontinuities in nutritional status for the children as measured by height for age and weight for age z-scores. The weight for height z-score however shows some discontinuity in Figure 3.4. However the effect of mother's education on the weight for height z-score for children is however insignificant in this thesis.

Figure 3.1: Women's Schooling by Birth Year Cohort



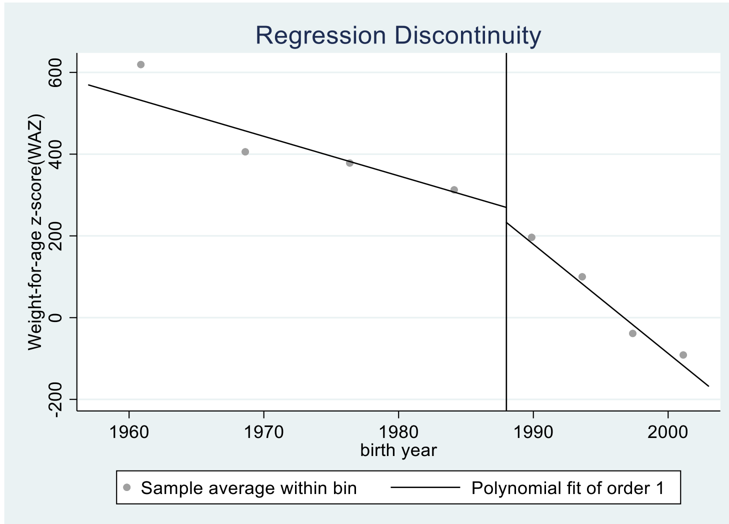
Notes: data from the 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The survey participants are limited to individuals who are 19 years of age or older.

Figure 3.2: Height for age z-score by birth year cohort



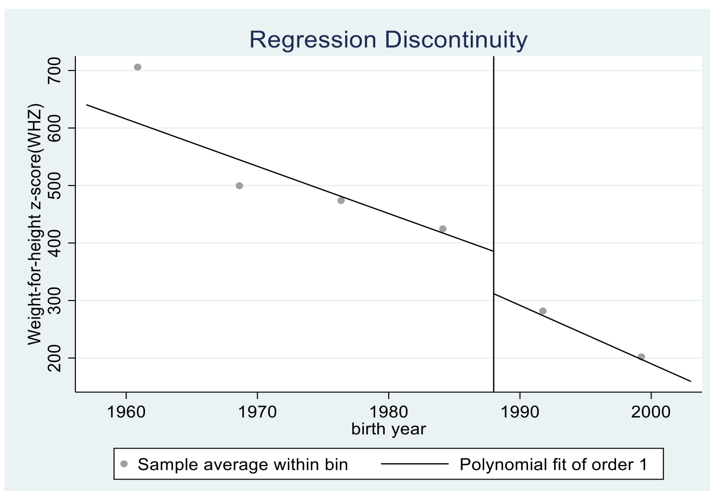
Notes: data from the 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The survey participants are limited to individuals who are 19 years of age or older.

Figure 3.3: Weight for age z-score by birth year cohort



Notes: data from the 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The survey participants are limited to individuals who are 19 years of age or older.

Figure 3.4: Weight for height z-score by birth year cohort



Notes: data from the 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The survey participants are limited to individuals who are 19 years of age or older.

3.3.4. First Stage Results – Effects of the UFPE policy on Education

The results of the first stage results i.e. the effect of the UFPE policy on maternal education are shown in Table 3.3. For its part the UFPE's objective was to increase enrollment, consequently reducing dropout rates thereby increasing the total average years of schooling completed. These first stage results confirm the success of these objectives. Table 3.3 displays the first-stage discontinuity indicated by the regression equation (1). The results are provided individually for the entire sample (columns 1 and 2), the poor subsample (columns 3 and 4) and the wealthy subsample (columns 5 and 6).

In discussing the results, we refer to the stricter specification which includes survey fixed effects and all controls. The results of the first stage show that for women who were assigned to treatment or benefited from the UFPE policy, the average effect of the policy was 0.68 more years of education compared to the their counterparts in the control cohort who were not assigned to treatment. The effect is economically large and statistically significant in the full sample and also in the poor subsample. The first stage F-statistic is 41.05[4]

***Table 3.3*: First Stage – Effect of the UFPE Policy on Maternal Education**

	Years of Schooling					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
1[Birth year ≥ 1988]	0.708***	0.683***	0.884***	0.862***	0.44*	0.4**
	(0.141)	(0.107)	(0.133)	(0.126)	(0.2444)	(0.193)
Mean contr. group	[5.88]		[3.61]		[7.36]	
Interaction: 1*Year of	0.07**	0.108***	0.157***	0.162***	-0.056	-0.007
Birth	(0.036)	(0.031)	(0.037)	(0.036)	(0.0719)	(0.9)
Intercept	5.919***	4.34***	4.368***	4.54***	8.673***	7.52***
	(0.075)	(0.22)	(0.069)	(0.232)	(0.125)	(0.347)
First Stage F-Statistic	25.34	41.05	43.97	46.96	3.25	4.3
Observations	15148	15123	10337	10337	4811	4811
Rsquared	0.015	0.36	0.046	0.138	0.002	0.188
Controls		Yes		Yes		Yes
Region Fixed Effects		Yes		Yes		Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

*Note: for direct interpretation of the coefficient on the discontinuity, the birth year distribution has been re-centered at the discontinuity year (1988). Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff. The sample is restricted to individuals 19 years and older at the time of the survey. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. Standard errors clustered at the survey level are presented in parentheses. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

[4] The first stage *F*-statistic exceeds 10, suggesting the instrument is not weak (Staiger and Stock, 1997).

3.3.5. Second-Stage Results

Having shown that the UFPE education policy had a substantially large economic effect on the educational attainment of women, we now present the second stage results. The results of the second stage which are the main findings in this paper, are presented in Table 3.4. The DHS surveys collects child health data for children under the age of five. We use the height for – age and weight for age z scores as dependent variables which are measured in standard deviation are available for all surveys. Further from these variables, we present results for stunting and underweight. Low height- for-age based on the WHO, indicates under nutrition or less than perfect health, whereas low weight-for-age is a product of height-for-age and weight-for-height scores for the child. The results are presented as follows: Columns 1, 3, 5 and 7 present results with survey fixed effects only. Columns 2, 4, 6 and 8 include survey fixed effects and also controls for region fixed effects and also a dummy for whether the mother lives in rural or urban areas. We report results for the stricter of the specifications which include controls, region and survey fixed effects.

We find large and statistically significant impacts of maternal years of schooling on height-for -age z-scores. For the IV results, Table 3.4 shows that an extra year of mother's schooling increases children's height-for-age by 0.11 standard deviations. The results equally show that an additional year of schooling by the mother decreases incidences of stunting by about 4 percentage points. These are similar estimates to those found by Chen and Li (2009), Hahn et al., (2018) and Nepal (2018). For the poor subsample, the results show that the effect of an extra year of mother's education is an increase in height-for-age z-score by 0.09 standard deviations. This is slightly lower than for the full sample. In the poor subsample, equally, chances of stunting decrease by about 4 percentage points. For the weight-for-age z-scores, the IV estimates show that controlling for age, children of women with additional schooling have 0.084 standard deviations more weight.

Equally, the incidences of being underweight decrease by 2 percentage points for children whose mothers have additional years of schooling. Our results equally show that each extra year of the mother's schooling increases the weight-for-height z-score of the child by 0.02 standard deviation. There is no effect on wasting in the full sample which signifies acute malnutrition episodes. In the poor subsample however, an additional years of schooling decreases the probability of wasting by 0.7 percentage points. These results are similar to the one by Keats (2018) and Gunes (2015) who find no significant effects of additional schooling on weight-for-height z-scores and wasting.

Comparative OLS also displayed in Table 3.4, columns 3 and 4. In all our comparisons, we refer to the stricter column 4 with controls. When compared with the OLS estimates, the IV estimates are larger and twice in magnitude compared to the OLS estimates. The OLS estimate for example shows that an additional year of mother's schooling, holding the age of the child constant, increases the height of the child by 0.052 standard deviation compared to 0.11 for the IV 2SLS estimate. This is consistent with many studies. The possible explanations for this trend include a downward bias of the OLS estimate brought about by variable omission bias or measurement error. Of particular concern are family characteristic variables such as ability which affect maternal schooling positively may bias the OLS estimates downwards.

Overall, the IV results do show that the mother's education has a positive and significant effect on the child's health. Specifically, the results show significant effects of additional mother's schooling on height-for-age z and weight-for-age z-scores. Further, incidences of stunting and being underweight significantly reduce as a result of the mother's schooling. On the contrary and in line with other papers, there are no significant effects of additional maternal schooling on weight for height z-scores and wasting.

Finally, we also explore the heterogeneous effects of education on type of place of residence, religion and the region of residence. The results are shown in Table A1. Consistent with the main results, an extra year of schooling increases height-for-age and weight for age z-scores for children of women living a rural area compared to those living in the urban area. We find no significant effects of maternal education on child outcomes for women in rural areas compared to those in urban areas. Equally, results show that additional maternal education decreases the likelihood of stunting and being underweight for children of women in rural areas compared to those in urban areas. We find scanty evidence of the effects of additional education on outcomes based on region and religion.

Table 3.4: Second Stage Results: IV Estimates

	Full Sample		OLS Sample		Poor Subsample		Wealthy Subsample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Controls		Yes		Yes		Yes		Yes
Region Fixed Effects		Yes		Yes		Yes		Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel A: Height-for - age z-score								
Schooling instrumented by: 1[Birth year ≥ 1988]	0.099***	0.114***	0.0522***	0.043***	0.098***	0.091**	0.126***	0.134***
	(0.009)	(0.012)	(0.005)	(0.005)	(0.021)	(0.029)	(0.023)	(0.024)
Mean of dep. var.	[6.23]		[6.23]		[4.99]		[8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.019	0.025	0.015	0.0221	0.004	0.013	0.14	0.019
Panel B: Stunting								
Schooling instrumented by: 1[Birth year ≥ 1988]	-0.03***	-0.037***	-0.0148***	-0.008***	-0.043***	-0.041***	-0.04***	-0.041***
	(0.003)	(0.004)	(0.014)	(0.002)	(0.007)	(0.009)	(0.006)	(0.007)
Mean of dep. var.	[6.23]		[6.23]		[4.99]		[8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.018	0.027	0.013	0.028	0.004	0.019	0.016	0.021
Panel C: Weight-for - age z-score								
Schooling instrumented by: 1[Birth year ≥ 1988]	0.066***	0.084***	0.0345***	0.02***	0.061***	0.099***	0.087***	0.095***
	(0.007)	(0.009)	(0.004)	(0.004)	(0.016)	(0.021)	(0.016)	(0.017)
Mean of dep. var.	[6.23]		[6.23]		[4.99]		8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.017	0.026	0.013	0.028	0.003	0.015	0.014	0.022

Panel D: Underweight								
Schooling instrumented by: 1[Birth year ≥ 1988]	-0.014***	-0.019***	-0.008***	-0.005***	-0.019***	-0.035***	-0.012***	-0.0144***
	(0.0016)	(0.005)	(0.001)	(0.001)	(0.005)	(0.01)	(0.004)	(0.004)
Mean of dep. var.	[6.23]		[6.23		[4.99]		[8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.008	0.014	0.007	0.017	0.003	0.013	0.003	0.006
Panel E: Weight-for - height z-score								
Schooling instrumented by: 1[Birth year ≥ 1988]	0.012	0.022*	0.004	0.001	-0.005	0.05*	0.022	0.027
	(0.008)	(0.011)	(0.004)	(0.005)	(0.02)	(0.025)	(0.02)	(0.021)
Mean of dep. var.	[6.23]		[6.23]		[4.99]		[8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.001	0.009	0.0002	0.009	0.001	0.034	0.008	0.007
Panel F: Wasting								
Schooling instrumented by: 1[Birth year ≥ 1988]	-0.006	-0.001	-0.0003	-0.0001	-0.007*	-0.008*	-0.003	-0.002
	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.003)
Mean of dep. var.	[6.23]		[6.23]		[4.99]		[8.96]	
Observations	12844		12844		8851		3993	
Rsquared	0.0001	0.006	0.0	0.01	0.001	0.01	0.0005	0.004

Notes: Each column's results are generated through an individual regression. The main explanatory variable (schooling) is instrumented by the binary indicator 1[Birth year≥ 1988]. *Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff. The sample is restricted to individuals 19 years and older at the time of the survey. Robust standard errors clustered at the survey level are presented in parentheses. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

3.4. Discussion

This first stage effect is similar to that of Keats (2018) and Makate (2016) who used the Ugandan DHS and Hahn et al., (2018) who used the Bangladeshi DHS. In the case of Uganda, several studies have examined the impact of the Universal Primary Education (UPE) policy. Keats (2018), Andriano and Monden (2019), and Nagashima and Yamauchi (2020) have reported findings on the effect of this policy on the duration of schooling. Specifically, Keats (2018) found that the UPE policy increased years of schooling by approximately 0.71 years, while Andriano and Monden (2019) observed an increase of 1.5 years. Similarly, Nagashima and Yamauchi (2020) reported a rise of 0.3 years in the duration of schooling as a result of the UPE policy. In the Nigerian context, Osili and Long (2008), Larreguy and Marshall (2017), and De Cao and La Mattina (2019) conducted studies that revealed the impact of the Free Primary Education (FPE) policy on educational attainment. Specifically, their findings indicated that the FPE programme resulted in an increase of 1.5, 0.6, and 2.1 years of schooling, respectively. According to Adu-Boahen and Yamauchi (2018), the implementation of the FCUBE policy in Ghana in 1996 resulted in an increase of one year in the duration of schooling. In general, the magnitude of the effect is consistent with the magnitudes observed in previous instances.

The magnitude of the effects increases as the bandwidths become narrower, while it decreases as the bandwidths become broader, as anticipated. The effect of the policy remains largely unchanged even after adding predetermined covariates. Equally, the mean of the control cohort group differs substantially between the poor and wealthy subgroups. Women from the poor subgroups without exposure to the UFPE educational policy had 3.6 years of education while their counterparts from the wealthy subgroups had more than double years of education at 7.4 years. Therefore, although the UFPE policy added the same number of years of education for both the poor and wealthy subsamples as the policy was implemented at a national level, it more than doubled the amount of education for the poor relatively. In summary, the first-stage findings suggest that the UFPE educational policy resulted in a large and substantial increase in female educational attainment.

3.4.1. Maternal Education improves child health

In accordance with previous research, the findings presented in Table 3.4 demonstrate a positive causal relationship between maternal education and indicators of child health, specifically height-for-age, weight-for-age, and weight-for-height z-scores. The full sample findings presented in Table 3.4 column 2 demonstrate that an increase in maternal schooling by one year has a positive impact on various nutritional indicators for children such as height-for-age and weight-for-age z scores, and reduces incidences of child stunting and being underweight. We find positive, large impacts of maternal years of schooling on height-for-age and weight-for-age z-scores. An extra year of mother's schooling increases children's height-for-age by 0.11 standard deviations. For the weight-for-age z-scores, the IV estimates show that controlling for age, children of women with additional schooling have 0.084 standard deviations more weight. On the contrary, we find weak evidence of additional schooling on weight-for-height z-scores.

The observed advantages of maternal education on child health in our study align with findings from various previous studies (Chen and Li, 2009; Hahn et al., 2018; Nepal, 2018; Le and Nguyen, 2020; Aldeman and Headey; 2017; Keats, 2018). The findings of this study are in opposition to the findings reported by Ali and Elsayed (2017), Lindeboom et al., (2009) and McCrary and Royer (2011). Their findings indicate that there is no significant impact of maternal education on child nutritional status. Our estimates further exhibit a lesser magnitude in comparison to the findings reported by Gunes (2015) in the context of Turkey.[5]

3.4.2. Potential Mechanisms

Numerous studies have attempted to uncover more direct paths via which maternal education can benefit child health. In Brazil, Thomas et al., (1991) investigated the role of income, mother's literacy and information processing, and the interaction of maternal schooling with community services. The authors conclude that nearly all of the relationship between maternal education and child height may be explained by the mother's access to knowledge (i.e., exposure to media).

More educated women do marry more educated men, in what is known as positive assortative mating. (Becker, 1973; Behrman & Rosenzweig, 2002). As a result of this union between two educated parents, children's health could substantially improve as opposed to a situation when the

[5] Our results are less compared to those found by Gunes (2015) for Turkey because in their paper, they use as a measurement of education, primary school completion. In most developing countries like Zambia, the rate of primary school completion is high and the returns to education falls by the level of schooling (Colclough 1982; Psacharopoulos 1985, 1994; Schultz 1988).

parents are less educated. Further, a high quality husband could be a mechanism via which mother's education positively affects child's health.

According to Glewwe (1999) and Thomas et al., (1991), information access, in particular access to information about family planning is one potential pathway through which mother's education influences child health. To begin, education gives women the ability to digest information more effectively, in particular the ability to learn and implement new approaches to child care. Furthermore, women who have received a higher level of education are more inclined to possess superior health literacy which enables them to "produce" children with improved health. (Aslam and Kingdon, 2012; Dinser et al., 2014; Le and Nguyen, 2020).

Working women have higher earnings than their non-working counterparts, allowing them to devote more resources to the welfare of their children. This work engagement coupled with the earnings in turn signifies a potential pathway through which the positive effects of education can be transmitted to the child.

This paper explores the potential channels via which mother's schooling affects child health. The potential channels are divided into four groups: (a) Fertility behaviour, (b) Husband's education, (c) Information access and health knowledge, (d) labour market outcomes.

Fertility Behaviour

To capture fertility behaviour, we use number of children born, age at first marriage, age at first birth and age at first sexual intercourse. The results are presented in Table 3.5. As shown, an extra year of maternal education is associated with a decrease in the number of children born by 0.23 births. Further, an extra year of mother's schooling increases her age at first birth by 0.35 years and well as her age at first sexual intercourse by 0.31 years. Equally, age at first marriage increases by about 0.42 years as a result of an additional year of schooling. Given this, a conclusion can be made that the decrease in the mother's fertility as well as her increase in the age at first marriage, the age at first birth and first sexual intercourse are responsible for the health of the child.

Husband characteristics

We measure husband's education as the total number of years of accumulated education. Results show that increasing mother's education leads to an increase in the husband's educational attainment by 0.77 years. Therefore positive assortative mating could be one way that the mother's education improves child health.

Information access and health knowledge

To measure information access, we proxy it using dummy variables for literacy, knowledge of ovulation, use of contraceptives. Further we create dummy variables for whether one frequently watches television, reads newspapers and listens to radio. The results show that an additional year of schooling increases the likelihood of listening to radio and watching TV by 7.1 and 13.3 percentage points respectively. Further, an extra year of education increases the likelihood of reading newspaper by 33 percent (An increase of 7.2 percentage points off a base of 22 percent). As part of health knowledge, an additional year of education increases the likelihood of contraceptive use and knowledge of ovulation cycle by 0.67 and 3.6 percent (an increase of 0.6 and 1.6 percentage points off a base of 89 and 44 percent respectively).

The use of contraceptives and delay in age at marriage together cancel out a possible "incarceration effect" and point to the fact that the extra education attained provides the women with the ability and desire to lessen pregnancies without necessarily using abstinence as an option. Our results show that literacy, knowledge of ovulation cycle, media information and use of contraceptives are critical in channeling the benefits of maternal education on child health.

Labour market outcomes

Surprisingly, there are no effects of maternal education on labour force participation on both the poor and wealthy subsample. Insufficient empirical support exists to establish a robust correlation between additional maternal education and labour market outcomes in Zambia, potentially attributable to the presence of structural labour market limitations and suboptimal levels of labour force engagement.

Table 3.5: Potential Mechanisms – Fertility related outcomes

	Full Sample	Poor	Wealthy
	(1)	(2)	(3)
Age at first Marriage			
Schooling instrumented by:	0.424***	0.082	0.593***
1[Birth year ≥ 1988]	(0.035)	(0.188)	(0.061)
Mean of dependent var.	18.21	17.51	19.17
Observations	13938	9703	4235
Rsquared	0.11	0.043	0.086
Number of children born			
Schooling instrumented by:	- 0.21***	-0.327***	-0.181***
1[Birth year ≥ 1988]	(0.015)	(0.036)	(0.024)
Mean of dependent var.	3.81	3.6	2.66
Observations	15132	10323	4809
Rsquared	0.22	0.19	0.15
Age at first birth			
Schooling instrumented by:	0.348***	-0.007	0.516***
1[Birth year ≥ 1988]	(0.028)	(0.051)	(0.053)
Mean of dependent var.	18.74	18.1	19.4
Observations	15132	10323	4809
Rsquared	0.086	0.016	0.079
Age at first sexual intercourse			
Schooling instrumented by:	0.314***	0.184***	0.417***
1[Birth year ≥ 1988]	(0.025)	(0.055)	(0.043)
Mean of dependent var.	16.56	16.01	17.35
Observations	11817	7721	4809
Rsquared	0.098	0.025	0.074
Husband education			
Schooling instrumented by:	0.767***	0.992***	0.848***
1[Birth year ≥ 1988]	(0.147)	(0.072)	(0.055)
Mean of dependent var.	7.6	6.51	10.54
Observations	13271	8797	3985
Rsquared	0.033	0.11	0.15
Currently working			
Schooling instrumented by:	-0.004	-0.009	0.002
1[Birth year ≥ 1988]	(0.005)	(0.458)	(0.008)
Mean of dependent var.	0.57	0.55	0.47
Observations	15104	10323	4809
Rsquared	0.057	0.067	0.036
Literacy			
Schooling instrumented by:	0.084***	0.126***	0.063***
1[Birth year ≥ 1988]	(0.004)	(0.01)	(0.006)
Mean of dependent var.	0.44	0.45	0.81
Observations	15132	10323	4809
Rsquared	0.18	0.09	0.07

Knowledge of ovulation			
Schooling instrumented by:	0.006*	0.011	-0.001
1[Birth year ≥ 1988]	(0.003)	(0.006)	(0.004)
Mean of dependent var.	0.89	0.88	0.91
Observations	15132	10323	4809
Rsquared	0.011	0.011	0.01
Watches Television			
Schooling instrumented by:	0.133***	0.024***	-0.017***
1[Birth year ≥ 1988]	(0.003)	(0.005)	(0.003)
Mean of dependent var.	0.32	0.048	0.051
Observations	15132	10323	4809
Rsquared	0.47	0.02	0.03
Reads Newspaper			
Schooling instrumented by:	0.072***	0.053***	0.082***
1[Birth year ≥ 1988]	(0.005)	(0.008)	(0.008)
Mean of dependent var.	0.58	0.014	0.41
Observations	15132	10323	4809
Rsquared	0.13	0.037	0.07
Listen to Radio			
Schooling instrumented by:	0.071***	0.128***	0.042***
1[Birth year ≥ 1988]	(0.005)	(0.011)	(0.008)
Mean of dependent var.	0.58	0.47	0.73
Observations	15132	10323	4809
Rsquared	0.089	0.053	0.039
Contraceptive Use			
Schooling instrumented by:	0.016***	0.036***	0.012
1[Birth year ≥ 1988]	(0.004)	(0.011)	(0.008)
Mean of dependent var.	0.44	0.4	0.55
Observations	15132	10323	4809
Rsquared	0.044	0.043	0.012

*Notes: All regressions include controls for region, religion and a rural dummy. Robust standard errors clustered at the survey level are presented in parentheses. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

3.4.3. Heterogeneity Effects: Poor versus Wealthy subsample results

We also explore the heterogeneous effects of maternal education on the potential mechanisms through which maternal education could affect child health outcomes. The results for the poor and wealthy subsamples are displayed alongside the full sample results in Table 3.5. The results show that there are no significant effects of additional schooling on age at first marriage in the poor subsample as opposed to the wealthy subsample which shows a significant increase of about 0.6

years. The decrease in total fertility for the poor subsample is larger than for the wealthy subsample by about 0.15 births. This can be explained by the already high fertility prevalent in the poor subsample compared to the wealthy subsample.

The results show no significant effects of extra years of schooling on age at first sex in the poor subsample. The affects are all driven by the wealthy subsample. As expected, the effect of maternal education on the age at first sexual intercourse is higher for the wealthy subsample as opposed to the poor subsample. Surprisingly, as with the full sample, there are no effects of maternal education on labour force participation. The impact of maternal education on literacy on the poor subsample is twice that on the wealthy subsample. We find no effect of maternal education on the knowledge of the ovulation cycle in both subsamples.

Regarding the effect of additional schooling on knowledge acquisition through access to media, we find that each extra year of education increases the likelihood of watching television in the poor subsample but decreases this likelihood in the wealthy subsample. Each extra year of schooling causes more people in the wealthy subsample to read newspapers than in the poor subsample. Interestingly, the effect of additional schooling on listening to the ratio is twice higher in the poor subsample than the wealthy subsample. The poor households in the rural areas listen more to the radio than the wealthy subsample, mostly in the urban areas. We find no effects of additional female schooling on use of contraceptives in the wealthy subsample. All the effects are being driven by poor subsample.

Finally, we also explore the heterogeneous effects of education on type of place of residence, religion and the region of residence. The results are shown in Table A1. Consistent with the main results, an extra year of schooling increases height-for-age and weight for age z-scores for children of women living a rural area compared to those living in the urban area. We find no significant effects of maternal education on child outcomes for women in rural areas compared to those in urban areas. Equally, results show that additional maternal education decreases the likelihood of stunting and being underweight for children of women in rural areas compared to those in urban areas. We find scanty evidence of the effects of additional education on outcomes based on region and religion.

3.5. Validity and Robustness Checks

3.5.1. Internal Validity

For us to use the UFPE policy as a quasi-experiment for purposes of obtaining policy estimates discussed in the previous results section, a few concerns have to be addressed. Am important concern is that parents may have advance information about the year and timing of the UFPE policy and therefore use this information to postpone the enrolment of their children into primary school so as to benefit from the policy. However, this is not likely to be the case as the UFPE policy was a deliberate national policy and was to be implemented at once countrywide with the aim of providing everyone with free primary education regardless of status or grade level. Further, if parents had postponed the enrolment of their children at entry level primary grade one, the gross enrolment ratio shown in would have declined around the 2000 and 2001 year close to the UFPE policy implementation but this is not the case.

The other concern regards the women who were very close to the official primary grade one starting year of 7 years. These are women who were 8, 9 or even 10 years at the time of UFPE policy implementation. Based on our cutoff, these are in the control cohorts although they are partially treated because the UFPE policy allowed drop outs and late entrants to benefit from the free primary education policy (fuzzy RDD). The effects of the UFPE educational policy would be significantly underestimated given that the full impact of the UFPE policy would be greater than our estimate if the effect of the school drop outs and late entrants is significant. The given DHS data however, is insufficient to quantify the impact of the late entrants and school drop outs who later went back to school.

3.5.2. Unconfoundedness

The other concern is the exclusion restriction. This entails that other government policies are most likely related in effect to the UFPE policy at designated cutoffs. Given this scenario, our causal estimate would constitute of the UFPE policy and other government policy. As far as the available information suggests, there has been no other policy introduced that shares the same eligibility requirements for treatment as the UFPE policy, nor has any other policy solely impacted a subset of the observations within the optimal bandwidth.

Here, the inclusion of local randomisation is an essential aspect of the regression discontinuity design. Therefore, it is imperative that outcome determinants, apart from the treatment (education), undergo continuous evolution (smooth assumption) at the cut-off point, encompassing women born prior to, during, and subsequent to the cut-off years. The major challenge with DHS surveys is that they lack retrospective data of the participants. To test the smoothness assumption (exclusion restriction) of the RDD design, we examine pretreatment covariates (Imbens and Lemieux, 2008; Jacob et al., 2012).

Religion, wealth status and type of place of residence are identified as the most significant pretreatment covariates in the dataset. The results are shown in Table 3.6 using a four sample *t*-test. The table examines the hypothesis that there are no significant differences in pretreatment variables between respondents located on either side of the cut-off point. Obtaining appropriate variables from survey data for conducting hypothesis testing of this nature is typically challenging. The most suitable variables in the dataset for testing purposes are religion, wealth status and type of residence. Religion is typically regarded as a relatively stable trait, with individuals frequently adhering to the religious beliefs of their families. Wealth status and type of place of residence are equally difficult to change over a short period of time.

In cases where treatment allocation is random, there should be no discernible differences among respondents in terms of pre-existing variables. Our results in the table confirm this underlying assumption. The coefficients on the predetermined covariates are insignificant. Therefore, we can assert that the selection of the sample and the assignment of treatment were not biased. To further confirm that there is no manipulation of the running variable, we run a density test, introduced by McCrary (2008).

Table 3.6: Pretreatment covariates

	Religion	Wealth
Coefficient	-0.162	-0.465
t	-0.36	-1.43
$Pr(\lvert T\rvert > \lvert t\rvert)$	0.718	0.152

Notes: The table tests the hypothesis that there are no significant differences in pretreatment variables between respondents on either side of the cut-off point. The underlying premise is that in cases where treatment allocation is randomized, there should be no discernible differences among respondents with respect to pre-treatment variables. The data presented confirms the hypothesis to be true.

3.5.3. External Validity

The limitation of our findings is that they can only be effectively interpreted within a specific optimal time frame of 6 years before and after the designated cutoff. Nevertheless, within the RDD framework, this issue is commonly encountered. However, the results of this study hold significance within the African and other nations that have yet to adopt UFPE-type policies and are seeking to attain the advantages associated with this educational policy.

Lastly, the draw back to our results lies in the fact that the results are only interpretable within a chosen optimal window/bandwidth of 6 years on both sides of the cutoff. However, in the RDD framework, this is a general problem. On the other hand, the findings of this study are nonetheless important in the context of African and other countries that have not yet implemented UFPE-type policies and are looking to solicit the benefits of this educational policy (Nishimura and Byamugisha, 2011; Bonney, 2021). Equally, for countries that have implemented the UPFE policies, there is lack of information on the health benefits of these free education policies. This study will therefore add to the body of knowledge, particularly for those countries. The outcome will also serve as a baseline for countries yet to implement the free primary education policies.

3.6. Robustness Checks

The assumptions that underpin our IV-estimates necessitate that women born (just) before and women born (just) after the cut-off are not different systematically regarding factors affecting outcomes for their children other than by their differential exposure to education. This suggests that women on each side of the threshold are comparable, thus adding controls should not affect the results. In other words, the only difference between these two groups is whether or not they attended primary school. Firstly, to address this issue, we include fixed effect and control variables.

To further test the validity of our results we estimate the results to varying functional form and varying bandwidth. According to the results in Table 3.7, estimating the results using different bandwidths does not alter the significance of the results. We re-estimate the first stage results with bandwidths of four, six eight and ten years on either side of the cutoff. Even from as low as a four year bandwidth, the re-estimated results show increases in the effect of the UFPE policy on educational attainment. After a bandwidth size of 6 years, the estimates are seen to be stable. The optimal results show that the UFPE policy led to an increase in the completed years of schooling by about 0.7 years. The results are therefore robust to varying birth year cohort on either side of the cutoff.

In addition to selecting the appropriate bandwidth, we conducted an analysis to assess the sensitivity of our findings to variations in functional form, as a means of further evaluating the robustness of our results. We modify our main equations (1) and (2) by incorporating higher order polynomials of the forcing variable, further allowing for interaction of the running variable, (centered at the cutoff) with the treatment indicator, $Z_i = 1[Year\ of\ Birth \geq c(1988)]$. We maintain the MSE-optimal bandwidth of six in all our re-estimations. The results, estimated using the main sample, are displayed in Table 3.8. Across the specifications, our estimates exhibit a consistent level of robustness in terms of both magnitude and statistical significance.

***Table 3.7*: First stage estimates with different bandwidths**

	Dependent variable: Years of schooling							
	4 year bandwidth		6 year bandwidth (optimal bandwidth)		8 year bandwidth		10 year bandwidth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1[Birth year $\geq$ 1988]	0.378**	0.452***	0.708***	0.685***	0.701***	0.674***	0.733***	0.69***
	(0.158)	(0.125)	(0.141)	(0.107)	(0.126)	(0.094)	(0.114)	(0.86)
Mean of dep. var.	6.31		6.26		6.23		6.28	
F-statistic	5.69	13.15	25.34	41.09	30.87	50.94	41.18	64.36
Observations	9927	9927	15148	15123	19403	19374	22651	22618
Controls	No	Yes	No	Yes	No	Yes	No	Yes

This table presents re-estimated first stage results. Columns (1), (3), (5), and (7) are devoid of control variables, while columns (2), (4), (6), and (8) incorporate control variables for place of residence, wealth status and religion. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The standard errors have been clustered at the survey clusters and are presented in parentheses $***$ *p < 0.01,* $**p$ *< 0.05,* $*p$ *< 0.1.*

Table 3.8: Main Results with different functional forms

	Linear	Linear Interaction	Quadratic	Quadratic Interaction	Cubic	Cubic Interaction
Height-for -age	0.111***	0.112***	0.113***	0.114***	0.113***	0.114***
z-score	(0.124)	(0.0123)	(0.0123)	(0.0124)	(0.0124)	(0.0123)
Stunting	-0.0368***	-0.0369***	-0.037***	-0.0374***	-0.0373***	-0.0377***
	(0037)	(0.0037)	(0.0037)	(0.0037)	(0.0037)	(0.0037)
Weight-for-age	0.0832***	0.0836***	0.0841***	0.0864***	0.0861***	0.0866***
z-score	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Underweight	-0.01967***	-0.0197***	-0.0198***	-0.0205***	-0.0204***	-0.0203***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Weight-for-	0.2219**	0.02205**	0.0224**	0.0252***	0.0248***	0.025**
height z-score	(0.011)	(0.0111)	(0.0111)	(0.011)	(0.11)	(0.011)
Wasting	-0.0005	-0.0006	-0.0007	-0.0007	-0.0007	-0.0007
	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Observations	12844	12844	12844	12844	12844	12844
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents re-estimates of the main results using different functional forms. All specifications include controls, region and survey fixed effects. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The estimates remain robust to varying functional forms. Robust standard errors have been clustered at the survey clusters and are presented in parentheses $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Lastly, to further test the validity and robustness of our estimates, we examine the running variable for possible manipulation. This test was developed by McCrary (2008) and tests for continuity of the running variable in the context of the RD design. The fundamental concept underlying manipulation testing in this particular context is that if there is no deliberate and organised manipulation of the unit's index in the vicinity of the cutoff point, the distribution of units should exhibit continuity in the proximity of this threshold value. Therefore a manipulation test is designed to formally establish the presence of a discontinuity in the density of units at a known cutoff point. The existence of such evidence is commonly construed as empirical evidence indicating self-selection or nonrandom sorting of units into control and treatment status. The results are presented in Table 3.9. The resulting plot is shown in lation of *the running variable.*

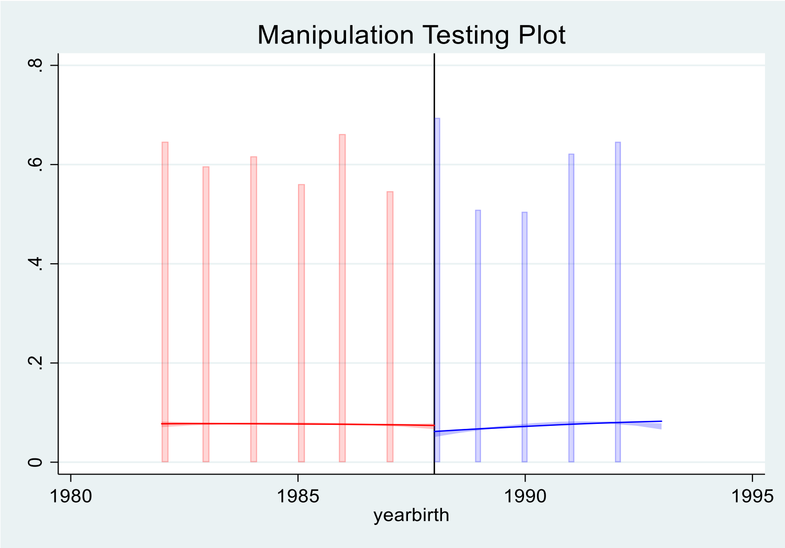
Figure 3.5.

Table 3.9: Manipulation Test of running variable

Method	T	P>|T|
Robust	-0.1557	0.8763

Notes: The outcome of the ultimate manipulation test indicates a T- value of -0.1557, accompanied by a p-value of 0.8763. We fail to reject the null hypothesis of no manipulation. Therefore the results lack statistical proof of any deliberate manipulation of the running variable.

Figure 3.5: Manipulation Testing Plot



This figure plots the density of the running variable against the year of birth. At the cutoff, there are no visible signs of any manipulation of the running variable. In other words, there is no bunching of the running variable at the cutoff.

We do not find any significant effect of the UFPE educational policy on men. Figure A1 plots the years of schooling for men with the year of birth. There does not appear to be any discontinuity for men at the cutoff. The UFPE educational policy did not seem to affect males as it did females. One plausible hypothesis is that the notable age disparity between spouses may account for this finding, as more than 90% of the fathers included in the study were born prior to 1988. Due to the unavailability of a valid instrument for the educational attainment of fathers, it is not possible to ascertain the causal impact of paternal education on child health. This finding also lends credence to the idea that school fees did not pose a significant obstacle to boys' primary education.

The study primarily concentrated on maternal education, so neglecting the potential significance of paternal education, despite its potential relevance (Chou et al., 2010). We find no exogenous variation in schooling for males as we did for females. Furthermore, it is widely acknowledged that there exists a substantial correlation between the educational attainment of husbands and women (Chou et al., 2010), hence introducing potential challenges in discerning any independent impacts. Additionally, the inclusion of fathers in the study would have resulted in a smaller sample size as it cannot always be guaranteed that the woman's partner is indeed the biological father of the child. However, it is important to recognise that the estimations regarding maternal education may have been overestimated as a result of not considering paternal education. (Breierova and Duflo, 2004; Chou et al., 2010). This remains an area for possible further research. Without the discontinuity (jump), the fundamental continuity assumption of the RDD is violated. This manifests itself in a weak first stage F-statistic. The first stage F-statistic for men shows a weak F-test of 13.78, suggesting a weak instrument problem (See Table A2).

Many other papers equally find no significant effect of the free primary education policies on men (Makate, 2016; Gunes, 2016; Keats, 2018; Odunowo, 2020; Wild and Stadelmann, 2020). Nonetheless, this constitutes a domain that warrants further investigation in the future. This explains why it has been difficult to do comparative causal studies between men and women. The free primary education policies do not seen to affect men in almost all studies.

3.7. Conclusion

This paper presents empirical evidence regarding the causal impact of maternal education on child health outcomes in Zambia through the utilisation of the 2002 universal free primary education (UFPE) policy as a natural experiment in a fuzzy regression discontinuity design (RDD) setting. As a result, we use instrumental variable (IV) estimations to measure the impact of additional maternal education on child health outcomes. This allows for reliable causal estimates. Additionally, this research investigates potential pathways through which maternal education exerts its influence on child health.

The findings indicate that the UFPE education policy had a positive impact on the educational achievements of women, irrespective of their socioeconomic status. Using this exogenous source of variation in education as an instrumental variable, we find significant causal impacts of maternal education on child health measured by three widely utilised indices namely height-for-age, weight-for-height and Weight-for-age z scores. The results equally show that an additional year of schooling by the mother decreases incidences of stunting and being underweight. We find no effects of additional maternal education on wasting. We find that women who possess higher levels of education tend to make a trade-off by sacrificing new fertility in order to allocate greater resources towards the upbringing and well-being of their existing children.[6] Specifically, children of mothers who possess higher levels of education exhibit improved nutritional status.

We further examined the underlying mechanisms through which maternal education contributes to improved child health outcomes. Several channels have been hypothesized as potential mechanisms through which maternal education could bring about alterations in child health outcomes. We show the possible mechanisms via which maternal education impacts child health. Our results indicate that mother's education improves child health through increased age at first marriage, reduced total number of children ever born, increased age at first birth and increased age at first sexual intercourse.

The education of women has a significant impact on enhancing the future achievements of the next generation. This effect is further amplified by positive assortative mating, whereby if a woman marries a highly educated partner, both individuals are more likely to have greater ambitions for the health of their children. The findings indicate that higher levels of education among women have a beneficial impact on the health of future generations. Also, education additionally empowers mothers to access information via television and newspapers, to gain understanding of the ovulation cycle, and to choose contraceptive techniques. Surprisingly, we find no effects of additional mother's schooling on labour market outcomes.

Furthermore, we provide evidence of the heterogeneous effects of a mother's education based on the socioeconomic status of the household, rural/urban status, region and religion. We find that the effect of an extra year of schooling on height-for-age z-scores is higher for the wealthy subgroup

[6] As a potential mechanism through which education affects child health, we find a significant reduction in total children ever born.

compared to the poor counterparts. The decrease in the probability of being underweight as a result of additional schooling is higher in the poor subsample as opposed to the wealthy subsample. On the other hand, we find no heterogeneities in the likelihood of stunting in both the poor and wealthy subgroups. Equally, the effect of additional years of schooling on weight for age z-scores is similar in both subgroups. We find scanty evidence of the effects of additional education on outcomes based on region and religion. Lastly, we also find evidence of heterogeneities by socioeconomic status in the effects of additional female schooling on the potential pathways through which maternal education affects child health outcomes.

The outcomes of this research are tailored to the participants within the chosen optimal window, and affected by the UFPE education policy, as per the study's design. However, it is possible to extrapolate the findings to the entire population, given that individuals under the age of 30 comprise a significant proportion of the Zambian population. The validity of these results was subjected to a number of robustness and falsification tests.

CHAPTER 4: Free primary education, female schooling and fertility: A regression discontinuity design

4.1. Introduction

The importance of education, particularly for women, has been recognised as the primary factor in generating favourable impacts on economic and social results. Education exerts significant influence on economic growth and income inequality (Mankiw, Romer, & Weil, 1992), as well as on worker productivity and earnings in the labour market (Card, 1999). Additionally, it has spillover effects on the productivity of other workers (Acemoglu & Angrist, 2000; Iranzo & Peri, 2009). Education has an impact on non-market results such as an individual's health, fertility, the long-term economic outcomes of their offspring, and the quality of their children, among other results (Grossman, 2006). Due to the significant influence of education on both the economy and personal welfare, numerous developing nations have allocated resources towards initiatives such as constructing schools, implementing conditional cash transfers, and enforcing compulsory schooling laws. These efforts aim to enhance educational opportunities and encourage individuals to pursue higher levels of education (Burde & Linden, 2013; Cui, Liu, & Zhao, 2019; Schultz, 2004).

The interplay between fertility and education is a crucial aspect of both poverty reduction and development. Schultz (2007) has provided evidence that the demographic transitions leading to decreased fertility have several advantages. The benefits that accrue to the mother include better health, more investment in human capital, and increased labor force participation. Meanwhile, the child benefits from improved nutrition, increased investment in education, and a range of improved health outcomes. The total fertility rate (TFR) of sub-Saharan Africa is more than twice that of any other region of the globe. According to the World Bank (2019), it is estimated that the population of Africa will experience a twofold increase by the year 2050, thereby constituting 23% of the global population. This represents a significant rise from the 7% recorded in 1960.

Promoting the education of girls and women has long been recognised as an effective measure for lowering fertility rates (Osili and Long 2008). An increasing amount of conclusive research from

SSA demonstrates that education, both at the primary and secondary levels, decreases the occurrence of adolescent pregnancies (Osili and Long, 2008; Baird, McIntosh, and Ozler, 2011; Duflo et al., 2015). Although a significant portion of these research primarily examined school girls, particularly in relation to teenage pregnancy, there is some evidence suggesting that education generally reduces fertility in general (Keats. 2018).

Economic theory proposes a number of strategies by which education might influence fertility decisions. First, education raises the returns to labour force participation, raising the opportunity cost of time-consuming activities such as childbearing and rearing and shift fertility preferences towards fewer children of higher quality (Becker and Lewis 1973; Moav, 2005). The influence of the income effect is amplified by assortative mating, whereby women with higher levels of education tend to marry men who are also highly educated and possess higher incomes. According to Becker's (1960) theory of quality-quantity trade-off, individuals with higher income tend to have a smaller number of children but with a higher quality. Second, education may raise a woman's awareness of and use of contraception and other healthy behaviors, which may result in decreased fertility (Grossman 1972; Rosenzweig and Schultz 1989). Education has the potential to enhance women's ability to negotiate within the context of a marital relationship, thereby increasing their likelihood of participating in decisions related to reproductive health. Furthermore, education may boost women's bargaining power and autonomy in fertility decisions (Mason, 1986). Third, education may have a direct effect on fertility reduction via the "incarceration effect." This relates to the fact that education increases women's time spent in school, hence diminishing or delaying their prospects for sexual activity and childbearing. However, such a delay in childbirth may be transitory and does not always affect total fertility. (Black et al, 2008). Lastly, education can also affect women's fertility through empowering them with necessary skills, including reading and numeracy, to access family planning information outside of the school context, for example, through the use of mass media channels (Glewwe 1999; Thomas et al, 1991).

While a substantial body of empirical evidence indicates a negative association between female education and fertility (Ainsworth et al., 1996; Martin 1995; Cochrane 1979), proving a causality is difficult due to the possibility of reverse causality and selection on unobservable factors. Omitted variable bias, for example, could confound the estimation (Osili and Long, 2008). Furthermore, if fertility decisions have a negative impact on education decisions, it is possible that fertility is endogenous (Angrist and Evans, 1999). Because of this, models that do not account for education's possible endogeneity are likely to provide erroneous estimates.

In light of the endogeneity issue associated with education, the optimal approach for investigating causality involves utilizing an exogenous determinant of educational attainment that is not correlated with fertility results. Several recent studies address these issues by examining the exogenous variation in school entry policies (McCrary and Royer 2011). More recently, in Sub-Saharan Africa, studies have used quasi-experimental methods taking advantage of exogenous gains in schooling caused by certain educational policies as an instrumental variable (IV) framework to identify the education effects (Breierova and Duflo 2004; Ferre 2009; Chicoine, 2012; Tequame and Tirivayi, 2015; Ozier, 2018; Ali and Gurmu 2018; Wild and Stadelmann, 2020). All of these recent causal studies find a significant negative impact of female education on fertility. Several academic studies have indicated that women with higher levels of education tend to have fewer children (Black, Devereux & Salvanes, 2008; Osili & Long, 2008; Lavy & Zablotsky, 2015; Ozier 2018; Mocan and Cannonier 2012; Chicoine, 2012). Nevertheless, there exists evidence that contradicts this association. The studies conducted by Braakmann (2011) and Fort, Schneeweis, and Winter-Ebmer (2011) have reported a favourable impact of education on fertility outcomes in the United Kingdom and Europe, respectively. Conversely, McCrary and Royer (2011) find no effect in the United States.

The paper uses the 2002 Universal Free Primary Education (UFPE) policy as an exogenous source of variation in education to estimate the causal effect of additional education on fertility outcomes. In filling this research gap, this study adds to the evidence by examining the heterogeneous effects by socioeconomic and rural/urban status. Few studies examine these heterogeneities in Sub-Saharan Africa (Adu-Boahen and Yamauchi 2018; Wild and Stadelmann, 2020). Further, given the fact that educated mothers do eventually catch up to their peers with fertility (Black et al, 2008), this study considers total (permanent) as opposed to teenage (temporal) fertility.

This paper explores the execution of Zambia's UFPE education policy reform as a quasi-experimental design, which, to the best of our knowledge, has not been utilised for evaluating potential impacts previously. The UFPE reform implementation offers a variation in educational attainment among age cohorts that is considered to be "as good as randomized" due to the similar background characteristics shared by women born on either side of the cutoff and the exogeneity of their year of birth (Lee and Lemieux, 2010). A fuzzy regression discontinuity design (RDD) is utilised in order to detect exogenous variation in female education. Subsequently, instrumental variable (IV) estimations are performed to investigate the influence of education on fertility. This study uses pooled survey data from the 2002, 2007, 2013/14 and 2018 ZDHS.[7]

Our results show that the UFPE educational policy had a positive impact on female schooling. Additionally, this exogenous surge in female educational outcomes leads to a decrease in fertility and fertility preferences, while simultaneously resulting in a delay in the age at which individuals have their first child. We further examine the potential pathways through which increased female education affects fertility outcomes. We find that an additional year of schooling delays the age at first marriage and first sexual intercourse and increases the probability of contraceptive use. There is also evidence of positive assortative mating. Finally our results also demonstrate the knowledge effect.

4.2. Data and Empirical Strategy

4.2.1. Data Sources

The data for this study came from three rounds of the Zambia Demographic and Health Surveys (ZDHS), which are nationally representative surveys of reproductive-aged women conducted in Zambia in 2007, 2013/14, and 2018. The DHS sample is drawn in two stages: first, enumeration regions are randomly selected from census files, and then households are randomly drawn from the list of households within each enumeration area (ZDHS, 2018). Data on socioeconomic and demographic variables, fertility, family planning, and maternal and child health are collected as part of the ZDHS, along with information on health care use, mortality and nutritional outcomes for children.

Specifically, the purpose of this research is to investigate the relationship between women's educational attainment and a variety of fertility outcomes. We examine the effect of female education on three fertility outcomes: the number of children ever born per woman, ideal (preferred) number of children, and the age of women at first birth. We also examine the potential alleyways through which female education affects fertility. We use the DHS wealth index score to

[7] The outcomes show no significant alteration in the event of non-pooling of data.

explore possible heterogeneities in the effect of education. The DHS wealth index score which is constructed by assessing households' living arrangements, such as the construction materials and sanitation facilities, as well as household assets such as radio and TV, bicycles, cars etc. and by placing them on a relative scale of wealth within the sample. The wealth index has five categories in the DHS. These are: poorest, poorer, middle, richer and richest. In line with other publications, we considered respondents from households scoring on and below the median wealth score as "poor", and respondents with a denoted household wealth level above the median as "wealthy".

The primary independent variable we are focusing on is the number of years women have spent in formal education. The DHS provide data on the total number of years of schooling completed by individuals, representing a continuous measure of educational achievement. For the sake of clarity and comparability with prior studies, we present the results based on years of completed education in this article. We also include a set of pretreatment covariates such as the respondent's, religious affiliation, region, and ethnicity and also whether the respondent lives in a rural or urban area. Table 4.1 shows the description of key variables.

Table 4.1: Description of key variables

Variable	Description
Mothers' education	Years of schooling
Fertility	Number of children ever born per woman,
	Ideal number of children
	Age of women at first birth.

Source: DHS 2007, 2013/2014 and 2018 pooled survey data for Zambia

4.2.2. Empirical Strategy

As per economic theory, there exists a correlation between education and fertility outcomes. The following is a simplistic parametric model that illustrates the correlation between educational achievement and fertility outcomes:

$$Y_i = \alpha + \beta Educ_i + \Theta X_i + \varepsilon_i \qquad (1)$$

where Y_i is a measure of fertility outcomes for individual i and $Educ_i$ represents the educational attainment of individual i measured as years of schooling. Equation (1) also includes a vector of control variables such as survey fixed effects, region fixed effects, socioeconomic status of the individual and a dichotomous variable equal to 1 if the individual lived in a rural area and 0 otherwise. Our main parameter is β. Estimating β using OLS is likely to lead to estimates that are biased upwards. Innate ability or personal preferences could predict both schooling experiences and fertility outcomes leading to endogeneity (simultaneity bias). Further the OLS estimates could exaggerate the impact of education because education serves as a partial substitute for omitted characteristics leading to biased estimates upward.

According to neoclassical fertility models, there exists a correlation between higher levels of female education and a trade-off between the quantity and quality of children, as posited by Becker (1960) and Becker and Lewis (1973). As a result, a variety of policy initiatives have been implemented with the goal of speeding up the demographic transition and thereby delivering these benefits sooner than would occur otherwise and education is viewed as a development tool in this regard (Schultz, 1993). Given the disproportionate burden of disease in SSA, estimating a causal relationship between education and fertility becomes more imperative. However, a causal relationship is hard to establish considering potential reverse causality and selection on unobservable factors like cognitive ability, or family preferences that could simultaneously determine both education and fertility preferences (Bledsoe et al., 1999).

In order to circumvent this endogeneity problem, we implore the implementation of the UFPE educational policy in 2002 which resulted in a dramatic increase in primary school enrollment and retention rates across the country. This introduction of free primary education (exogenous source of variation in educational attainment) allowed more Zambians than usual to attend primary education as the main barrier (school fees) had been abolished.

The exogenous increase in enrollment resulting from the UFPE policy is consistent with a regression discontinuity design (RDD) for determining the causal influence of education on fertility outcomes. The RDD design compares individuals who were "just-treated, i.e., young enough to benefit from the fee removal (age 14 at policy implementation) versus individuals who were just-too-old" (age 15) (Keats, 2018; Behrman et al., 2017; Wild and Stadelmann, 2020; Aguero and Bharadwaj, 2014).

Our technique is predicated on the assumption that cohorts of individuals near the cut-off (on either side) are comparable, with the only probable variation in outcomes attributable to the surge or exogenous source of variation in education (Gelman and Imbens 2014; Lee and Lemieux 2010; Imbens and Lemieux 2008).

Late enrollment in school, extended absence, and grade repetition are all fairly prevalent in Zambia. Individuals who were older than the primary school exit age were subject to the policy due to their later enrolment in lower primary school grades. Others continued their primary school education until their late ages. Furthermore, several younger individuals, particularly women had dropped out of school permanently owing to work or early marriage, and as a result, they were unable to take advantage of the UFPE program.

We use the age – specific nature of the UFPE reform to sort individuals into treatment and control groups. Due to the fact that entry time into primary school is influenced by a lot of other factors, some unobserved, we use individual's birth year rather than the year in which she first entered primary school as the rating (forcing) variable. Since the forcing variable is year of birth, it is highly improbable that the assignment to treatment is a deterministic function of it. Given such non-compliance, a fuzzy RDD as opposed to a sharp RDD is used for estimations. Our analysis first evaluates if this UFPE policy has an effect on educational attainment and subsequently on fertility outcomes.

4.2.2.1. UFPE education policy and educational attainment

In order to evaluate the effect of the UFPE policy on educational outcomes, we estimate the following first stage regression:

$$S_i = \beta_0 + \beta_1 T_i + f(Birthyear_i - c) + \Theta X_i + \varepsilon_i \qquad (2)$$

where T_i denotes the treatment, for individual i, $f(.)$ is a polynomial function of the forcing variable (birth year), and c denotes the cut-off year (1988). $T_i = 1(Birthyear \geq c(1988))$. The vector X_i includes pretreatment covariates including survey and region fixed effects, place of residence and socioeconomic status while ε_i represents the idiosyncratic error term. The vector β_1, denotes the average causal effect of the UFPE educational policy reform on years of schooling.

4.2.2.2. UFPE Educational Policy and fertility Outcomes

Two regression models are employed to determine the impact of the UPFE educational policy reform on fertility outcomes. Initially, the reduced form of equation (1) is estimated. This approach assesses the impact of the policy reform on a cohort that was exposed to it, i.e., the *"intent to treat" (ITT)* effect, under the assumption that the monotonicity condition holds. This means that the educational policy implemented by UFPE incentivises individuals to extend their duration of education and does not result in premature dropout. For this, we regress the outcome of interest Y_i on equation (2) as follows:

$$Y_i = \beta_0 + \gamma_1 T_i + f(Birthyear_i - c) + \lambda X_i + \varepsilon_i \qquad (3)$$

Where the parameter γ_1 denotes the reform's average causal effect on fertility and other related outcomes. T_i is a dummy variable that indicates if an individual was impacted by the UFPE educational reform; it is equal to one if the individual's year of birth is greater than or equal to c, and to zero otherwise ($T_i = 1(Birthyear \geq c(1988))$).

Second, we estimate the *"treatment on the treated"* (TOT) impact for the compliers, which is the local average treatment effect for the compliers evaluated at the cutoff threshold. To have such a causal estimate, we employ 2SLS with T_i as the instrumental variable for educational attainment (years of schooling).

The causal effect of an extra year of schooling in the fuzzy setting is defined as "the difference in outcome from a regression on the treatment-determining variable (birth year) divided by the difference in treatment (an extra year of primary schooling) from a regression on the treatment-determining variable, both evaluated at the cutoff value, c" (Hahn, Todd, and Klaauw 2001; Imbens and Lemieux 2008; Lee and Lemieux 2010) as follows:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} \mathrm{E}[Y \mid x] - \lim_{x \uparrow c} \mathrm{E}[Y \mid x]}{\lim_{x \downarrow c} \mathrm{E}[education \mid x] - \lim_{x \uparrow c} \mathrm{E}[education \mid x]}$$

This is equivalent to a 2SLS regression model. Finally, we estimate the second stage regression by regressing the outcome variable Y_i on predicted values of schooling $\widehat{S_i}$ from the first stage regression in equation (1):

$$Y_i = \gamma_0 + \gamma_1 \hat{S}_i + f(Birthyear_i - c) + \theta X_i + \varepsilon_i \qquad (4)$$

Where γ_1, estimates the causal impact of schooling on fertility outcomes. This is the local average treatment effect (LATE) for compliers. Both the first and second stage regressions (equations 2 and 4) should have the same order of the polynomial and bandwidth for estimation of standard errors as in a 2SLS framework.

4.3. Results

4.3.1. Data and Descriptive Statistics

The descriptive statistics are shown in Table 4.2. The presented table displays statistical data for the entire sample, as well as for two sub-samples categorized as "poor" and "wealthy". These sub-samples were created by dividing the household wealth into two distinct groups. Of the five categories in the wealth variable, the poorest, poorer and middle categories are reclassified as "poor". The richer and richest categories are reclassified as "wealthy".

The table shows that the mean years of education for the control group in the full sample is 6.5 years. Consistent with expectations, the treatment group exhibits a higher mean educational attainment of 7.6 years compared to the control group. The wealthy subsample exhibit more years of education for both control and treatment cohorts compared to the poor counterparts. Regarding the two subsamples, as a result of being able to benefit from free primary education, (Being born in or after 1988) the poor subsample treatment cohort exhibit an immediate and large improvement in educational achievements in comparison with the control cohorts born earlier (Being born in 1987 and earlier). The evidence of the discontinuity (jump) is equally shown in

Figure ***4.1***. The full sample also shows that about 61 percent of the women are literate and that the husband has more years of schooling compared to the wife.

The table also shows that about 51 percent of the sample live in the rural area. The mean desired number of children is around 4 while the mean total number of children ever born is around 2. Lastly, the table also shows vital statistics for the mean age at survey, age at first sex, age at first birth and age at first marriage.

Table 4.2: Summary Statistics

	Full Sample		**Poor**		**Wealthy**	
	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)
Number of observations	9156	7667	4014	4235	5085	3489
Years of completed education	6.52 [3.97]	7.64 [3.59]	4.62 [3.01]	5.98 [3.09]	8.03 [3.89]	9.62 [3.09]
Complete Primary	0.51	0.39	0.65	0.58	0.4	0.19
Literacy: able to read sentence	0.49	0.61	0.3	0.45	0.63	0.8
Husband's years of schooling	7.94 [3.94]	8.41 [3.68]	6.34 [3.36]	7.09 [3.23]	9.85 [3.76]	10.8 [3.2]
Age at survey	26.15 [6.1]	22.87 [4.07]	28.86 [4.27]	23.1 [3.97]	24.1 [6.49]	22.54 [4.16]
Age at first birth	18.6 [3.07]	18.21 [2.54]	18.27 [2.72]	17.85 [2.223]	19 [3.4]	18.8 [2.87]
Age at first sex	16.55 [2.74]	16.6 [2.48]	16.07 [2.31]	15.97 [2.12]	16.94 [2.99]	17.37 [2.66]
Age at first marriage	18.26 [3.66]	17.98 [3.06]	17.74 [3.32]	17.45 [2.79]	18.91 [3.93]	18.95 [3.29]
Residence (Rural = 1)	0.56	0.52	0.83	0.82	0.34	0.16
Total number of children born	2.55 [2.01]	1.64 [1.49]	3.81 [1.86]	2.09 [1.54]	1.58 [1.65]	1.08 [1.2]
Desired number of children	4.53 [1.84]	4.19 [1.6]	5.24 [1.85]	4.65 [1.69]	3.99 [1.65]	3.64 [1.3]
Wealth Index combined (1-5)	3.12 [1.41]	3.16 [1.43]	2.01 [0.82]	2.02 [0.82]	4.51 [0.5]	4.53 [0.5]
Works (Yes =1)	0.53	0.42	0.59	0.47	0.48	0.34

Notes: The study employs data 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The analysis utilizes a bandwidth size of 6 birth year cohorts on both sides of the cut-off. The subsamples denoted as 'poor' and 'wealthy' comprise of female individuals whose wealth levels are either below or above the median, respectively. The statistical analysis incorporates the DHS sample weights and takes into consideration the pooling effect across survey rounds.

4.3.2. Bandwidth Selection

Following the recommendations by Abadie et al. (2017) and Nunn (2010) standard errors are clustered at the level of survey clusters. This approach to clustering yields standard errors that are robust against arbitrary patterns of within-cluster variation and covariation (Cameron and Miller, 2015). The standard errors are clustered at the survey cluster level since this is the level at which units in the sample were selected and there are clusters in the population that are not represented in the sample. Clustering at survey clusters is useful because while respondents are identical across cohorts, their characteristics may differ by location; in the context of this study, social and economic status can vary much within and between clusters. A section of the literature also clusters at the level of running variables. The point estimates remain robust to clustering at the level of the running variable. The strength of the instrument, however, becomes sensitive to alternative bandwidths in this approach.

Recent academic literature on the regression discontinuity design supports the preference for employing local linear regression estimation instead of global polynomial estimation. We perform an estimation of a local linear regression model on both sides of the threshold, allowing for variation in the slopes on either side of the threshold (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). By employing the 2SLS method, we estimate the Local Average Treatment Effect at the Cutoff (LATE) specifically for individuals who comply with the treatment assignment.

In order to determine the most suitable bandwidth, we employ the mean squared error (MSE-optimal) bandwidth method, utilizing the local polynomial bandwidth selection method proposed by Cattaneo et al. (2014; 2017). The proposed methodology yields an optimal bandwidth of 6.31 years on both sides of the cutoff. Hence, all estimations in the first and second stage regressions were performed with a bandwidth of 6 years on both sides of the cutoff.

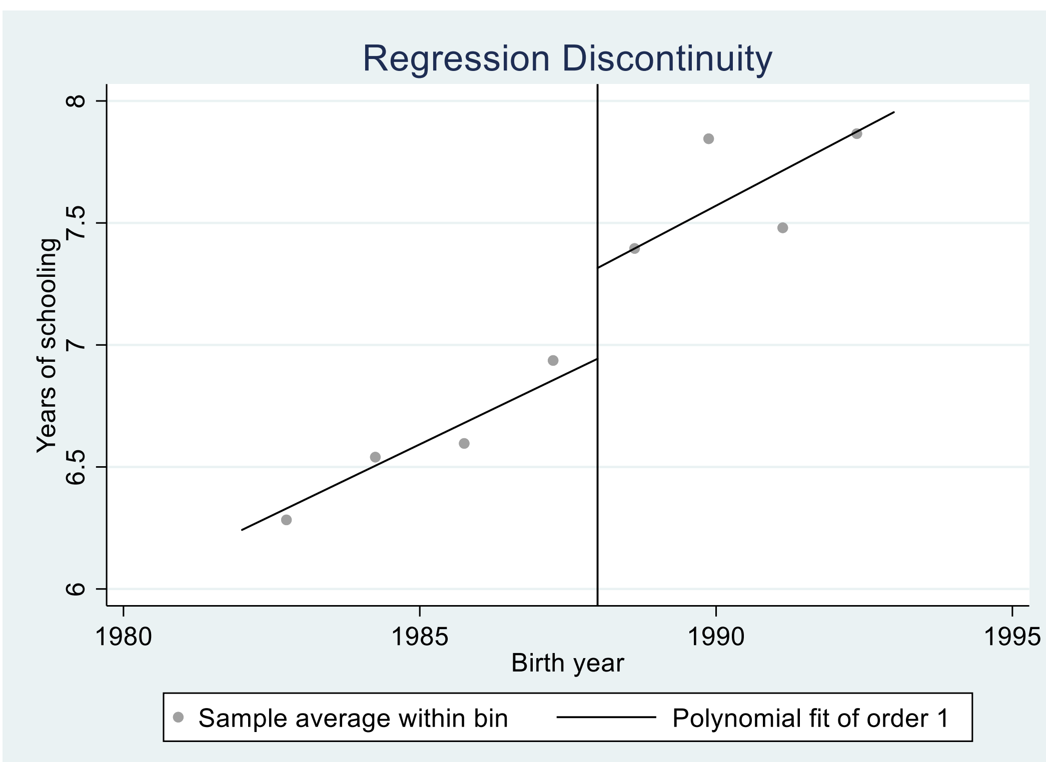
4.3.3. First Stage Discontinuity in educational attainment

We begin by showing empirical evidence that the UFPE educational reform yielded favorable outcomes in terms of educational attainment, as indicated by an increase in the number of years of schooling. This surge in educational attainment is shown in

Figure ***4.1*** which provides evidence to support the hypothesis that the implementation of the UFPE policy had a significant impact on the educational attainment of women. The figure shows that women who had access to cost-free primary education, specifically those born in 1988 or later, exhibited a significant and prompt improvement in their educational attainment in comparison to the groups born earlier. The educational attainment of female individuals is observed to increase with the rise of birth year cohorts subsequent to the cut-off point, owing to the fact that girls born in later years have a greater number of primary schooling years available to them without incurring fees. In contrast to women's education by birth year cohort, there are no discontinuities for children born, both actual and preferred as shown in

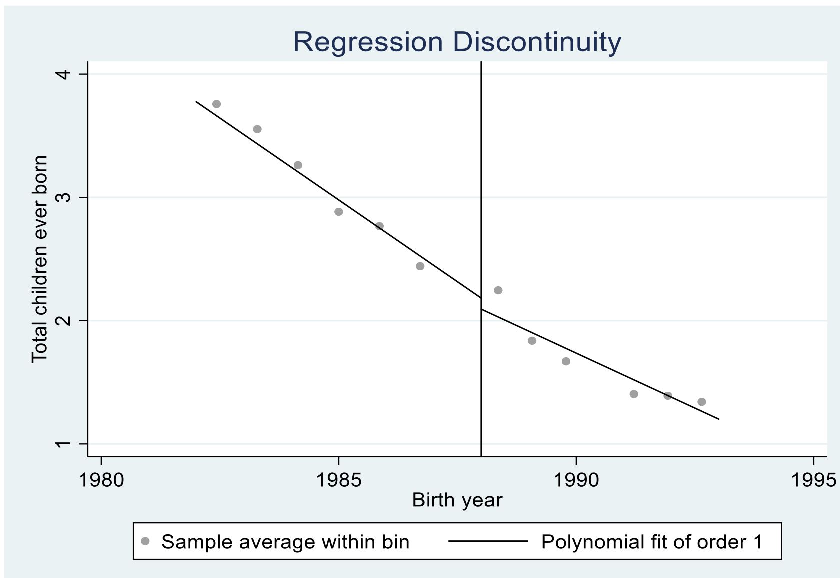
Figure 4.2 and Figure 4.3.

Figure 4.1: Women's Schooling by Birth Year Cohort



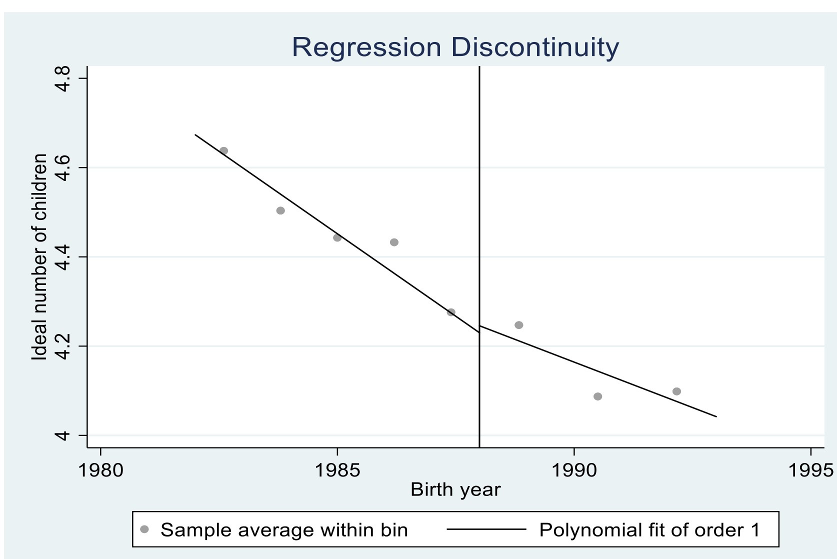
Source: Pooled data from the 2007, 2013/14, and 2018 DHS surveys

Figure 4.2: Total children ever born by birth year cohort



Source: Pooled data from the 2007, 2013/14, and 2018 DHS surveys

Figure 4.3: Preferred number of children by birth year cohort



Source: Pooled data from the 2007, 2013/14, and 2018 DHS surveys

4.3.4. First Stage Results – Effects of the UFPE policy on years of schooling

The results of the first stage are displayed in Table 4.3. The results present evidence of a jump at the discontinuity. The results are provided individually for the entire sample (columns 1 and 2), the Poor (columns 3 and 4) and the Wealthy (columns 5 and 6) subsamples. Regarding the discussion of outcomes, we make reference to the more rigorous specification that encompasses survey fixed effects and all predetermined covariates. The 2002 UFPE educational reform had a discernible impact on educational attainment, with the earliest cohort to have benefited from the reform being those born in 1988. Consequently, there is a noticeable (jump in years of schooling) at the discontinuity with an increase in educational attainment among individuals born in 1988.

The findings from the first stage results indicate that among female participants who were either assigned to the treatment group or experienced positive outcomes from the UFPE policy, the mean impact of the policy was an increase of 0.68 years in educational attainment relative to their peers in the control group who were not assigned to the treatment. For the full sample, the jump in the discontinuity is economically large and statistically significant. The first stage *F-statistic* at 53.53 is above the minimum figure of 10 suggesting that the instrument is not weak, pointing to a strong first stage relationship (Staiger and Stock, 1997).

In the poor subsample, the effect of the UFPE educational policy (jump in years of schooling at the discontinuity) on years of schooling is 0.81 years while in the wealthy subsample, the effect is 0.96 years. The first stage effect is slightly larger in the wealthy subsample as compared to the poor subsample. This could be because of the fact that although the UFPE educational policy was a national wide policy and applicable to all regardless of socioeconomic status, it was for the wealthy subsample who were already well off an added stimulus.

The effect of the policy remains largely unchanged even after adding predetermined covariates. Equally, the mean of the control cohort group differs substantially between the poor and wealthy subgroups. Women from the poor subgroups without exposure to the UFPE educational policy had 5.3 years of education while their counterparts from the wealthy subgroups had more years of schooling at 8.7 years. Consequently, while the policy implemented by UFPE resulted in the addition of equal years of education for both the wealthy and poor subsamples, it led to a relatively greater increase in educational attainment for the latter in relative terms. Therefore, the first stage findings indicate that the educational policy implemented by UFPE led to a significant and substantial rise in the educational attainment of females.

Table 4.3: First Stage Estimates – Effect of the UFPE Policy on Maternal Education

	Years of Schooling					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
1[Birth year ≥ 1988]	0.802***	0.675***	0.807***	0.795***	1.376***	0.964***
	(0.107)	(0.092)	(0.109)	(0.106)	(0.149)	(0.144)
Mean of dep. var.	[7.03]		[5.32]		[8.68]	
Interaction: 1*Year of Birth	0.117***	0.116***	0.198***	0.188***	0.043	0.047
	(0.294)	(0.0265)	(0.031)	(0.03)	(0.04)	(0.039)
Intercept	6.587***	11.432***	4.49***	7.32***	8.016***	10.98***
	(0.064)	(0.258)	(0.669)	(0.315)	(0.083)	(0.332)
First Stage F-Statistic	55.14	53.53	54.76	56.28	84.78	45.04
Observations	16799	16799	8238	8238	8574	8561
Rsquared	0.022	0.242	0.048	0.102	0.043	0.123
Controls		Yes		Yes		Yes
Region Fixed Effects		Yes		Yes		Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

*Note: for direct interpretation of the coefficient on the discontinuity, the birth year distribution has been re-centered at the discontinuity year (1988). Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff. Standard errors clustered at the survey level are presented in parentheses. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

4.3.5. Effects of Education on fertility outcomes

Having shown that the UFPE education policy had a substantial and positive effect on years of schooling, we now show the causal effect of this increase in years of schooling on fertility outcomes using 2SLS (Treatment on the treated). The second stage results are shown in Table 4.4. The full sample 2SLS results show that each additional year of schooling results in a reduction in the total number of children born by 0.259 births. This reduction in fertility represents a 12.1% reduction from mean of 2.14 births. This discovery aligns with prior research conducted within the framework of developing nations. Several studies, including Breierova and Duflo (2004), Osili and Long (2008), and Duflo et al. (2015), have demonstrated a negative relationship between increased levels of education and fertility rates in Indonesia, Nigeria, and Kenya, respectively.[8]

The findings from the IV-estimates indicate a negative relationship between education and women's desired lifetime fertility. The results show that an additional year of schooling reduces the ideal (desired) number of children by 0.264 births. This represents a 6% reduction off the baseline of 4.38 children. This finding is consistent with estimates from Masuda and Yamauchi (2020), Keats (2018) and Behrman (2015b).

An additional year of schooling also increases the age at first birth. Each additional year of education increases the age at first birth by 0.225 years. This delay is equivalent to 1.2% of the mean age at first birth of 18.24 years.

To examine the heterogeneities in the effect of years of schooling on fertility outcomes, we split the full sample into two subsamples: the poor and wealthy subsamples. The results are shown in column 3 and 4 for the poor subsample and column 5 and 6 for the wealthy subsamples.[9] The results for the subsamples are similar to the full sample results even after including controls. The results in the full sample (column 2) are driven by both the poor and wealthy subsamples (column 4 and 6, respectively). Overall, increased female education lowers the average number of children (actual and ideal), and it increases the age at first birth.

The comparative OLS estimates together with the reduced form (intention to treat) results are presented in column 7 and 8 of **Error! Reference source not found.** respectively[10]. The OLS estimates are derived from the equation specification (1). As previously mentioned, OLS estimates may suffer from bias as a result of unobserved factors that simultaneously affect both educational attainment and fertility choices. The reduced form estimates are based on equation (3). The

[8] Additional research with similar results in both magnitude and sign include Duflo et al. (2006), Ferré (2009), Chicoine (2012) and Ozier (2015) for Kenya, Tequame and Tirivayi (2015) for Ethiopia, Masuda and Yamauchi (2017) and Keats (2018) for Uganda and Hahn et al., (2018) for Bangladesh.

[9] We regard respondents whose households have a wealth score that is equal to or lower than the median to be "poor," and we consider respondents whose families have a wealth level that is higher than the median to be "wealthy."

[10] We present the results for full sample with controls, region and survey fixed effects only.

findings derived from both the OLS and reduced form estimations consistently indicate that increased education for women is associated with a decrease in both the total number of children born and the desired number in the full subsample. The OLS estimates show that increased female education reduced the total number of children ever born by 0.15 births while the reduced form estimates a reduction in the total number of children ever born by 0.47 births. This translates to a reduction in fertility of about 7.2 and 21.9 percent respectively (coefficient over sample mean). Both the OLS and reduced form estimates also show a reduction in the desired number of children as a result of increased female education by 0.1 and 0.18 births representing a 2.3 and 4.1 percent reduction respectively. Lastly, the naïve OLS estimates also show that an additional year of schooling increases the age at first birth by about 0.24 years while the reduced form effect is insignificant.

Table 4.4: Second Stage (2SLS), Reduced form and OLS Estimations

	Full sample		**Poor Subsample**		**Wealthy subsample**		**OLS**	**Reduced form**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Controls		Yes		Yes		Yes	Yes	Yes
Region Fixed Effects		Yes		Yes		Yes	Yes	Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel A: Total children born								
Schooling instrumented by: 1[Birth year ≥ 1988]	-0.252***	-0.259***	-0.28***	-0.321***	0.198***	-0.288***	-0.154***	-0.468***
	(0.013)	(0.0198)	(0.033)	(0.039)	(0.021)	(0.063)	(0.006)	(0.043)
Mean of dep. var.	[2.14]		[2.93]		[1.38]		[2.14]	
Observations	16814		8243		8571		16828	16828
Rsquared	0.12	0.15	0.18	0.18	0.06	0.09	0.19	0.15
Panel B: Ideal number of children								
Schooling instrumented by: 1[Birth year ≥ 1988]	-0.284***	-0.264***	-0.119***	-0.485***	-0.193***	-0.152***	-0.1***	-0.178***
	(0.01)	(0.061)	(0.035)	(0.075)	(0.019)	(0.057)	(0.005)	(0.041)
Mean of dep. var.	[4.38]		[5.32]		[3.86]		[4.38]	
Observations	16268		7918		8350		16275	16290
Rsquared	0.095	0.12	0.024	0.06	0.03	0.04	0.17	0.12
Panel C: Age at first birth								
Schooling instrumented by: 1[Birth year ≥ 1988]	0.244***	0.225***	0.011	0.219**	0.332***	0.287***	0.24***	-0.09
	(0.022)	(0.027)	(0.043)	(0.117)	(0.051)	(0.061)	(0.01)	(0.088)
Mean of dep. var.	[18.24]		[18.07]		[18.93]		[18.24]	
Observations	12579		7326		5253		12593	

Rsquared	0.03	0.034	0.01	0.015	0.015	0.02

*Notes: Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The robust standard errors, which are clustered in survey clusters, are enclosed within parentheses. Significance at the 10% level is denoted by *, 5% level by **, and at the 1% level is by ***. The years of birth are r-centered at the discontinuity (1988). Each column's results are generated through an individual regression. The main explanatory variable (schooling) is instrumented by the binary indicator* 1[Birth year≥ 1988]. *Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff.*

4.4. Discussion

The findings consistently demonstrate that the implementation of the education policy has had a strong positive effect on the number of years of schooling completed by the cohorts affected, with an estimated increase of around 0.7 years. This first stage effect is similar to that of Keats (2018) and Makate (2016) who used the Ugandan DHS and Hahn et al., (2018) who used the Bangladeshi DHS. In the case of Uganda, several studies have examined the impact of the Universal Primary Education (UPE) policy (Keats, 2018; Andriano and Monden, 2019; and Nagashima and Yamauchi, 2020) have reported findings on the effect of this policy on the duration of schooling. Specifically, Keats (2018) found that the UPE policy increased years of schooling by approximately 0.71 years, while Andriano and Monden (2019) observed an increase of 1.5 years. Similarly, Nagashima and Yamauchi (2020) reported a rise of 0.3 years in the duration of schooling as a result of the UPE policy. In the Nigerian context, a number of papers (Osili and Long, 2008; Larreguy and Marshall, 2017; De Cao and La Mattina, 2019) conducted studies that revealed the impact of the Free Primary Education (FPE) policy on educational attainment. Specifically, their findings indicated that the FPE programme resulted in an increase of 1.5, 0.6, and 2.1 years of schooling, respectively. The implementation of the FCUBE policy in Ghana in 1996 resulted in an increase of one year in the duration of schooling (Adu-Boahen and Yamauchi, 2018). In general, the magnitude of the effect is consistent with the magnitudes observed in previous instances.

The magnitude of the effects increases as the bandwidths become narrower, while it decreases as the bandwidths become broader, as anticipated. The effect of the policy remains largely unchanged even after adding predetermined covariates. Equally, the mean of the control cohort group differs substantially between the poor and wealthy subgroups. Women from the poor subgroups without exposure to the UFPE educational policy had 3.6 years of education while their counterparts from the wealthy subgroups had more than double years of education at 7.4 years. Therefore, although the UFPE policy added the same number of years of education for both the poor and wealthy subsamples as the policy was implemented at a national level, it more than doubled the amount of education for the poor relatively. In summary, the first-stage findings suggest that the UFPE educational policy resulted in a large and substantial increase in female educational attainment.

4.4.1. The Effects of Education on Fertility Outcomes

The results are presented in Table 4.4. The results obtained for OLS, RF, and 2SLS analyses consistently indicate that female education reduces both the actual born and the ideal number of children. This relationship holds true for both the full sample and the subset of poor and wealthy subsamples. The three estimates also consistently show that additional female schooling also increases the age at first birth across the full sample in column 2 and also the poor and wealthy subsamples in columns 4 and 6 respectively.

The findings from the RF estimates in column 8 indicate a negative relationship between female education and the total number of children ever born, with a reduction of 0.47 births. Similarly, the 2SLS estimates suggest that for each additional year of schooling, there is a decrease in total births by 0.26. The aforementioned impacts result in a decrease in fertility of about 21.9 and 12.1 percent respectively, as calculated by dividing the coefficient by the sample mean. The findings from the reduced form estimates indicate that female education causes a decrease in the average number of intended children by 0.18. Similarly, the 2SLS estimates predict a drop of 0.264, which corresponds to around 6 percent. Moreover, there exists a correlation between schooling and a rise in the average age at which individuals have their first child, with an approximate increase of 0.23 years, equivalent to a 1.2 percent increment.

The average outcomes we have observed align with the findings of other recent research who have all reported declines in the intended number of children ranging from -0.11 to -0.34, which is consistent with our estimated coefficient of -0.264 (Behrman 2015b; Keats 2018; Masuda and Yamauchi 2020). In contrast to the existing body of research, our paper demonstrates that the observed effect is mostly influenced by women belonging to the lower socio-economic segment of the population as displayed in columns 3 to 6. In comparison to women with lower levels of education, women who received treatment exhibited a reduction of 6 percent in their fertility desire.

Comparing the OLS and 2SLS estimates in Table 4.4, column 2 and column 7 respectively, the 2SLS estimates are larger in magnitude than the corresponding OLS estimates. There exists a potential for bias in the OLS estimates due to the exclusion of variables, such as family preferences. The OLS estimates are prone to an upward bias as a result of unobserved ability. Surprisingly, our estimates are biased downwards. However, this is not strange as most of the studies in the estimation between education and health point to this conclusion (Cutler and Lleras-Muney 2010).

The rationale behind our evaluation lies in the type of the policy intervention, being a supply-side measure designed to enhance accessibility to free education. Due to the nature of our instrumental variable, the UFPE educational policy has a limited impact on women. Consequently, the treatment effect is expected to surpass the estimates obtained through OLS regression. The compliers exhibit distinctive characteristics in comparison to the general population, rendering them more adept at evaluating the trade-offs associated with pursuing more education and engaging in childbirth.

The implementation of supply-side interventions, such as the education reform implemented in Zambia, enables us to assess the impact of education on health outcomes by modifying the educational decisions of individuals who, in the absence of the reform, would typically have limited access to schooling, creating an exogenous source of variation unrelated to an individual's capabilities. However, as Card (2001) has previously argued, it is important to consider the average impact of education on health for the specific group affected by the reform, rather than focusing solely on the average marginal effect of education for the entire population, from a policy perspective. These results are consistent to those found in the context of developing countries by Adu-Boahen and Yamauchi (2018), Keats (2018), Hahn (2018), Aguero and Bharadwaj (2014) and Tequame and Tirivayi (2015).

4.4.2. Heterogeneous effects: Poor versus Wealthy subsample results

The effect of an additional year of schooling on the total number of children ever born in the poor and wealth subsamples is similar although high in magnitude on the poor subsample. The reduction in total children ever born as result of an extra year of schooling is 10.9 percent for the poor subsample compared to 20.9 percent for the wealthy subsample. Regarding the effect of an extra year of female schooling on the ideal number of children, the effect is more than twice higher in the poor subsample compared to the wealthy subsample. Among the subgroup of individuals with lower socioeconomic status, a 9% reduction in fertility preference was observed. Conversely, among the wealthy subgroup, a fall of 4% in fertility preference was noted.

In general, female education has been found to have a significant impact on reducing both the average number of children, both in terms of effectiveness and desired outcomes, as well as increasing the age at which women have their first child.

4.4.3. Potential pathway mechanisms for the full sample

Having shown the causal effect of years of schooling on fertility outcomes, we now examine the underlying mechanisms through which education negatively affects fertility outcomes. Numerous potential pathways have been proposed to establish a connection between female education and fertility outcomes. These pathways include shifts in opportunity costs as outlined in neoclassical models (Becker and Lewis, 1973), the impacts of positive assortative mating (Behrman and Rosenzweig (2002) and knowledge appropriation about health and family planning (Glewwe, 1999).

We examine the potential correlation between female education and the following factors: (i) Age at first marriage (ii) Age at first sex (iii) Assortative mating (iv) Labour market participation (v) Literacy (vi) Knowledge of the ovulation cycle (vii) Access to media and (viii) Contraceptive use. Using an RDD framework, the underlying potential mechanisms are shown in Table 4.5.

Age at first marriage and sex

We explore the potential scenario wherein the pursuit of female education may diminish the prospects for women to participate in the institution of marriage. Column (1) shows that an additional year of schooling delays the age at first marriage by about 0.35 years. The results also show that women who get more schooling are likely to postpone their age at first sexual intercourse by about 0.18 years.

According to Black et al. (2008), individuals who engage in regular attendance at educational institutions are more likely to experience limitations in terms of the time available for engaging in sexual intercourse, as a substantial portion of their time is dedicated to schooling. This can be investigated by directly analysing the impact on abstinence as a potential mechanism. Education necessitates a proportional augmentation in the duration of women's schooling, thus reducing their prospects for engaging in sexual relationships, entering into marriage, or engaging in childbirth.

This phenomenon bears resemblance to the concept known as the "incarceration effect" observed in adolescent females, as discussed by Black et al., (2008).

The results are displayed in panel C column1 and show that the effect of additional schooling on the likelihood of never having sex are statistically significant. Each extra year of schooling reduces the probability of never having sex by 26 percent (7.1 percentage points) in the full sample. Therefore abstinence is a mechanism through which education lowers fertility choices. The heterogeneous effects also confirm abstinence as a potential mechanism with a decrease in the likelihood of ever having sex as a result of an extra year of schooling by 35.2 and 31.8 percent for the poor and wealthy subsample respectively. This finding aligns with the research conducted by Black et al (2008). Our results support the perspective put forth by Black et al. (2008) regarding the "incarceration effect," which suggests that educational institutions serve as a deterrent for women to partake in activities that may result in unintended pregnancies. Our findings contradict the research conducted by Keats (2018) and Adu-Boahen and Yamauchi (2018), which suggests that there is no significant difference in the sexual behavior of women attending school compared to those who are not. Our findings suggest that the postponement of childbearing is attributable to a delay in the onset of sexual activity among females.

Assortative mating

We investigate assortative mating as a potential mechanism by studying the impact of female education on the quality of marriage. When examining the impact of education on marriage quality, it is important to note that individuals marrying are not entirely exogenous. The education of women influences their decision to get into marriage, as well as the decision of their husbands, based on local cultural standards. The results in panel D column (1) also show evidence of positive assortative mating. An additional year of female schooling increases husband's education by about 0.76 years. Women with more education are more likely to marry males with more education. This is consistent with other research findings (Breierova and Duflo, 2004; Lavy and Zablotsky, 2015; Ali and Gurmu, 2018; Hahn, 2018).

Labour market participation

Regrettably, the DHS surveys do not offer comprehensive data regarding individuals' pay and employment records. We can only utilise data regarding the probability of the woman's participation in the labour market by considering her work status during the survey and in the preceding 12 months. Panel E presents data pertaining to the labour market outcomes of women. The labor force participation rate for women is relatively low, standing at approximately 42%. Furthermore, it appears that higher levels of education do not have a significant impact on labour market outcomes. We find no statistically significant effects of additional education on labour force participation in all forms.

Literacy

The results are presented in panel F. The findings indicate that the inclusion of an extra year of education resulted in a 10.9 percentage point increase in women's literacy rates, representing a 20.2 percent increase relative to the mean. This is consistent with the results obtained in similar studies conducted in other developing nations, such as those conducted by Keats (2018) and Behrman (2015a).

Knowledge effect

Education has a crucial role in equipping women with the necessary skills to effectively process information. Education has an impact on fertility due to the acquired capacity to comprehend the implications of high fertility. For instance, women who have received a formal education are more inclined to regularly engage with the media. Consequently, if the media provides detailed explanations about the detrimental consequences of early marriage and teenage pregnancy, it will enhance the level of knowledge among educated adolescent girls. Schooled women will most likely frequent access to media to acquire knowledge related to fertility outcomes, helping them to make better informed decisions. Our results also demonstrate the knowledge effect. An extra year of schooling increases the probability of reading the newspaper, watching television and having knowledge of the ovulation cycle by 4.2, 9.2 and 1.1 percentage points respectively. Lastly, each extra year of schooling increases the likelihood of contraceptive use by 3.5 percent (an increase of 1.2 percentage points of a base of 34 percent). Consequently, educating girls and women about family planning methods is a crucial method for reducing fertility rates.

4.4.4. Heterogeneous effects of education on potential mechanisms

The age at first marriage as a potential mechanism through which female schooling negatively affects fertility is also significant in the poor as well as the wealthy subsamples. According to our results, the wealthy households take longer to marry compared to the poor households. The age at first sex mechanism seems to operate only in the wealthy subsample and is insignificant in the poor households.

The effect of education on labour market participation is insignificant on both the poor and wealthy households. This finding aligns with similar research (Masuda and Yamauchi, 2017). They discover that there is no statistically significant relationship between maternal education and the likelihood of engaging in employment or being employed by an external entity in Uganda.

Regarding literacy, the impact of additional female schooling on individuals from different socioeconomic backgrounds showing the heterogeneous effects of additional schooling between the poor and wealthy subgroups is demonstrated in columns (2) and (3) of panel F. The literacy rate among the poor women experienced a notable increase of 27% as a result of their educational pursuits, whereas wealthy women witnessed a comparatively modest gain of 11%.

It is important to acknowledge that there exists a notable disparity in literacy rates between the poor and wealthy women. Specifically, the literacy rate among the wealthy women stands at 79%, but it is comparatively lower at 49% among the poor women as shown in the descriptive statistics in Table 4.5. This implies that the provision of free schooling effectively equipped individuals from low-income backgrounds with fundamental skills, perhaps obtained at earlier stages of education. Finally, access to media through reading newspapers seems to only operate in the wealthy subsample and not in the poor subsample.

In conclusion, given that the outcomes of this research are tailored to the participants within the chosen optimal window, on either side of the cut off, and affected by the UFPE education policy, as per the study's design, an examination of whether there exists a sustained impact on fertility would be of interest. Therefore, conducting a comprehensive investigation into the long-term returns of the UFPE educational policy would be a valuable avenue for future research.

Table 4.5: Potential Mechanisms – Fertility related outcomes

	Full Sample	Poor	Wealthy
	(1)	(2)	(3)
Panel A: Age at first Marriage			
Schooling instrumented by:	0.348***	0.335***	0.552***
1[Birth year ≥ 1988]	(0.056)	(0.061)	(0.117)
Mean of dependent var.	18.15	17.6	18.9
Observations	11647	6841	4750
Rsquared	0.05	0.01	0.026
Panel B: Age at first sexual intercourse			
Schooling instrumented by:	0.314***	-0.004	0.156*
1[Birth year ≥ 1988]	(0.025)	(0.051)	(0.082)
Mean of dependent var.	16.557	16.1	17.1
Observations	12275	6112	6112
Rsquared	0.053	0.003	0.031
Panel C: Abstinence (=1)			
Schooling instrumented by:	-0.071***	-0.088***	-0.089***
1[Birth year ≥ 1988]	(0.007)	(0.009)	(0.012)
Mean of dependent var.	0.27	0.25	0.28
Observations	16814	8171	8571
Rsquared	0.014	0.024	0.022
Panel D: Husband education			
Schooling instrumented by:	0.756***	0.667***	0.677***
1[Birth year ≥ 1988]	(0.0075)	(0.086)	(0.119)
Mean of dependent var.	8.13	6.7	10.12
Observations	10854	6073	4355
Rsquared	0.22	0.055	0.072
Panel E: Currently working			
Schooling instrumented by:	-0.04	0.004	-0.001
1[Birth year ≥ 1988]	(0.01)	(0.013)	(0.014)
Mean of dependent var.	0.47	0.53	
Observations	16814	8171	8571
Rsquared	0.02	0.01	0.025
Panel F: Literacy			
Schooling instrumented by:	0.109***	0.132***	0.0844***
1[Birth year ≥ 1988]	(0.008)	(0.01)	(0.01)
Mean of dependent var.	0.54	0.49	0.79
Observations	16814	8171	8571
Rsquared	0.12	0.06	0.079
Panel G: Knowledge of ovulation			
Schooling instrumented by:	0.011*	-0.004*	0.017***
1[Birth year ≥ 1988]	(0.002)	(0.002)	(0.003)
Mean of dependent var.	0.89	0.98	0.98

Observations	16814	8171	8571
Rsquared	0.001	0.01	0.028
Panel H: Watches Television			
Schooling instrumented by:	0.092***	0.001	0.109***
1[Birth year ≥ 1988]	(0.003)	(0.007)	(0.0126)
Mean of dependent var.	0.41	0.13	0.67
Observations	16814	8171	8571
Rsquared	0.32	0.01	0.18
Panel I: Reads Newspaper			
Schooling instrumented by:	0.042***	0.001	0.081***
1[Birth year ≥ 1988]	(0.008)	(0.008)	(0.014)
Mean of dependent var.	0.3	0.17	0.43
Observations	16814	8171	8571
Rsquared	0.083	0.01	0.04
Panel J: Listen to Radio			
Schooling instrumented by:	0.013	-0.025**	0.033**
1[Birth year ≥ 1988]	(0.008)	(0.011)	(0.013)
Mean of dependent var.	0.59	0.48	0.71
Observations	16814	8171	8571
Rsquared	0.047	0.003	0.031
Panel K: Contraceptive Use			
Schooling instrumented by:	0.012***	0.014	0.059***
1[Birth year ≥ 1988]	(0.03)	(0.008)	(0.005)
Mean of dependent var.	0.34	0.36	0.31
Observations	16814	8171	8571
Rsquared	0.004	0.002	0.023

*Notes: Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. Robust standard errors clustered at the survey level are presented in parentheses. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

4.5. Robustness and validity

4.5.1. Internal Validity

In order to utilize the UFPE educational policy as a quasi-experiment for the purpose of obtaining policy estimates, it is necessary to address several concerns. A significant issue arises from the possibility that parents may possess prior knowledge regarding the year and timing of the UFPE educational policy, which they could exploit to delay their children's enrollment in primary school in order to take advantage of the policy. However, it is improbable that this is the scenario, given that the UFPE educational policy was a purposeful, nationwide policy intended to be uniformly implemented with the objective of granting free primary education to all individuals, irrespective of their grade level or social standing.

Another issue is women who were close to the 7-year primary grade 1 beginning age. These women were 8–10 years when the UFPE policy was implemented. Our cutoff places these in the control cohorts, although the UFPE program allowed dropouts and late entrants to benefit from the free primary education policy (fuzzy RDD). If school dropouts and late entries are considerable, the whole impact of the UFPE educational strategy would be bigger than our estimate. Late entrants and dropouts who returned to school cannot be quantified using DHS data.

Another important consideration is the exclusion restriction. This implies that there is a probable correlation between the policy at specific thresholds and other government policies. In light of the presented scenario, our causal estimation would comprise the UFPE educational policy alongside other governmental policies. Based on the currently available information, it appears that no other policy has been implemented that possesses identical eligibility criteria for treatment as the UFPE educational policy. Furthermore, there is no evidence to suggest that any other policy has exclusively affected a specific subset of observations within the optimal bandwidth.

The inclusion of local randomisation is a fundamental component of the RDD. Therefore, outcome determinants, apart from the treatment (education), must undergo continuous evolution (smoothing assumption) at the cut-off, encompassing women born prior to, during, and subsequent to the cut-off years (Jacob et al., 2012). In order to evaluate the validity of the smoothing assumption, (exclusion restriction), within the RDD, we analyze the pretreatment covariates as suggested by Imbens and Lemieux (2008) and Jacob et al. (2012).

Acquiring suitable variables from survey data for the purpose of conducting hypothesis testing of this nature is difficult. The dataset identifies religion and region as the most significant pretreatment covariates. Religion is typically regarded as a relatively stable trait, with individuals frequently adhering to the religious beliefs of their families. Region (place of residence) is equally difficult to change over a short period of time. In instances where treatment allocation is randomized, it is expected that there will be no observable variations among participants with regard to pre-existing variables. The results are shown in Table 4.6. In line with our apriori expectations, coefficients associated with the predetermined covariates lack statistical significance. Therefore, it can be posited that the process of sample selection and treatment assignment was devoid of bias.

Finally, in order to assess the accuracy and reliability of our estimations, we investigate the running variable for potential manipulation. The test utilised in this study was developed by McCrary (2008) and is designed to assess the continuity of the running variable within the framework of the RD design. The underlying principle of the manipulation testing in this specific context is that if there is no intentional and systematic manipulation of the unit's index near the cutoff point, the distribution of units should demonstrate continuity in the vicinity of this threshold value. Therefore, a manipulation test is devised with the purpose of formally establishing the existence of a discontinuity in the density of units at a predetermined threshold. The presence of such evidence is frequently interpreted as empirical evidence suggesting the occurrence of self-selection or nonrandom assignment of units to control and treatment status. The findings are displayed in Table 4.7. The plot illustrating the outcome is depicted in Figure 4.4.

***Table 4.6*: Pretreatment covariates**

	Region	Religion
Coefficient	0.088	0.123
t	1.29	0.53
$Pr(\|T\| > \|t\|)$	0.198	0.598

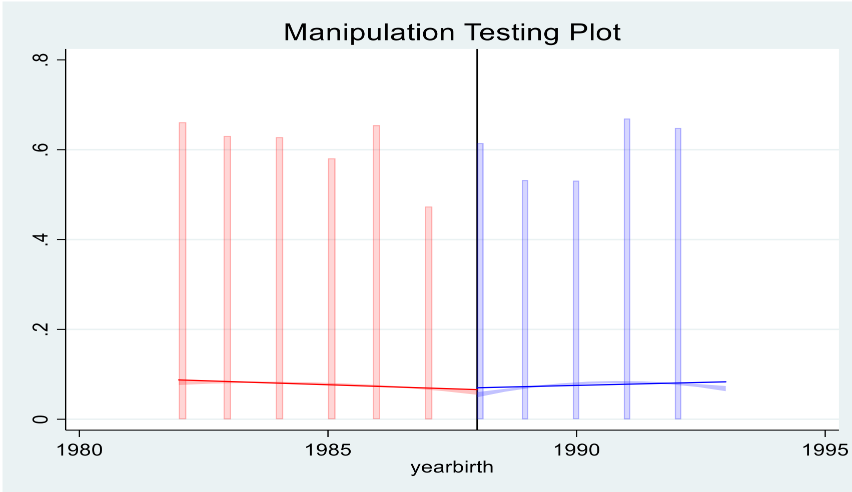
Notes: This table evaluates the hypothesis that there are no statistically significant variations in the pretreatment variables among respondents located on either side of the cut-off. The fundamental assumption is that when treatment allocation is randomised, there should be no observable disparities among participants in terms of pre-treatment variables. The data presented provides support for the validity of the hypothesis.

Table 4.7: Manipulation Test of running variable

Method	T	P>|T|
Robust	0.096	0.9235

Notes: The outcome of the ultimate manipulation test indicates a T- value of 0.096, accompanied by a p-value of 0.9235. We fail to reject the null hypothesis of no manipulation. Therefore the results lack statistical proof of any deliberate manipulation of the running variable.

Figure 4.4: Manipulation Testing Plot



This figure plots the density of the running variable against the year of birth. At the cutoff, there are no visible signs of any manipulation of the running variable. In other words, there is no bunching of the running variable at the cutoff.

4.5.2. External Validity

The limitation of our findings is that they can only be effectively interpreted within a specific optimal time frame of 6 years before and after the designated cutoff. Nevertheless, within the RDD framework, this issue is commonly encountered. However, the results of this study hold significance within the African and other nations that have yet to adopt UFPE-type policies and are seeking to attain the advantages associated with this educational policy.

4.6. Robustness Checks

The assumptions underlying our instrumental variable estimates require that women born slightly before and slightly after the cut-off do not exhibit systematic differences in factors influencing fertility outcomes, except for their varying levels of educational exposure. This implies that the women on both sides of the threshold exhibit similar characteristics, indicating that the inclusion of control variables would not have an impact on the outcomes. To clarify, the sole distinction between these two cohorts lies in their attendance or non-attendance of primary education.

In order to enhance the robustness of our findings, we conduct additional analyses by estimating the results using different functional forms and varying bandwidths. Based on the findings presented in Table 4.8, the statistical significance of the results remains unchanged when employing various functional forms for estimation. The first stage results are re-estimated using bandwidths of four, six, eight, and ten years on both sides of the cutoff. Even when considering a relatively low bandwidth of four years on either side of the cutoff, the re-estimated findings indicate a rise in the impact of the UFPE policy on educational achievement. The optimal findings indicate that the UFPE educational policy resulted in a significant rise in the average number of years of completed education, with an approximate increase of 0.7 years. Consequently, the findings remain consistent when considering birth year cohorts on either side of the cutoff.

Table 4.8: First stage estimates with different bandwidths

	Dependent variable: Years of schooling							
	4 year bandwidth		6 year bandwidth (optimal bandwidth)		8 year bandwidth		10 year bandwidth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1[Birth year ≥ 1988]	0.651***	0.579***	0.802***	0.677***	0.654***	0.566***	0.826***	0.703***
	(0.122)	(0.0106)	(0.107)	(0.092)	(0.1)	(0.086)	(0.089)	(0.078)
Mean of dep. var.	[7.05]		[7.03]		[7.25]		[7.25]	
F-statistic	28.45	30.14	55.14	53.89	42.76	41.83	86.89	81.17
Observations	10959		16823		19460		23999	
Controls	No	Yes	No	Yes	No	Yes	No	Yes

*This table presents re-estimated first stage results. Columns (1), (3), (5), and (7) are devoid of control variables, while columns (2), (4), (6), and (8) incorporate control variables for place of residence, region and religion. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The standard errors have been clustered at the survey clusters and are presented in parentheses, ***p < 0.01, **p < 0.05, *p < 0.1.*

Furthermore, alongside the selection of the varying bandwidths, we evaluated the impact of different functional forms on our findings. This additional analysis aimed to enhance the reliability and stability of our results. The primary equations (4) is adjusted to include higher order polynomials of the forcing variable, thereby enabling the consideration of the interaction between the running variable (centered at the cutoff) and the treatment indicator, denoted as $Z_i = 1[Year\ of\ Birth \geq c(1988)]$. In all of our re-estimations, we consistently maintain a bandwidth of six that is optimised according to the mean squared error (MSE) criterion. The findings, as estimated using the primary sample, are presented in Table 4.9. Throughout the specifications, our estimates demonstrate a consistent level of robustness in terms of both the magnitude of the effects and their statistical significance.

Table 4.9: Main Results with different functional forms

	Linear	Linear Interaction	Quadratic	Quadratic Interaction	Cubic	Cubic Interaction
Total children	-0.223***	-0.225***	-0.227***	-0.235***	-0.232***	-0.238***
ever born	(0.021)	(0.021)	(0.02)	(0.02)	(0.021)	(0.02)
Observations	16814	16814	16814	16814	16814	16814
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Ideal number	-0.325***	-0.327***	-0.329***	-0.333***	-0.332***	-0.335***
of children	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)
Observations	16268	16268	16268	16268	16268	16268
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age at first	0.223***	0.229***	0.232***	0.234***	0.233***	0.241***
birth	(0.28)	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)
Observations	12579	12579	12579	12579	12579	12579
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents re-estimates of the main results using different functional forms. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. All specifications include survey fixed effects. The estimates remain robust to varying functional forms. Robust standard errors have been clustered at the survey clusters and are presented in parentheses $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

4.7. Conclusion

This paper examines the causal relationship between female education and fertility by utilising Zambia's free primary education policy (UFPE) of 2002 as a quasi-experiment. We use a plausibly exogenous increase in female schooling by utilising a fuzzy (RDD) to establish a causal link between education and fertility while exploring the underlying mechanisms. Our results show that the UFPE educational policy had a positive impact on female schooling. We also find heterogeneous effects of the effects of the UFPE education policy by socioeconomic status. The poor households exhibit large increases in female schooling compared to their wealthy counterparts. This points to the conclusion that the UFPE education policy positively impacted the poor more than it did for the wealthy households, offering significant evidence of heterogeneity in treatment influenced by wealth status.

Additionally, this exogenous surge in female schooling leads to a decrease in fertility and fertility preferences, while simultaneously resulting in a delay in the age at which individuals have their first child. We find heterogeneous effects of education on fertility and fertility preferences by socioeconomic status. Specifically, it has been shown that women from poor households and residing in rural areas exhibit more significant alterations in their fertility patterns and preferences, specifically in terms of the magnitude of the impact.

Given that the UFPE educational policy affected female education at the primary level, this implies that the attainment of primary years of schooling may have a greater impact on changing women's fertility behaviour and related characteristics, compared to the schooling achieved at later stages of education. Although wealthier women tend to attain at least some primary education, women from lower socioeconomic backgrounds in developing nations may not receive even minimal levels of learning without such policy efforts. This implies that, in addition to promoting increased education, Free Primary Education (FPE) policies can also have a role in influencing changes in fertility patterns, particularly among the poor.

This paper further uncovers the potential pathways through which female education affects fertility. Our evaluation suggests that education affects fertility through positive assortative mating. Educated women are more likely to marry males with more education. Further, female schooling affects fertility through delayed the age at first marriage, first sexual intercourse and also through increasing the probability of contraceptive use.

Our results also demonstrate the knowledge effect. More educated women are likely to frequently access the media to acquire knowledge related to fertility outcomes, helping them to make better informed decisions. Increased female education also results in elevated levels of literacy. Insufficient empirical support exists to establish a robust correlation between additional female education and labour market outcomes in Zambia, potentially attributable to the presence of structural labour market limitations and suboptimal levels of labour force engagement.

The high fertility rate in Zambia, along with the prevalence of child marriage, continues to present substantial threats to the well-being of women and children in one of the poorest countries on Earth. The results of our study indicate that expanding the availability of female education can serve as a viable policy measure to reduce fertility rates. These findings are especially significant in numerous developing nations that are affected by the problem of early child marriage and underdevelopment.

CHAPTER 5: The effect of female education on HIV seroprevalence status and knowledge in Zambia: Universal Primary Education as a natural experiment

5.1. Introduction

While human immunodeficiency virus (HIV) infections are declining in many nations, it is worth noting that around 730,000 people in Eastern and Southern Africa contracted the virus in 2019 (UNAIDS, 2020). HIV/AIDS poses a significant health threat to young women in Sub-Saharan Africa (SSA). HIV remains a significant global health concern, as evidenced by an approximate 1.3 million individuals who were infected in the year 2022 (UNAIDS 2023).

HIV/AIDS presents an unparalleled obstacle to development and human well-being, particularly in Sub-Saharan Africa. The epidemic has significantly reduced life expectancy and caused the loss of millions of individuals in their most productive years in numerous countries. It has extinguished the optimism of leading fulfilling and fruitful lives for many infants, children, and young adults.

Over the past decade, Zambia has made progress in the HIV response. According to UNAIDS (2020), annual HIV infections (for all ages) in Zambia have declined from 60,000 in 2010 to 51,000 in 2019. Despite the progress, the HIV burden remains high and disproportionately affects females. In 2019, it was estimated that there were 26,000 new HIV infections among women 15+ years, compared to 19,000 among their male counterparts. The Zambia Demographic and Health Survey (ZDHS 2018) reports that HIV prevalence among females aged 15-49 years is 14.2 per cent, compared to 7.5 per cent for males of the same age.

Given the significant challenge of providing antiretroviral treatment to an increasingly large population of individuals infected with HIV, it is imperative for international efforts to prioritize prevention strategies in addressing the HIV epidemic. Consequently, the implementation of formal education, with a particular emphasis on the education of girls, has been widely recognized as a preventive measure to mitigate the transmission of HIV within society. It is widely believed that increased education can equip individuals with the necessary knowledge and skills to mitigate the risk of transmission through the adoption of preventive measures such as condom usage and other precautionary behaviors. Nevertheless, there is a scarcity of causal evidence to support this assertion (Vandemoortele J. and Delamonica E., 2000).

This article examines the correlation between women's educational achievement and their vulnerability to HIV infection. The correlation between socioeconomic status and health outcomes, specifically measured through educational attainment, is widely acknowledged among researchers (Grossman, 2006; Montez and Friedman, 2015). However, the connection between education and the risk of contracting HIV is a subject of considerable debate.

Education has the potential to mitigate the risk of HIV transmission by facilitating greater exposure to knowledge regarding HIV and its prevention strategies (Aguero and Bharadwaj, 2014; De Walque, 2009). Education also has the potential to mitigate the risk of HIV transmission by enhancing cognitive abilities necessary for making informed decisions and fostering improved financial stability (Psacharopoulos, 1994; Schultz, 1960; Willis, 1973). Education has the potential to mitigate the risk of HIV transmission by reducing women's engagement in transactional sex and enhancing their ability to form relationships with partners who have a lower risk of transmitting the virus (Baird et al., 2012; Behrman and Rosenzweig, 2002; Rosenberg et al., 2015). Nevertheless, it is plausible that education could potentially lead to an expansion in an individual's sexual network, a lengthening of the duration of premarital sexual activity, and an escalation in transactional sex among males (Kohler and Thornton, 2011; Case and Paxson, 2013).

The difficulty lies in determining the causal relationship between schooling and the risk of HIV infection due to the strong association between educational attainment and various factors, such as socioeconomic status, psychological traits and selection on unobservable factors like cognitive ability and family preferences. These factors pose challenges in observational studies as they are complex to fully control for and can also influence the risk of HIV infection. A number of studies in Sub-Saharan Africa (SSA) have established causal interpretations between schooling and HIV risk (Hargreaves et al., 2008; Alsan and Cutler, 2013; Behrman, 2015; De Neve et al., 2015; Dupas, 2011; Aguero and Bharadwaj, 2014; Gummerson, 2013; Gummerson et al., 2013; Duflo et al., 2015; Lindskog and Durevall, 2021). This paper therefore, adds to this growing list of literature establishing causal interpretations between schooling and HIV seroprevalence status and investigates the potential mechanisms through which schooling affects HIV related outcomes.

Research has shown that there is a negative relationship between educational achievement and HIV infection risk. It has been argued that education functions as a "social vaccine" in preventing the transmission of the illness (Behrman, 2015a, 2015b; De Neve, Fink, Subramanian, Moyo, &

Bor, 2015a,b). Nevertheless, despite significant advancements in increasing the availability of formal education in recent years, there has not always been proportional enhancement in important sexual and reproductive health measures (Bongaarts & Casterline, 2013; Mensch, Singh, & Casterline, 2005).

Establishing causality between education and HIV seroprevalence status, is a complex task that poses significant challenges. Consequently, there is a scarcity of scholarly literature that investigates potential mechanisms that could potentially connect these two variables. Few papers establish causality in Sub-Saharan Africa. According to Hardee et al. (2014), several studies conducted in Sub-Saharan Africa have demonstrated a negative correlation between educational attainment and the likelihood of contracting HIV. As depicted, establishing a causal interpretation presents challenges. Nevertheless, several studies have employed diverse methodologies in order to establish causal interpretations.

Aguero and Bharadwaj (2014) find no statistically significant association between education and HIV seroprevalence status. On the other hand, Duflo et al. (2015) using an HIV information campaign that emphasized abstinence, mitigated the prevalence of sexually transmitted infections. Behrman (2015) finds that an additional year of education is associated with a reduced likelihood of adult women testing positive for HIV in both Malawi and Uganda. De Neve et al. (2015) report a decrease in the overall risk of HIV infection for each additional year of secondary schooling.

To analyze this correlation from a causal perspective, we employ an instrumental variables methodology that leverages age-specific exposure to an educational reform implemented in Zambia. We take advantage of the enactment of the Universal Free Primary Education (here after UFPE) policy by the Zambian government in 2002. We use this exogenous source of variation in educational attainment, created by this age – specific nature of the reform to address the endogeneity problem of schooling thereby estimating the causal impact of schooling on HIV seroprevalence status and other outcome of interest such as HIV knowledge and number of sex partners.

We contribute to the literature by adding a new piece of evidence to the few papers done on the causal findings between educational attainment and HIV related outcomes in SSA. We provide a new piece of evidence from Zambia, a country with high prevalence of HIV in SSA (National prevalence of about 11% among adults i.e.15-59 years (UNAIDS, 2020). The paper also adds to the few articles that examine causality between educational attainment and HIV related outcomes.

5.2. Data and Empirical Strategy

5.2.1. Data Sources

This study uses data from three rounds of the Zambia Demographic and Health Surveys (ZDHS), which are nationally representative surveys of reproductive-aged women conducted in Zambia in 2007, 2013–2014, and 2018. In addition these three rounds, of the ZDHS also collect vital information relating to HIV status and knowledge. For HIV status, HIV tests were also administered to all men and women in the household between the ages of 15 and 49, and blood sample on eligible men and women was collected in these surveys households from rural and urban areas. The DHS sample is drawn in two stages: first, enumeration regions are randomly selected from census files, and then households are randomly drawn from the list of households within each enumeration area (ZDHS, 2018).

5.2.2. Empirical Strategy

To examine the impact of female educational attainment on HIV related outcomes, we consider the following OLS equation (1):

$$H_i = \psi + \beta Edu_i + \lambda \theta_i + \varepsilon_i \qquad (1)$$

The variable H_i is an indicator of HIV-related behaviour (HIV seroprevalence status, number of sexual partners or HIV transmission knowledge) for individual i and Edu_i denotes the educational achievement of individual i, measured by the number of years of schooling completed. The vector θ_i is a combination of control variables and other characteristics such as region and survey fixed effects. Therefore, our interest is parameter β. Employing OLS to estimate this parameter is prone to an upward bias as a result of potential reverse causality and unobservable variables (individual ability) that impact both education and the HIV related outcome variables. For instance, the relationship between promiscuity and schooling might be attributed to factors such as economic levels, individual preferences, and societal standards. These complex variables pose challenges in accurately determining the bias in OLS regression analysis. Equally, it can lead to a downward bias due to a potential error in the measurement of educational attainment.

One potential approach for mitigating these biases involves introducing a source of exogenous variation in schooling, which may subsequently be utilized to establish a causality. In order to address this endogeneity issue, we employ the adoption of the 2002 UFPE educational policy in Zambia. This policy resulted in a dramatic rise in both enrolment and retention rates causing an exogenous source of variation in educational attainment. This exogenous spike in enrollment resulting from the UFPE policy aligns with a regression discontinuity design, which is used to identify the causal impact of education on HIV related outcomes. Therefore, we examine the cohort of women who were born at a time that allowed them to take advantage of the UFPE program, and compare them to a group of women who were born just a few years prior and so did not have the opportunity to benefit from the policy. Utilizing the year of birth rather than the year of initial enrolment in primary school is expected to yield more dependable outcomes, as the latter is influenced by a multitude of factors, including latent variables that impact both educational duration and subsequent achievements.

Considering the fact that the rating variable in our study is determined by the year of birth, it is improbable that the treatment assignment is solely determined by this variable. In addition, the treatment effect, which involves an extra year of primary education, is found to be less than 1 for those born at the cutoff year. This characteristic aligns with the concept of a fuzzy RDD, (Lemieux and Imbens, 2008). Therefore, the fuzzy RDD framework is employed to quantify the impacts of schooling.

The estimand in a fuzzy RDD is similar to an IV estimate and is equivalent to the 2SLS estimator in an IV framework. The causal effect of an extra year of schooling in the fuzzy RDD setting is defined as the "difference in outcome from a regression on the treatment-determining variable (birth year) divided by the difference in treatment (an extra year of primary schooling) from a regression on the treatment-determining variable, both evaluated at the cutoff value, c" (Hahn, Todd, and Klaauw 2001; Imbens and Lemieux 2008; Lee and Lemieux 2010) as follows:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} \mathrm{E}\left[H \mid birthyear\right] - \lim_{x \uparrow c} \mathrm{E}\left[H \mid birthyear\right]}{\lim_{x \downarrow c} \mathrm{E}\left[educ \mid birthyear\right] - \lim_{x \uparrow c} \mathrm{E}\left[educ \mid birthyear\right]} \qquad (2)$$

According to Hahn et al. (2001) and employing the dichotomous indicator: $Z_i = 1[Year\ of\ Birth \geq c(1988)]$, the fuzzy RDD estimand is identical to the 2SLS estimator β_{1RD} as follows:

$$Educ = \alpha_0 + \alpha_1 Z + f(Birthyear - c) + \Pi_1 \Gamma + \varepsilon \qquad (3)$$

$$H = \beta_0 + \beta_{1RD} educ_i + g(Birthyear_i - c) + \Pi_2 \Gamma + u \qquad (4)$$

The functions $f(.)$ and $g(.)$ are polynomials and equations (3) and (4) correspond to the first and second stage regressions in the 2SLS framework, respectively. The vector Γ corresponds to pretreatment covariates. The polynomial order and the bandwidth have to be similar for both the first and second stages of the 2SLS framework. By imposing the condition that $f(.)$ and $g(.)$ are of the same order, the resulting standard errors are identical to those in a 2SLS framework.

5.3. Results

5.3.1. Data and Descriptive Statistics

Table 5.1 displays the descriptive statistics. The table provided presents statistical data for the complete sample, as well as for two sub-samples classified as “poor” and “wealthy”. The sub-samples were generated by the process of separating the household wealth into two discrete categories. Of the five categories in the wealth variable, the poorest, poorer and middle categories are reclassified as “poor”. The richer and richest categories are reclassified as “wealthy”.

According to the data presented in Table 5.1, the average number of years of schooling for the control group in the complete sample is 6.5 years. As anticipated, the treatment group demonstrates a greater average educational attainment of 7.6 years in comparison to the control group. The wealthy subsample demonstrates a greater number of years of schooling in both the control and treatment groups, in comparison to their less affluent counterparts. In relation to the two subsamples, the treatment cohort of the poor subsample, who have had access to free primary education (specifically those born in or after 1988), demonstrates a significant and immediate enhancement in educational accomplishments when compared to the control cohorts born prior to 1988. The presence of a discontinuity, or leap in educational attainment is similarly demonstrated in Figure 5.1. The full sample reveals that approximately 61 percent of the female population possesses literacy skills, and further indicates that husbands tend to have a higher level of educational attainment than their wives. Approximately 51 percent of the sample population resides in rural areas. The average desired number of children is approximately 4, although the average total number of children ever born is approximately 2. Finally, the table additionally presents essential demographic data pertaining to the average age at survey, age at first sex, age at first birth, and age at first marriage.

Table 5.1: Summary Statistics

	Full Sample		Poor		Wealthy	
	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)	Control Cohort (born from 1982 -1987)	Treatment Cohort (born from 1988 - 1993)
Number of observations	9156	7667	4014	4235	5085	3489
Years of completed education	6.52 [3.97]	7.64 [3.59]	4.62 [3.01]	5.98 [3.09]	8.03 [3.89]	9.62 [3.09]
Complete Primary	0.51	0.39	0.65	0.58	0.4	0.19
HIV risk (HIV positive =1)	0.172	0.115	0.132	0.09	0.227	0.145
Number of sexual partners	2.16	2	2.12	2	2.2	2
HIV knowledge: Condom use	0.79	0.8	0.82	0.79	0.76	0.81
HIV knowledge: Transmission during childbirth	0.81	0.83	0.85	0.81	0.78	0.85
HIV knowledge: HIV transmitted by mosquito bite?	0.63	0.7	0.57	0.59	0.68	0.81
HIV transmission mitigated by sex with 1 partner?	0.89	0.91	0.93	0.91	0.86	0.93
Literacy: able to read sentence	0.49	0.61	0.3	0.45	0.63	0.8
Age at first sex	16.55 [2.74]	16.6 [2.48]	16.07 [2.31]	15.97 [2.12]	16.94 [2.99]	17.37 [2.66]
Age at first marriage	18.26 [3.66]	17.98 [3.06]	17.74 [3.32]	17.45 [2.79]	18.91 [3.93]	18.95 [3.29]
Residence (Urban = 1)	0.44	0.48	0.17	0.18	0.66	0.84

Wealth Index combined (1-5)	3.12 [1.41]	3.16 [1.43]	2.01 [0.82]	2.02 [0.82]	4.51 [0.5]	4.53 [0.5]

Notes: The study employs data from the, 2007, 2013/2014, and 2018 Zambia DHS Female Surveys. The analysis utilises a bandwidth size of 6 birth year cohorts on both sides of the cut-off. The subsamples denoted as 'poor' and 'wealthy' comprise of female individuals whose wealth levels are either below or above the median, respectively. The statistical analysis incorporates the DHS sample weights and takes into consideration the pooling effect across survey rounds.

5.3.2. Bandwidth Selection

Following the recommendations by Abadie et al. (2017) and Nunn (2010) standard errors are clustered at the level of survey clusters. This approach to clustering yields standard errors that are robust against arbitrary patterns of within-cluster variation and covariation (Cameron and Miller, 2015). The standard errors are clustered at the survey cluster level since this is the level at which units in the sample were selected and there are clusters in the population that are not represented in the sample. Clustering at survey clusters is useful because while respondents are identical across cohorts, their characteristics may differ by location; in the context of this study, social and economic status can vary much within and between clusters. A section of the literature also clusters at the level of running variables. The point estimates remain robust to clustering at the level of the running variable. The strength of the instrument, however, becomes sensitive to alternative bandwidths in this approach.

Recent academic literature on the regression discontinuity design supports the preference for employing local linear regression estimation instead of global polynomial estimation. We perform an estimation of a local linear regression model on both sides of the threshold, allowing for variation in the slopes on either side of the threshold (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). By employing the 2SLS method, we estimate the Local Average Treatment Effect at the Cutoff (LATE) specifically for individuals who comply with the treatment assignment.

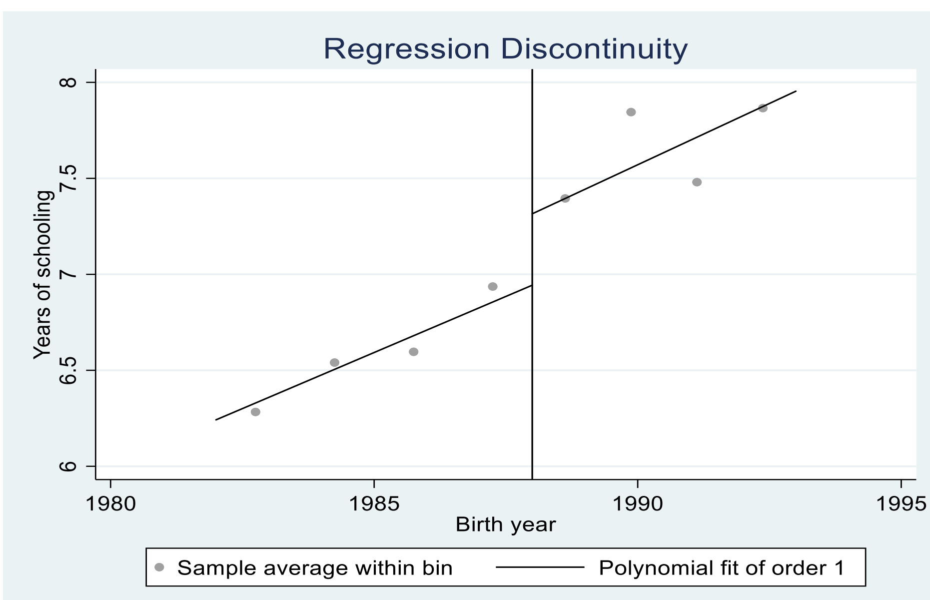
In order to determine the most suitable bandwidth, we employ the mean squared error (MSE-optimal) bandwidth method, utilizing the local polynomial bandwidth selection method proposed by Cattaneo et al. (2014; 2017). The proposed methodology yields an optimal bandwidth of 6.31 years on both sides of the cutoff. Hence, all estimations in the first and second stage regressions were performed with a bandwidth of 6 years on both sides of the cutoff.

5.3.3. First Stage Discontinuity in educational attainment

The initial step involves presenting empirical evidence that demonstrates the positive effects of the educational reform implemented at UFPE. These outcomes are primarily reflected in the increased educational attainment, as seen by a rise in the number of years of schooling. The increase in educational achievement is demonstrated in Figure 5.1 which presents empirical evidence that substantiates the hypothesis positing that the adoption of the UFPE policy exerted a substantial influence on the educational achievement of females.

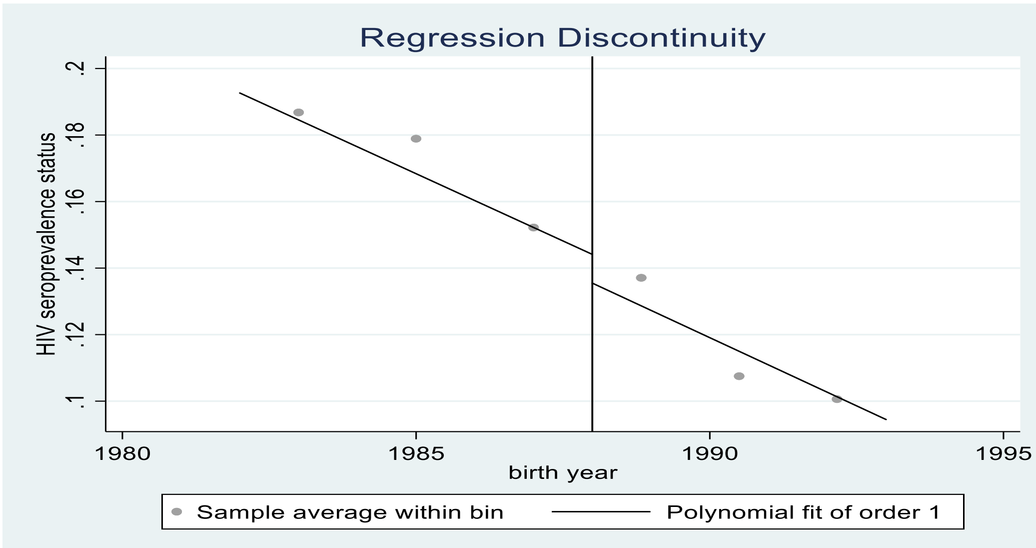
The findings depicted in Figure 5.1 demonstrate that females who were provided with unrestricted access to primary education without any associated costs, particularly those born in or after 1988, had a noteworthy and immediate enhancement in their number of primary schooling years when compared to cohorts born prior to this period. The correlation between birth year cohorts and the educational achievement of females indicates that as cohorts progress beyond the cutoff point, there is a noticeable increase in educational attainment. This can be attributed to the fact that girls born in later years have access to a higher number of years of primary education without incurring any fees. In contrast to years of female schooling, there are no noticeable discontinuities for HIV seroprevalence status by birth year cohort as shown in Figure 5.2.

Figure 5.1: Women's Schooling by Birth Year Cohort



Source: Pooled data from the 2007, 2013/14, and 2018 DHS surveys

Figure 5.2: HIV Seroprevalence status by birth year cohort



Source: Pooled data from the 2007, 2013/14, and 2018 DHS surveys

5.3.4. First Stage Results – Effects of the UFPE policy on years of schooling

The first stage estimates are presented in Table 5.2. The findings provide empirical support for a jump at the discontinuity. The results are provided individually for the entire sample (columns 1 and 2), the poor (columns 3 and 4) and the wealthy (columns 5 and 6) subsamples. In relation to the discourse on outcomes, we allude to the more stringent specification that incorporates survey fixed effects and all covariates. The UFPE educational reform implemented in 2002 had a noticeable influence on educational achievement, particularly among individuals born in 1988 who constitute the initial cohort to have experienced the effects of this reform. As a result, a discernible surge (jump in schooling years) in years of formal education is observed at the point of discontinuity, accompanied by an upturn in educational achievement among females born in and after 1988.

The first stage findings suggest that female participants who were assigned to the treatment group or experienced positive outcomes from the UFPE policy had a mean increase of 0.67 years in educational attainment compared to their counterparts in the control group who were not assigned to the treatment. For the full sample, the jump in the discontinuity is economically large and statistically significant. The first stage F-statistic, which measures the strength of the instrumental variable, exceeds the threshold of 10, indicating that the instrument is not weak. This finding suggests a robust first stage, as noted by Staiger and Stock (1997).

In the poor subsample, the UFPE educational policy resulted in an increase of 0.81 in the years of schooling. Conversely, in the wealthy subsample, the effect of the policy was a 0.96-year increase in years of schooling. The first stage impact exhibits a significantly greater magnitude within the wealthy subsample in comparison to the poor subsample. This may be attributed to the fact that while the educational policy implemented by UFPE had a nationwide scope and was intended to benefit individuals irrespective of their socioeconomic background, it mostly served as an additional incentive for the wealthy subsample who were already financially secure.

The first stage impact revealed in this study has resemblance to the findings of Keats (2018) and Makate (2016) in Uganda, Ali and Elsayed (2017) in Egypt, and Makate (2016) in Malawi. The impact of the policy stays predominantly unaltered even upon the inclusion of predetermined covariates. Similarly, there is a significant disparity in the mean of the control cohort group between the poor and wealthy subsamples. Women belonging to the poor subsamples who did not have access to the UFPE educational policy had an average of 5.3 years of education, but their counterparts from the wealthy subsample had a higher average of 8.7 years of schooling. As a result, the UFPE educational policy yielded equitable years of education for both the wealthy and poor subsamples. However, it notably contributed to a comparatively higher growth in educational achievement for the latter, when considering relative terms. Therefore, the first stage findings indicate that the educational policy implemented by UFPE led to a significant and substantial rise in the educational attainment of females.

Table 5.2: First Stage Estimates -– Effect of the UFPE Policy on Maternal Education

	Years of Schooling					
	Full		Poor		Wealthy	
	(1)	(2)	(3)	(4)	(5)	(6)
1[Birth year ≥ 1988]	0.802***	0.675***	0.807***	0.795***	1.376***	0.964***
	(0.107)	(0.092)	(0.109)	(0.106)	(0.149)	(0.144)
Mean of dep. var.	[7.03]		[5.32]		[8.68]	
Interaction: 1*Year of Birth	0.117***	0.116***	0.198***	0.188***	0.043	0.047
	(0.294)	(0.0265)	(0.031)	(0.03)	(0.04)	(0.039)
Intercept	6.587***	11.432***	4.49***	7.32***	8.016***	10.98***
	(0.064)	(0.258)	(0.669)	(0.315)	(0.083)	(0.332)
First Stage F-Statistic	55.14	53.53	54.76	56.28	84.78	45.04
Observations	16799	16799	8238	8238	8574	8561
Rsquared	0.022	0.242	0.048	0.102	0.043	0.123
Controls		Yes		Yes		Yes
Region Fixed Effects		Yes		Yes		Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

*Note: for direct interpretation of the coefficient on the discontinuity, the birth year distribution has been re-centered at the discontinuity year (1988). Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff. Standard errors clustered at the survey level are presented in parentheses. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

5.3.5. Second Stage Results: Education and HIV related outcomes

5.3.5.1. Education and HIV risk

After establishing the significant and favorable impact of the implemented UFPE education policy, we proceed to examine the causal relationship between the resulting increase in years of schooling and its effect on HIV related outcomes. The OLS, RF (Intention to treat) and 2SLS (Treatment on the treated) results for the full sample are presented in **Error! Reference source not found.**. The OLS results show that each additional year of schooling is associated with a 0.3 percentage lower risk of HIV infection. As previously mentioned, the estimates obtained by OLS may be subject to bias as a result of unobserved factors that have a simultaneous impact on both education and sexual behavior choices. For instance, the association between promiscuity and schooling may be influenced by factors such as economic levels, individual preferences, and societal values. These complex interrelationships might introduce challenges when attempting to ascertain and mitigate bias in OLS estimations. **Error! Reference source not found.** equally displays the reduced form effect i.e. the intention to treat effect where HIV seroprevalence status was directly regressed on the instrument and predetermined covariates.

The UFPE reform had a significant protective effect on those women exposed to the reform. The findings from the reduced form (RF) effects indicate that women who were subjected to the reform exhibited a 3.8 percentage point decrease in the likelihood of being infected with HIV compared to women who were not exposed to the reform. The findings from the 2SLS estimates show that each extra year of female education has a protective effect against HIV seroprevalence status and lowers HIV seroprevalence status by 5.7 percentage points. The results are statistically significant.

5.3.5.2. Education and number of sexual partners

In Table 5.3, we also present the effects of years of schooling on the number of sexual partners. The existence of unobserved factors that simultaneously influence decisions regarding schooling and sexual behavior may introduce bias in the OLS estimates. The reduced form effects show that the UFPE reform had a significant effect on the number of sexual partners for those that benefited from it compared to the older cohort. In the reduced form, (intention to treat), we directly regress the HIV seroprevalence status on the instrument. The results show that the UFPE educational instrument had a significant effect in reducing the number of sexual partners. Each additional year of schooling had an 11.2 percentage point reduction in the likelihood of contracting HIV. The results are statistically significant. The 2SLS (treatment on the treated) estimates show that each additional year of schooling reduces the number of sexual partners by 8.6 percent (-0.172/2.10). We show the full samples for each survey year showing the full sample consenting to testing and the refusal rates by residence and province. These are shown in Table 13, Table A14 and Table 15.

Table 5.3: Years of Schooling, HIV seroprevalence status (HIV positive = 1) and number of sexual partners - Full sample

	Full Sample: Dependent Variable – Years of schooling		
	OLS	RF	2SLS
	(1)	(2)	(3)
Panel A: HIV seroprevalence status	-0.003***	-0.038***	-0.057***
	(0.001)	(0.01)	(0.015)
Mean of Dep. var.		[0.143]	
Observations		13637	
Rsquared	0.033	0.042	0.042
Panel B: Number of sexual partners	0.012**	-0.112***	-0.172***
	(0.005)	(0.044)	(0.065)
Mean of Dep. var.		[2.01]	
Observations		13591	
Rsquared	0.023		0.025
Controls	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes
Survey Fixed Effects	Yes	Yes	Yes

*Notes: Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The robust standard errors, which are clustered in survey clusters, are enclosed within parentheses. Significance at the 10% level is denoted by *, 5% level by **, and at the 1% level is by ***. The years of birth are r-centered at the discontinuity (1988). Each column's results are generated through an individual regression. The main explanatory variable (schooling) is instrumented by the binary indicator* 1[Birth year≥ 1988]. *Each regression only includes people who were born between 1982 and 1993 i.e. a six year bandwidth is chosen on either side of the cutoff.*

5.3.6. Education and HIV Knowledge

Does education correlate with a greater understanding of HIV?

We now investigate the potential correlation between female education and a greater understanding of the modes of transmission for HIV. The 2SLS results for the effect of additional years of schooling on knowledge about HIV transmission are displayed in Table 5.4. The findings show that within the instrumental variable estimation framework, there exists a statistically significant association between educational attainment and the level of HIV knowledge. On being asked whether a healthy individual can be infected with HIV, each additional year of schooling raises the likelihood of giving the correct answer by over 3 percentage points in the full sample and poor subsample while in the wealthy subsample the likelihood of giving the correct answer raises by over 8 percentage points. On whether condom use decreases risk of contracting HIV, each extra year of schooling increases the probability of giving the correct answer by 2.4 percentage points in the full sample. Regarding the question of whether one had detailed knowledge about HIV, an additional year of schooling raises the probability of a correct answer by 15.4 percent (0.071/0.46)[11] in the full sample.

Table 5.4: Years of Schooling and Knowledge of HIV

	Full Sample	**Poor**	**Wealthy**
	(1)	(2)	(3)
Panel A: Ever heard of HIV?	0.005***	0.008***	-0.0
	(0.002)	(0.003)	(0.002)
Mean of Dep. var.	[0.99]	[0.98]	[0.99]
Observations	16783	8243	8571
Rsquared	0.012	0.008	0.011
Panel B: Do you have detailed knowledge of HIV?	0.071***	0.076***	0.047***
	(0.008)	(0.01)	(0.011)
Mean of Dep. var.	[0.46]	0.38	0.54
Observations	16783	8243	8571
Rsquared	0.056	0.017	0.038
Panel C: Can a healthy individual be infected with HIV?	0.032**	-0.035*	0.083***
	(0.014)	(0.021)	(0.018)
Mean of Dep. var.	0.81	0.77	0.85
Observations	16783	8243	8571
Rsquared	0.047	0.038	0.052
Panel D: Does Condom use decrease risk of contracting HIV?	0.024***	0.037***	0.014
	(0.006)	(0.008)	(0.009)
Mean of Dep. var.	[0.79]	[0.8]	0.78
Observations	16783	8243	8571
Rsquared	0.01	0.01	0.026
Panel E: Is HIV transmitted by mosquito bites?	0.083***	0.027	0.14***
	(0.019)	(0.026)	(0.026)
Mean of Dep. var.	16783	8243	8571

[11] Based on the DHS, one is considered to have detailed knowledge of HIV if they answer correctly to all the following questions: (1) Does Condom use decrease risk of contracting HIV? (2) Is HIV transmission mitigated by sex with 1 partner? (3) Can a healthy individual be infected with HIV? (4) Is HIV transmitted by mosquito bites?

Observations	[0.66]	[0.59]	0.74
Rsquared	0.059	0.018	0.061
Panel F: Is HIV transmitted during childbirth?	0.057***	-0.016	0.086***
	(0.014)	(0.018)	(0.023)
Mean of Dep. var.	[0.82]	[0.83]	0.81
Observations	16783	8243	8571
Rsquared	0.018	0.025	0.042
Panel G: Is HIV transmission mitigated by sex with 1 partner?	0.034***	-0.023	0.054***
	(0.012)	(0.015)	(0.017)
Mean of Dep. var.	[0.9]	0.92	0.89
Observations	16783	8243	8571
Rsquared	0.012	0.015	0.048
Controls	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes
Survey Fixed Effects	Yes	Yes	Yes

*Note: Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. For the dependent variables, a value of 1 indicates a correct answer, while 0 indicates an incorrect answer. All regressions include controls for region, religion and an urban dummy. Robust standard errors clustered at the survey level are presented in parentheses. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

5.4. Discussion

The findings consistently demonstrate that the implementation of the education policy has had a strong positive effect on the number of years of schooling completed by the cohorts affected, with an estimated increase of around 0.7 years.

This first stage effect is similar to that of Keats (2018) and Makate (2016) who used the Ugandan DHS and Hahn et al., (2018) who used the Bangladeshi DHS. In the case of Uganda, several studies have examined the impact of the Universal Primary Education (UPE) policy (Keats, 2018; Andriano and Monden, 2019; and Nagashima and Yamauchi, 2020) have reported findings on the effect of this policy on the duration of schooling. Specifically, Keats (2018) found that the UPE policy increased years of schooling by approximately 0.71 years, while Andriano and Monden (2019) observed an increase of 1.5 years. Similarly, Nagashima and Yamauchi (2020) reported a rise of 0.3 years in the duration of schooling as a result of the UPE policy. In the Nigerian context, a number of papers (Osili and Long, 2008; Larreguy and Marshall, 2017; De Cao and La Mattina, 2019) conducted studies that revealed the impact of the Free Primary Education (FPE) policy on educational attainment. Specifically, their findings indicated that the FPE programme resulted in an increase of 1.5, 0.6, and 2.1 years of schooling, respectively. The implementation of the FCUBE policy in Ghana in 1996 resulted in an increase of one year in the duration of schooling (Adu-Boahen and Yamauchi, 2018). In general, the magnitude of the effect is consistent with the magnitudes observed in previous instances.

The magnitude of the effects increases as the bandwidths become narrower, while it decreases as the bandwidths become broader, as anticipated. The effect of the policy remains largely unchanged even after adding predetermined covariates. Equally, the mean of the control cohort group differs substantially between the poor and wealthy subgroups. Women from the poor subgroups without exposure to the UFPE educational policy had 3.6 years of education while their counterparts from the wealthy subgroups had more than double years of education at 7.4 years. Therefore, although the UFPE policy added the same number of years of education for both the poor and wealthy subsamples as the policy was implemented at a national level, it more than doubled the amount of education for the poor relatively. In summary, the first-stage findings suggest that the UFPE educational policy resulted in a large and substantial increase in female educational attainment.

5.4.1. Second Stage (2SLS) Results

Having established that the UFPE educational policy had positive effects on female education, this paper utilises the exogenous variation in female schooling created as a source of exogeneity to establish a causal link between female schooling and adult women's HIV seroprevalence status, demonstrating a negative impact by employ an instrumental variable approach using an RDD framework. Through the implementation of an education policy change, we demonstrate that primary schooling significantly reduces the likelihood of contracting HIV infection in Zambia. The effects remained consistent when subjected to a variety of robustness tests for reliability. The estimates we obtained are slightly greater, but they align well with the substantial negative associations between education and the risk of HIV in multivariate OLS analysis. The 2SLS estimates we derive are considered causal because they are not susceptible to the influence of unobserved factors, such as psychological qualities or unmeasured variables, which might distort research examining the relationship between education and HIV seroprevalence status.

This article adds to the literature by expanding upon the previous research conducted by De Neve et al. (2015), Aguero and Bharadwaj (2014), Duflo et al. (2015), De Walque (2009) and other scholars to establish a causal relationship between female education and HIV seroprevalence status. We expand upon the discoveries made by Baird et al., (2012) by illustrating the impact of primary education, in contrast to secondary school, on the HIV status of adults. This differentiation is significant as numerous students, especially female students, are unable to access secondary education due to socioeconomic or familial limitations.

The UFPE reform had a substantial safeguarding impact on women who were subjected to the reform. In the intention to treat (reduced form) we regressed HIV seroprevalence directly on the UFPE instrument. The findings displayed in Table 5.3 suggest that women who experienced the reform had a 3.8 percentage point reduction in the probability of contracting HIV compared to women who did not benefit from the reform. This result is similar to De Neve et al. (2015) who find that in the case of Botswana, women who were subjected to the reform were 7 percentage points less susceptible to being HIV positive compared to those who were not subjected to the reform.

Within the framework of the 2SLS IV models, for each additional year of schooling resulting from the reform, there was a 5.7 percentage point decrease in the probability of testing positive for HIV. The results of our paper are comparable to De Neve et al. (2015) who find that each additional year of schooling induced by the reform reduced HIV infection risk by 8 percentage points. Our results are also comparable to Behrman (2015) who found that every additional year of education reduces the likelihood of an adult woman in Malawi testing positive for HIV by 6 percentage points, and in Uganda by 3 percentage points. The results are also similar to Lindskog and Durevall (2021) who utilises a similar exogenous variation in educational attainment as a result of a one-year exogenous increase in junior secondary schooling. Their 2SLS results indicate that each additional year of schooling decreases the likelihood of HIV seroprevalence status in Botswana by about 6.5 percentage points. Our results do not align with those of Aguero and Bharadwaj (2014) who do not find any significant effect of additional schooling on HIV seroprevalence status in the case of Zimbabwe. Table 5.3 also shows that for the 2SLS estimates, each additional year of schooling reduces the number of sexual partners by 8.6 percent (-0.172/2.10). Our results are inconsistent with those of Behrman (2015) and Aguero and Bharadwaj (2014) who find scanty evidence of the effects of additional schooling on number of sexual partners.

Comparing the two estimates the OLS and the 2SLS, the 2SLS estimate, on average is larger in magnitude than the OLS result. Possibly, the OLS estimates are erroneous because of the potential omission of covariates, such as family preferences. An alternative explanation is that the IV estimates capture the impact of female education on a specific subgroup of the population (compliers). This subgroup would not have access to free primary schooling if it were not for the UFPE educational programme. It is plausible that the extra years of schooling provided by this policy had a significant effect in reducing HIV seroprevalence status.

We further analyse the relationship between additional schooling and HIV transmission knowledge. We examine whether women with more schooling know more about HIV that those with less schooling. The results are displayed in Table 5.4. For some inquiries such as, "Ever heard of HIV?" The magnitude of the impact is rather modest due to the widespread awareness of HIV among the general population. On being asked whether one has detailed knowledge of HIV, each extra year of schooling increases the likelihood of giving the correct answer by 7.1 percentage points in the full sample. This result is also consistent in the poor and wealthy subsample. On being asked whether a heathy individual can be infected with HIV, an additional year of schooling increases the likelihood of giving the correct answer by over 3 percentage points in the full sample and the wealthy subsample by about 8 percentage points but surprisingly decreases the likelihood of giving the correct answer in the poor subsample by about 3.5 percentage points. On being asked whether condom use decreases risk of contracting HIV, each additional year of schooling increases the probability of giving the correct answer in the full sample and the poor subsample but is insignificant in the wealthy subsample. Finally on being asked whether HIV transmission is mitigated by having sex with 1 partner, each additional year of schooling increases the likelihood of giving the correct answer in the full sample and wealthy subsample by about 3.7 and 6.1 percent respectively.

Education plays a crucial role in fostering thorough understanding and minimising myths surrounding HIV. In general, the findings presented in Table 5.4 indicate that schooling significantly influences a majority of HIV knowledge measures. The 2SLS estimates confirm that educated women possess a greater understanding and knowledge of HIV transmission.

5.4.2. Heterogeneous Effects

We examine the effects of years of schooling on the HIV related outcomes in the poor and wealthy subsamples.[12] Table 5.5 displays the subsample results of the effects of education on HIV seroprevalence status. The OLS results for the poor subsample show contradictory effects of additional schooling on HIV seroprevalence status. Each extra year of schooling increases HIV risk in the poor subsample. The OLS estimates as earlier discussed are however potentially biased as a result of unobserved variables which could simultaneously affect both educational and sexual behaviour choices. For the 2SLS estimate, each extra year of schooling reduces HIV risk by about 5 percentage point. The wealthy subsample on the other hand shows a further reduction of about 6 percentage points in the probability of testing for HIV as a result of an additional years of schooling within the 2SLS framework.

Table 5.5: Years of Schooling, HIV seroprevalence status subsample (HIV positive = 1) and Number of sexual partners Subsample

	Dependent variable: Years of schooling			
	Poor Subsample		Wealthy subsample	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Panel A: HIV seroprevalence status	0.002*	-0.048***	-0.01***	-0.064****
	(0.001)	(0.016)	(0.002)	(0.025)
Mean of Dep. var.	[0.11]		0.18	
Observations	7615		6026	
Rsquared	0.037	0.041	0.021	0.028
Panel B: Number of sexual partners	0.021***	-0.012	0.008	-0.228**
	(0.007)	(0.093)	(0.008)	(0.092)
Mean of Dep. var.	[1.95]		[1.79]	
Observations	7832		5732	
Rsquared	0.036	0.036	0.012	0.015
Controls	Yes		Yes	
Region Fixed Effects	Yes		Yes	
Survey Fixed Effects	Yes		Yes	

[12] We regard respondents whose households have a wealth score that is equal to or lower than the median to be "poor," and we consider respondents whose families have a wealth level that is higher than the median to be "wealthy."

*Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. Robust standard errors clustered at the survey level are presented in parentheses. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

There are heterogeneous effects equally of the years of schooling on the number of sexual partners. The results are equally displayed in Table 5.5. There is a significant reduction in the number of sexual partners for the wealthy subsample compared to the poor subsample within the 2SLS framework. Each extra year of schooling lowers the number of sexual partners by about 13 percent (-0.228/1.79). This high reduction in number of sexual partners for the wealthy is statistically significant and could be explained by the higher years of schooling in the wealthy subsample as opposed to the poor subsample.

5.4.3. Potential Transmission Mechanisms

We also investigate the mechanism by which female education may influence one's HIV seroprevalence status and other HIV related outcomes, employing a consistent methodology to establish causal relationships. We begin by examining the impact of education on various indicators pertaining to sexual behaviour during adolescence and adulthood, namely age at first sexual debut and age at first marriage as these are associated with the risk of contracting HIV. The results are presented in Table 5.6. We find significant effects of each additional year of schooling on these potential mechanisms of contracting HIV. Each extra year of schooling increases age at first sexual debut and age at first marriage by 19.5 and 32.8 percentage points, respectively. The age at first marriage mechanism as a potential pathway through which education affects HIV outcomes seems to operate in both the poor and wealthy subsamples. In Zambia, there is a correlation between schooling and the age at which individuals get married. This correlation suggests that female education delay partnerships that could potentially expose women to risks. This finding is similar to that of Behrman (2015). Each extra year of schooling has no significant effect on the age at first sex in the poor subsample. Therefore, this effect as a mechanism operates only through the wealthy subsample. This result also aligns with Behrman (2015) who finds no effect of additional education on sexual debut for Malawi and Uganda. The education of the husband seem to play a significant role in reducing HIV related risks on both the poor and wealthy subsamples.

Another vital mechanism is literacy. Literacy holds significance as a rudimentary indicator of abilities or cognitive advancement that occurs within educational institutions. We find significant positive effects of additional years of schooling on literacy by about 20 percent. This mechanism operates through the poor as well as the wealthy subsamples.

Finally, we investigate the correlation between educational attainment and the utilisation of several media platforms as sources of information on HIV related outcomes. Female education substantially increases the likelihood of watching television and reading newspaper in both the poor and wealthy subsample. The findings from the 2SLS estimation indicate a statistically significant 15% rise in the probability of individuals engaging in newspaper reading in the full sample. Equally, listening to radio and watching television are significant pathways via which education can affect HIV related outcomes although listening to the radio operates only in the full sample.

Lastly we find no effect of additional schooling on the probability of having used a condom in the last sex. This analysis is insufficient in its study of potential mechanisms, as there are likely other pathways via which schooling influences adult HIV related outcomes. Additional research should thoroughly investigate the intricate mechanisms by which education impacts HIV related outcomes, paying close attention to the variations observed in different circumstances.

Table 5.6: Transmission Mechanisms

	Full Sample	Poor	Wealthy
	(1)	(2)	(3)
Panel A: Age at first Marriage			
Schooling instrumented by:	0.328***	0.25***	0.2*
1[Birth year ≥ 1988]	(0.056)	(0.162)	(0.11)
Mean of dependent var.	18.15	17.61	18.9
Observations	11632	6897	4750
Rsquared	0.052	0.02	0.02
Panel B: Age at first sexual intercourse			
Schooling instrumented by:	0.195***	0.079	0.232***
1[Birth year ≥ 1988]	(0.045)	(0.054)	(0.077)
Mean of dependent var.	16.57	16.02	17.11
Observations	12258	6160	6115
Rsquared	0.058	0.013	0.033

Panel C: Husband education			
Schooling instrumented by:	0.766***	0.621***	0.414***
1[Birth year ≥ 1988]	(0.075)	(0.086)	(0.12)
Mean of dependent var.	8.13	6.68	10.18
Observations	10464	6124	4355
Rsquared	0.22	0.056	0.067
Panel D: Literacy			
Schooling instrumented by:	0.109***	0.13***	0.055***
1[Birth year ≥ 1988]	(0.008)	(0.01)	(0.01)
Mean of dependent var.	0.54	0.49	0.08
Observations	16377	8243	8571
Rsquared	0.12	0.058	0.074
Panel E: Watches Television			
Schooling instrumented by:	0.095***	0.017**	0.114***
1[Birth year ≥ 1988]	(0.008)	(0.007)	(0.012)
Mean of dependent var.	0.41	0.13	0.67
Observations	16738	8243	8571
Rsquared	0.33	0.02	0.19
Panel F: Reads Newspaper			
Schooling instrumented by:	0.044***	0.014*	0.04***
1[Birth year ≥ 1988]	(0.007)	(0.008)	(0.013)
Mean of dependent var.	0.3	0.17	0.43
Observations	16783	8243	8571
Rsquared	0.085	0.01	0.037
Panel G: Listen to Radio			
Schooling instrumented by:	0.017**	-0.006	0.013
1[Birth year ≥ 1988]	(0.008)	(0.011)	(0.011)
Mean of dependent var.	0.6	0.48	0.71
Observations	16783	8243	8571
Rsquared	0.057	0.018	0.033
Panel H: Condom use during last sex			
Schooling instrumented by:	0.004	0.002	0.014
1[Birth year ≥ 1988]	(0.006)	(0.006)	(0.01)
Mean of dependent var.	0.15	0.1	0.78
Observations	13010	8239	8571
Rsquared	0.02	0.004	0.026

*Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. Robust standard errors clustered at the survey level are presented in parentheses. ***, **, and * represent significance at the 1, 5 and 10 percent level, respectively.*

5.5. Identification, internal and external validity

5.5.1. Internal Validity

To effectively employ the educational policy of UFPE as a natural experiment for the aim of getting policy estimates, it is imperative to acknowledge and address certain concerns. One notable concern is from the potential for parents to hold previous information on the year and time of the UFPE educational policy. This knowledge might be utilised by parents to intentionally postpone their children's enrolment in primary school, with the aim of capitalising on the benefits offered by the policy. Nevertheless, it is highly unlikely that this is the case, as the UFPE educational policy was a deliberate and comprehensive initiative designed to be nationally applied, aiming to provide free basic education to all citizens, regardless of their academic level or socioeconomic status.

An additional concern pertains to females who were in close proximity to the commencement age of primary school grade 1, which spans a duration of 7 years. The aforementioned individuals were between the ages of 8 and 10 at the time when the UFPE educational policy was put into effect. The control and treatment cohorts are based on the 1988 birth year cutoff criteria. However, it is important to note that the UFPE programme made exceptions for dropouts and late entrants, allowing them to also receive the benefits of the free primary education policy. This introduces a potential source of uncertainty in our RDD, making it a fuzzy as opposed to a sharp RDD. If there is a significant number of school dropouts and late entrants, the overall impact of the UPFE educational policy may exceed our initial estimation. However, it is not possible to quantify late entrants and dropouts who have re-enrolled in school using the data provided by the DHS.

Another important consideration is the exclusion restriction. This suggests that there may be a potential association between the policy implemented at certain thresholds and other governmental policies. Given the aforementioned scenario, our causal estimation would encompass the UFPE educational policy in conjunction with many other governmental programmes. Based on the existing information, it seems that there are no other policies in place that have the same qualifying criteria for treatment as the UFPE educational policy. Moreover, no empirical data exists to indicate that any alternative strategy has solely impacted a particular subset of observations within the optimal bandwidth.

Local randomisation is an essential element of the regression discontinuity design. Therefore, outcome determinants, apart from the treatment (education), should undergo continuous evolution (smoothing assumption) at the cut-off, encompassing women born prior to, during, and subsequent to the cut-off years. In order to evaluate the validity of the smoothing assumption, (exclusion restriction), within the RDD, we analyze the pretreatment covariates as suggested by Imbens and Lemieux (2008) and Jacob et al. (2012).

Obtaining appropriate variables from survey data for the purpose of doing hypothesis testing of this kind presents challenges. Within the DHS dataset, religion and region are identified as the most prominent pretreatment covariates. Religion is commonly perceived as a somewhat enduring characteristic, wherein individuals often maintain adherence to the religious ideas inherited from their families. The process of changing one's region of residency is equally challenging within a limited timeframe. In situations when treatment allocation is randomised, it is anticipated that there will be no discernible differences among participants in relation to pretreatment covariates. The findings are presented in Table 5.7. Consistent with our earlier assumptions, the coefficients related to the specified covariates do not exhibit statistical significance. Therefore, we argue that the procedure of sample selection and treatment assignment was free from bias.

***Table 5.7*: Pretreatment covariates**

	Region	Religion
Coefficient	-0.066	0.088
t	-0.27	1.29
$Pr(\lvert T\rvert > \lvert t\rvert)$	0.788	0.198

Notes: This table evaluates the hypothesis that there are no statistically significant variations in the pretreatment variables among respondents located on either side of the cut-off. The fundamental assumption is that when treatment allocation is randomized, there should be no observable disparities among participants in terms of pre-treatment variables. The data presented provides support for the validity of the hypothesis.

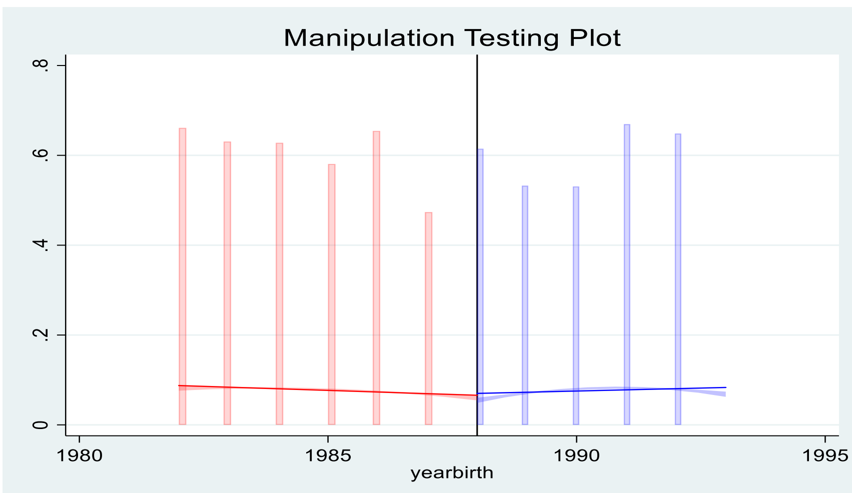
Lastly, in order to evaluate the precision and dependability of our estimations, we undertake an examination of the running variable to identify any potential manipulation. The assessment tool employed in this research was created by McCrary (2008) and is specifically designed to evaluate the consistency of the running variable within the context of the RD design. The fundamental concept of this manipulation testing is that if there is no deliberate and systematic manipulation of the unit's index in close proximity to the cutoff point, the distribution of units should exhibit continuity around this threshold value. The existence of such evidence is often construed as empirical evidence indicating the likelihood of self-selection or nonrandom assignment of units to control and treatment conditions. The results are presented in Table 5.8. The corresponding figure depicting the outcome is presented in Figure 5.3.

Table 5.8: Manipulation Test of running variable

Method	T	P>|T|
Robust	0.096	0.9235

Notes: The outcome of the ultimate manipulation test indicates a T- value of 0.096, accompanied by a p-value of 0.9235. We fail to reject the null hypothesis of no manipulation. Therefore the results lack statistical proof of any deliberate manipulation of the running variable.

Figure 5.3: Manipulation Testing Plot



This figure plots the density of the running variable against the year of birth. At the cutoff, there are no visible signs of any manipulation of the running variable. In other words, there is no bunching of the running variable at the cutoff.

5.5.2. External Validity

One potential constraint of our research findings is their applicability, which is limited to a certain ideal temporal window of six years prior to and following the defined cutoff. However, this is common within the RDD framework. Nevertheless, the findings of this study are of great importance for African nations and other countries that have not yet implemented UFPE-type policies. These nations are striving to achieve the benefits that are linked to this educational policy. Extensive research in the field of economics has provided evidence for the endogeneity of education, partly due to the fact that individuals have agency in choosing their level of schooling. Measurement errors related to the duration of formal schooling have the potential to introduce endogeneity. Nevertheless, it is anticipated that the extent of measurement errors in education data will be small, hence ensuring its dependability.

5.6. Robustness Checks

To bolster the reliability of our findings, we undertake further analyses by estimating the outcomes using alternative functional forms and employing varying bandwidths. The statistical significance of the results stays consistent across different functional forms for estimation, as seen by the data reported in Table 5.9. The first stage results are re-estimated using bandwidths of four, six, eight, and ten years on both sides of the cutoff. Even when considering a relatively low bandwidth of four years on either side of the cutoff, the re-estimated findings indicate a rise in the impact of the UFPE policy on educational achievement. The optimal findings indicate that the UFPE educational policy resulted in a significant rise in the average number of years of completed education, with an approximate increase of 0.7 years. Therefore, the findings remain consistent when considering birth year cohorts on either side of the cutoff.

Table 5.9: First stage estimates with different bandwidths

	Dependent variable: Years of schooling							
	4 year bandwidth		6 year bandwidth (optimal bandwidth)		8 year bandwidth		10 year bandwidth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1[Birth year ≥ 1988]	0.651***	0.573***	0.802***	0.671***	0.819***	0.707***	0.972***	0.827***
	(0.122)	(0.105)	(0.107)	(0.092)	(0.099)	(0.086)	(0.089)	(0.078)
Mean of dep. var.	[7.05]		[7.03]		[7.08]		[7.08]	
F-statistic	28.45	29.5	55.14	53.05	68.15	67	118.71	113.39
Observations	10959		16823		22042		27202	
Controls	No	Yes	No	Yes	No	Yes	No	Yes

This table presents re-estimated first stage results. Columns (1), (3), (5), and (7) are devoid of control variables, while columns (2), (4), (6), and (8) incorporate control variables for place of residence, region and religion. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. The standard errors have been clustered at the survey clusters and are presented in parentheses $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

In addition, alongside the selection of the varying bandwidths, we evaluated the impact of different functional forms on our findings. The purpose of conducting this supplementary analysis was to enhance the dependability and consistency of our findings. The primary equations (4) is modified to include higher order polynomials of the forcing variable, thereby enabling the consideration of the interaction between the running variable (centered at the cutoff) and the treatment indicator, denoted as $Z_i = 1[Year\ of\ Birth \geq c(1988)]$. Throughout all of our estimations, we continually uphold a bandwidth of six that has been optimised based on the mean squared error (MSE) criterion. The findings, as estimated using the primary sample, are presented in Table 5.10. In the specifications provided, our estimations consistently exhibit consistent level of robustness, as evidenced by the magnitude of the observed impacts and their statistical significance.

Table 5.10: Main Results with different functional forms

	Linear Interaction	Quadratic Interaction	Cubic Interaction
HIV risk	-0.057***	-0.059***	-0.061***
	(0.015)	(0.018)	(0.19)
Observations	13642	13642	13642
Controls	Yes	Yes	Yes
Number of sexual partners	-0.173***	-0.222***	-0.189***
	(0.065)	(0.078)	(0.084)
Observations	13595	13595	13595
Controls	Yes	Yes	Yes

Notes: This table presents re-estimates of the main results using different functional forms. All specifications include controls, region and survey fixed effects. The estimates remain robust to varying functional forms. Pooled survey data from the 2007, 2013/14 and 2018 DHS is used. The controls include socioeconomic status, rural/urban status, region and religion. Robust standard errors have been clustered at the survey clusters and are presented in parentheses $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

5.7. Conclusion

This paper examines the causal relationship between education and three HIV related outcomes: HIV seroprevalence status, number of sexual partners and knowledge about HIV by utilizing Zambia's free primary education policy (UFPE) of 2002 as a natural experiment, specifically focusing on the context of HIV in Africa. We utilise the plausibly exogenous increase in the accessibility of free primary education as a source of exogeneity to examine the causal impact of education on HIV related outcomes in Zambia.

The findings indicate that the UFPE educational policy had a positive impact on female schooling. There are three sets of results. The first set of results shows that an increase in female education leads to a decrease in HIV seroprevalence status. By use of an enacted educational policy, we establish that primary education substantially decreases the probability of testing positive for HIV infection in Zambia. The UFPE reform had a substantial safeguarding impact on women who were subjected to the reform as evidenced by the intention to treat (reduced form) results. Our second set of results show that female schooling reduces the number of sexual partners. Our third set of results show that female education increases knowledge about HIV and related outcomes.

Education plays a crucial role in fostering thorough understanding and minimising myths surrounding HIV. In general, the findings indicate that schooling significantly influences a majority of knowledge measures. The 2SLS estimates confirm that women with higher levels of education, possess a greater understanding of the mechanisms via which HIV is transmitted.

The paper also considers the heterogeneous effects of education by examining the effects of years of schooling on the HIV related outcomes in the poor and wealthy subsamples. Specifically, each extra year of schooling reduces HIV risk in the poor subsample. The wealthy subsample shows a further reduction in the likelihood of testing for HIV as a result of an additional years of schooling.

The OLS results for the poor subsample show contradictory effects of additional schooling on HIV risk. Each extra year of schooling increases HIV risk in the poor subsample. The OLS estimates as earlier discussed are however potentially biased as a result of unobserved variables which could simultaneously affect both educational and sexual behaviour choices.

Lastly, we also explore the various pathways through which female education can impact an individual's HIV seroprevalence status and other outcomes related to HIV. We find significant effects of female years of schooling on age at first sex, age at first marriage, literacy and the utilisation of several media platforms as potential sources of information on HIV and as potential mechanisms of contracting HIV. Regarding heterogeneous effects, our findings indicate that female schooling does not have a statistically significant impact on the age of first sexual intercourse in the poor subsample. This suggests that this particular mechanism exclusively functions inside the wealthy subsample. The husband's level of education has been found to significantly reduce HIV-related risks among individuals from both the poor and wealthy subgroups. Our findings indicate that female education played a significant role in the high literacy rates seen in both the full sample and the subgroups of both poor and wealthy households. Finally, we demonstrate that female education impacts the utilisation of media channels as sources of knowledge regarding HIV-related outcomes.

CHAPTER 6: Conclusion

This section provides a concise overview of the key findings presented in chapters 3, 4 and 5, and examines the potential policy implications arising from these results. Within each chapter, we present substantiating evidence to validate the identification assumptions.

Education, particularly maternal education, is widely recognised as a crucial policy instrument for addressing global issues, such as enhancing child health, high fertility and the HIV/AIDS pandemic. This thesis investigates the causal impact of maternal education on three urgent challenges in Zambia: child health, fertility and the HIV/AIDS pandemic. The endogeneity problem is present in all of the three empirical chapters. To mitigate the endogeneity problem, this paper utilizes the implementation of Zambia's Universal Free Primary Education (UFPE) policy implemented in 2002 which created an exogenous source of variation in education as a quasi – experiment in a Regression Discontinuity Design framework.

Education serves as a crucial tool for development policy. Specifically, the chapter examined the effects of a nationwide strategy to eliminate tuition fees on the outcomes of women. The thesis demonstrated that educational achievement was positively influenced for individuals across all social groups. However, the behavioural improvements resulting from increased schooling were primarily observed among the poorest segment of the sample. Poor women benefit more from added schooling in the form of increased literacy, remunerated employment opportunities as well as lowered desired and realized fertility. These effects of added schooling are less documented for women from wealthier households. This validates the idea that providing education to groups that were previously overlooked yields substantial additional benefits to the educational system. These findings offer explicit guidance for policymakers, donor agencies, and governments: Targeted interventions aimed at persons with little financial resources may offer more cost-effectiveness and eventually promote fairness more effectively than broad programmes that apply to everyone indiscriminately.

Chapter 3 examines the impact of maternal education on child health outcomes in Zambia. The third chapter of this thesis revealed profound and enlightening discoveries. The elimination of tuition fees for primary school in Zambia has had a significant effect on the educational achievement of women. The results of the first stage show that for women who were assigned to treatment or benefited from the UFPE policy, the average effect of the policy was 0.68 more years of education compared to the their counterparts in the control cohort who were not assigned to treatment. We find large and statistically significant impacts of maternal years of schooling on height-for-age, weight-for-height and Weight-for-age z scores. The findings similarly demonstrate that an extra year of education for the mother reduces the occurrence of stunting and underweight conditions. We observe no discernible impact of increased maternal education on the prevalence of wasting.

We demonstrate the potential processes through which maternal education influences child health. The findings of our study suggest that a higher level of maternal education has a positive impact on child health by delaying the age at which women get married, reducing the overall number of children they have, and delaying the age at which they give birth for the first time and engage in sexual intercourse for the first time. We also find evidence of positive assortative mating.

In addition, education empowers mothers by enabling them to access information through television and newspapers, acquire knowledge about the ovulation period, and make informed decisions about contraceptive methods. Remarkably, we observe no discernible impact of increased maternal education on employment outcomes. Additionally, we present evidence of the heterogeneous impacts of a mother's education based on “poor versus wealthy” criterion, whether rural/urban status, region and religion. Our analysis reveals that the impact of an additional year of female education on height-for-age z-scores is more pronounced in the wealthy subgroup than in the poor subgroup. However, we observe no heterogeneous effects in the probability of stunting among both the poor and wealthy subgroups. The estimation results of this chapter are largely consistent with those of previous studies, especially with tuition fee abolition policies on primary education in Africa.

Chapter 4 of this thesis reviews insightful findings. We examine the effect of female schooling on fertility. The chapter estimates the causal effect of education on three fertility outcomes: the number of children ever born per woman, ideal (preferred) number of children, and the age of women at first birth. The first set of our main results indicate that there is a negative relationship between female education and the total number of children ever born. Our second set of the main results equally show a negative relationship between female education and women’s intended lifelong fertility. Lastly, our third set of main results show that female schooling is found to delay the age at first birth.

We also find heterogeneous effects of education on socioeconomic status. The effect of an additional year of schooling on the total number of children ever born in the poor and wealth subsamples is similar although high in magnitude on the poor subsample. Specifically, it has been shown that women from poor households and residing in rural areas exhibit more significant alterations in their fertility patterns and preferences, specifically in terms of the magnitude of the impact. Among the potential mechanisms through which female schooling affects fertility, we find that female schooling affects fertility through: (a) age at first sex and marriage: The results also show that women who get more schooling are likely to postpone their age at first sexual intercourse and marriage (b) Assortative mating: Our results confirm positive assortative mating. An additional year of female schooling increases husband's education (c) Labour market participation: There is no evidence to suggest that female schooling has a major impact labour market participation (d) Literacy: Women's literacy rates also increased substantially (e) Knowledge effect: Additionally, our findings illustrate the impact of the knowledge effect. Additional years of education enhance the probability of engaging in activities such as reading newspapers, watching television, and possessing knowledge about the ovulation cycle.

Chapter 5 of this thesis examines the effect of increased primary schooling on adult women's HIV knowledge, risk and seroprevalence status in Zambia. We examine the causal impact of female schooling on three specific HIV related outcomes: HIV seroprevalence status, HIV transmission knowledge and number of sex partners.

Our first set of results in chapter 5 show that each additional year of schooling resulting from the reform led to a reduction in the likelihood of testing positive for HIV. Our second set of results in chapter 5 show that each additional year of schooling reduces the number of sexual. In our third set of results in chapter 5, we examine the relationship between additional schooling and HIV transmission knowledge. We investigate the correlation between women's education and their knowledge about HIV, specifically determining if women with more education possess greater knowledge about HIV compared to those with less education. We offer rare causal estimates about the impact of education on knowledge about health and health behaviour in relation to HIV. This evidence substantiates the concept regarding the allocative efficiency of education. Our findings indicate that individuals with higher levels of education are more inclined to possess intricate and nuanced knowledge on HIV. Specifically, receiving primary education significantly influences one's level of understanding of HIV. We find heterogeneous effects of education on HIV related outcomes. Specifically, each extra year of schooling reduces HIV risk in the poor subsample. The wealthy subsample shows a further reduction in the likelihood of testing for HIV as a result of an additional years of schooling. Lastly, we also explore the various pathways through which female education can impact an individual's HIV seroprevalence status and other outcomes related to HIV. We find that female education significant delays age at first sexual debut and age at first marriage. For heterogeneous effects, we find that female schooling has no significant effect on the age at first sex in the poor subsample, signifying that this mechanism operates only through the wealthy subsample. The education of the husband is found to have a substantial impact on decreasing HIV-related risks among both the poor and wealthy and subgroups. We also found that female schooling was largely responsible for the high literacy levels in the main sample and also in the poor and wealthy subsamples. Lastly, we show that female schooling influences the usage of media platforms as sources of information on HIV related outcomes.

Policy Implications

The results of this study have significant consequences for policy-making. The findings indicate that education reforms can significantly contribute to enhancing formal education achievement among females, hence leading to improvements in child health, lowering fertility and HIV seroprevalence status. Investing in policies that promote the education of girls yields significant long-term advantages. It is evident that enhancing the educational level of mothers has a crucial role in the development of children's human capital and serves as a means to disrupt cycles of poverty.

Our estimation findings reveal the importance of maternal education. This is especially pertinent considering the enduring impact of childhood health on an individual's entire lifespan. Hence, measures aimed at enhancing women's educational chances can potentially provide favourable outcomes for both their own well-being and the economy. This is because when children are raised in good health, they are more likely to receive education and contribute productively to society.

Based on the supporting evidence regarding the incarceration effect, it is recommended that the government implement mandatory education in order to decrease the occurrence of teenage pregnancy. Assessing the correlation between education and reduced HIV incidence would highlight a significant indirect effect of education in underdeveloped nations. Therefore, as our understanding and collection of data on HIV status and HIV-related behaviours improves, these empirical exercises are likely to provide further insights into this significant connection.

Limitations and Future Research

Prior research has shown that the level of education attained by fathers has a crucial role in enhancing the well-being of their children (Breierova and Duflo, 2004; Chou et al., 2010; Ali and Elsayed, 2018). Nevertheless, this work primarily emphasised maternal education, disregarding paternal education due to the nature of our instrument. Excluding information on the father's education will result in an inflated estimation of the impact of the mother's education (Breierova and Duflo 2004; Chou et al. 2010). While the nature of our instrumental variable does not allow us to estimate the impact of paternal education on child health separately in this study, future research can explore the causal impact of paternal education on child health and equally estimate the causal impact of adult male education on HIV related outcomes.

Further, the other limitation of the DHS is that it does not contain the variable that measures the number of children living within the household. This, as a limitation in the DHS makes it difficult to estimate the quantity-quality of children as posited by Becker (1960). Also, the DHS does not directly measure household income. This limitation is however mitigated by the DHS wealth index score which is constructed by assessing households' living arrangements. The study was unable to analyse the long-term characteristics of the study population because the cohorts affected by UFPE policy are still young at the time of writing this thesis. Nevertheless, it seems that UFPE has the potential to enhance human capital, as demonstrated in the instance of Uganda (Keats, 2014). Therefore, the potential gains from the UFPE policy over an extended period of time could be significant.

Considering that several African countries have initiated the implementation of free secondary education, it would be intriguing for future research to investigate the causal impact of education at secondary level. Additional education at the secondary level could perhaps yield a contrasting outcome. Finally, the results from all the three chapters are robust to an array of robustness and falsification tests. These estimates, however, pertain specifically to the subset of women whose education is impacted by the UFPE reform, within the optimal bandwidth. Given that the external validity of our empirical strategy is limited, future research needs to investigate whether our findings can be confirmed in other settings.

References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? National Bureau of Economic Research.

Acemoglu, D., & Angrist, J. (2000). How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual*, *15*, 9–59. https://doi.org/10.1086/654403

Adu-Boahen, Emmanuel, and Chikako Yamauchi (2018). The effect of Female Education on Adolescent Fertility and Early Marriage: Evidence from Free Compulsory Universal Basic Education in Ghana. Journal of African Economies 27 (2): 227–48. https://doi.org/10.1093/jae/ejx025.

Aguero, J. M., & 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.

Ainsworth, M., K. Beegle, and A. Nyamete (1996). The Impact of Women's Schooling on Fertility and Contraceptive Use: A Study of Fourteen Sub-Saharan African Countries. The World Bank Economic Review 10 (1): 85–122. https://doi.org/10.1093/wber/10.1.85.

Alderman, H. and Headey, D.D., (2017). How important is parental education for child nutrition? *World Development*, *94*, pp.448-464.

Ali, F. R. M., & Elsayed, M. A. A. (2017). The Effect of Parental Education on Child Health: Quasi-Experimental Evidence from a Reduction in the Length of Primary Schooling in Egypt. Health Economics, 27(4), 649–662.

Ali, F. R. M., & Gurmu, S. (2018). The impact of female education on fertility: a natural experiment from Egypt. *Rev Econ Household*, *109*, 681–712. https://doi.org/10.1007/s11150-016-9357-6

Alsan, M. M., & Cutler, D. M. (2013). Girls' education and HIV risk: evidence from Uganda. Journal of health economics, 32(5), 863-872.

Andriano, L., & Monden, C. W. S. (2019). The Causal Effect of Maternal Education on Child Mortality: Evidence from a Quasi-Experiment in Malawi and Uganda. Demography, 56(5), 1765–1790.

Angrist, J. D. and Pischke, J. S. (2009). Mostly Harmless Econometrics: An Empiricists

Angrist, J. D. and W. N. Evans (1999). Schooling and labor market consequences of the 1970 state abortion reforms. Research in labor economics 18, 75–113.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. 'Identification of Causal Effects Using Instrumental Variables'. Journal of the American Statistical Association 91 (434): 444–55.

Aslam, M., & Kingdon, G. G. (2012). Parental education and child health–understanding the pathways of impact in Pakistan. World Development, 40 (10), 2014–2032.

Baird SJ, Garfein RS, McIntosh CT, Ozler B. (2012). Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. Lancet 379: 1320–29.

Baird, S., C. McIntosh, and B. Özler (2011). Cash or condition? Evidence from a cash transfer experiment. The Quarterly Journal of Economics, qjr032.

Becker, G. S. (1960). An economic analysis of fertility, demographic and economic change in developed countries: a conference of the universities. National Bureau Committee for Economic Research, 209.

Becker, G. S., & Lewis, H. G. (1973). On the Interaction between the Quantity and Quality of Children. Journal of Political Economy, 81(2, Part 2), S279–S288.

Behrman JR, Rosenzweig MR. (2002). Does increasing women's schooling raise the schooling of the next generation? American Economic Review; 92: 323–34.

Behrman, J.A., (2015). The effect of increased primary schooling on adult women's HIV status in Malawi and Uganda: Universal Primary Education as a natural experiment. Social science & medicine, 127, pp.108-115.

Behrman, Julia Andrea, Amber Peterman, and Tia Palermo. 2017. 'Does Keeping Adolescent Girls in School Protect Against Sexual Violence? Quasi-Experimental Evidence from East and Southern Africa'. Journal of Adolescent Health 60 (2): 184–90. https://doi.org/10.1016/j.jadohealth.2016.09.010.

Behrman, Julia Andrea. 2015a. 'The Effect of Increased Primary Schooling on Adult Women's HIV Status in Malawi and Uganda: Universal Primary Education as a Natural Experiment'. Social Science & Medicine 127 (February): 108–15. https://doi.org/10.1016/j.socscimed.2014.06.034.

Behrman, Julia Andrea. 2015b. 'Does Schooling Affect Women's Desired Fertility? Evidence From Malawi, Uganda, and Ethiopia'. Demography 52 (3): 787–809. https://doi.org/10.1007/s13524-015-0392-3.

Bentaquet-Kattan, R. (2006). Implementation if the Free Basic Education Policy. Education Working paper Series No. 7, World Bank, Washington DC.

Black S. E., Devereux P. J., Salvanes K. G. (2008) 'Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births', Economic Journal, 118 (430): 1225–1054.

Black, D. A., Galdo, J. and Smith, J. A. (2007). Evaluating the worker profiling and reemployment services system using a regression discontinuity approach, American Economic Review 97(2): 104–107.

Bledsoe C. H., Casterline J. B., Johnson-Kuhn J. A., Haaga J. G. (1999) 'Critical Perspectives on Schooling and Fertility in the Developing World', Washington D.C.: National Academy Press. Retrieved from http://www.nap.edu/catalog/6272.html.

Bongaarts, John, & Casterline, John (2013). Fertility transition: Is sub-saharan Africa different? Population and Development Review, 38(s1), 153–168.

Bonney, E. (2021). Zimbabwe's parliamentary delegation understudies ghana's FCUBE programme. Daily Graphic.

Boyle, Siobhan, Andy Brock, John Mace and Mo Sibbons (2002). Reaching the Poor: the Costs of Sending Children to School. United Kingdom Department for International Development.

Braakmann, N. (2011). The causal relationship between education, health and health related behaviour: Evidence from a natural experiment in england. Journal of Health Economics 30 (4), 753–763.

Breierova, L., & Duflo, E. (2004). The impact of education on fertility and child mortality: Do fathers really matter less than mothers? National Bureau of Economic Research.

Burde, D., & Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, *5*(3), 27–40. https://doi.org/10.1257/app.5.3.27

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. The Stata Journal 14 (4), 909{946.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295{2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82: 2295–2326.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. The Stata Journal 17 (2), 372{404.

Cao, R., Ye, W., Liu, J., Chen, L., Li, Z., Ji, H., & Liu, P. (2024). Dynamic influence of maternal education on height among Chinese children aged 0–18 years. *SSM-Population Health*, *26*, 101672.

Card, D. (1999). Chapter 30. The causal effect of education on earnings. In *Handbook of Labor Economics*: *Vol. 3 PART* (pp. 1801–1863). https://doi.org/10.1016/S1573-4463(99)03011-4

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." Econometrica 69, no. 5:1127–60.

Case A, Paxson C. HIV risk and adolescent behaviors in Africa. (2013). American Economic Review; 103: 433–38.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. The Stata Journal, 18(1), 234-261.

Cattaneo, M. D., M. Jansson, and X. Ma (2019). lpdensity: Local polynomial density estimation and inference. arXiv preprint arXiv:1906.06529 .

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. Journal of the American Statistical Association 115 (531), 1449{1455.

Central Statistical Office (CSO), Ministry of Health (MOH), Tropical Diseases Research Centre (TDRC), University of Zambia, and Macro International Inc. 2009. Zambia Demographic and Health Survey 2007. Calverton, Maryland, USA: CSO and Macro International Inc.

Chen, Y., & Li, H. (2009). Mother's education and child health: Is there a nurturing effect?. Journal of Health Economics, 28(2), 413–426.

Chicoine L. (2012) 'Education and Fertility: Evidence from a Policy Change in Kenya', IZA Discussion Paper 6778.

Chicoine, (2016). Free Primary Education, Schooling, and Fertility: Evidence from Ethiopia. In IZA Discussion Paper No. 10387. Bonn: The Institute for the Study of Labor (IZA). https://doi.org/10.2139/ssrn.2868563.

Chicoine, (2020). Free Primary Education, Fertility, and Women's Access to the Labor Market: Evidence from Ethiopia. The World Bank Economic Review, 19.

Chou, S. Y., Liu, J. T., Grossman, M., & Joyce, T. (2010). Parental education and child health: Evidence from a natural experiment in Taiwan. American Economic Journal: Applied Economics, 2(1), 33–61.

Cochrane, Susan Hill (1979). Fertility and Education: What Do We Really Know? World Bank Staff Occasional Papers 26. Baltimore: Published for the World Bank [by] Johns Hopkins University Press.

Colclough C (1982). The impact of primary schooling on economic development: a review of the evidence. World Dev 10(3):167–185.

Colin A. Cameron and Douglas L Miller (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2):317–372, 2015. Companion, Princeton University Press.

Cui, Y., Liu, H., & Zhao, L. (2019). Mother's education and child development: Evidence from the compulsory school reform in China. *Journal of Comparative Economics*, *47*(3), 669–692. https://doi.org/10.1016/j.jce.2019.04.001

Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. The Quarterly Journal of Economics, 118(4), 1495–1532.

Cutler, David M., and Adriana Lleras-Muney. (2010). "Understanding Differences in Health Behaviors by Education." Journal of Health Economics 29, no. 1:1–28.

De Cao, E., & La Mattina, G. (2019). Does Maternal Education Decrease Female Genital Cutting? AEA Papers and Proceedings, 109, 100–104.

De Neve, J. W., Fink, G., Subramanian, S. V., Moyo, S., & Bor, J. (2015). Length of secondary schooling and risk of HIV infection in Botswana: Evidence from a natural experiment. The Lancet Global Health, 3(8), e470–e477.

De Neve, J. W., Fink, G., Subramanian, S. V., Moyo, S., & Bor, J. (2015a). Length of secondary schooling and risk of HIV infection in Botswana: Evidence from a natural experiment. Lancet Global Health, 3(8), e470–e477. https://doi.org/10.1016/S2214-109X(15)70161-0.

De Neve, Jan Walter, Günther Fink, S. V., Subramanian, Sikhulile Moyo, & Jacob, Bor (2015b). Length of secondary schooling and risk of HIV infection in Botswana: Evidence from a natural experiment. Lancet Global Health, 3(8), e470–e477. https://doi.org/10.1016/S2214-109X(15)70161-0.

De Neve, Jan-Walter, & Kawachi, Ichiro (2017). Spillovers between siblings and from offspring to parents are understudied: A review and future directions for research. Social Science & Medicine, 183, 56–61.

De Walque D. Does education affect HIV status? Evidence from five African countries. (2009). World Bank Economic Review; 23: 209–33.

Dinser, M. A., Kaushal, N., & Grossman, M. (2014). Women's education: Harbinger of another spring? Evidence from a natural experiment in Turkey. World Development, 64, 243–258.Adu Boahen, E., & Yamauchi, C. (2018). The effect of female education on adolescent fertility and early marriage: evidence from free compulsory universal basic education in Ghana. Journal of African Economies, 27(2), 227-248.

Duflo E., Dupas P., Kremer M., Sinei S. (2006) 'Education and HIV/AIDS Prevention: Evidence from a Randomized Evaluation in Western Kenya',World Bank Policy ResearchWP Series 4024.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. (2015). Education, HIV, and Early Fertility Experimental Evidence from Kenya. American Economic Review 105 (9): 2757–97. https://doi.org/10.1257/aer.20121607.

Dupas, P. (2011). Health behavior in developing countries. American Economic Review; 3(1), 425-449.

Durevall, D., Lindskog, A., & George, G. (2019). Education and HIV incidence among young women in KwaZulu-Natal: An association but no evidence of a causal protective effect. PloS One, 14(3), e0213056.

Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications, Chapman and Hall, London, New York and Melbourne.

Ferré, Céline (2009). Age At First Child: Does Education Delay Fertility Timing? The Case of Kenya. In Policy Research Working Paper. WPS 4833. The World Bank. https://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-4833.

Fort, M., N. Schneeweis, and R. Winter-Ebmer (2011). More schooling, more children: Compulsory schooling reforms and fertility in europe. Technical report, Economic Series, Institute for Advanced Studies.

Gelman, A. and G. Imbens (2014). Why high-order polynomials should not be used in regression discontinuity designs. Technical report, National Bureau of Economic Research.

Gelman, Andrew, and Guido Imbens. (2019). 'Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs'. Journal of Business & Economic Statistics 37 (3): 447–56. https://doi.org/10.1080/07350015.2017.1366909.

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.

Grepin, K. A., & Bharadwaj, P. (2015). Maternal education and child mortality in Zimbabwe. Journal of Health Economics, 44, 97–117.

Grossman, M. (1972). The demand for health: A theoretical and empirical investigation. New York: Columbia University Press for the National Bureau of Economic Research.

Grossman, M. (1972a). On the concept of health capital and the demand for health. The Journal of Political Economy, 80(2), 223.

Grossman, M. (1972b). The demand for health: A theoretical and empirical investigation. National Bureau of Economic Research Book, 72, 1–10.

Grossman, M. (2006). Education and nonmarket outcomes. Handbook of the Economics of Education, 1, 577–633.

Grossman, M. (2015). The relationship between health and schooling: What's new? (No. w21609). National Bureau of Economic Research.

Gummerson, E., (2013). Have the educated changed HIV risk behaviours more in Africa? African Journal of AIDS Research. 12 (3), 161-172.

Gummerson, E., Maughan-Brown, B., & Venkataramani, A. (2013). Who is taking up voluntary medical male circumcision? Early evidence from Tanzania. AIDS, 27(16), 2657-2659.

Gunes, P. M. (2015). The role of maternal education in child health: Evidence from a compulsory schooling law. Economics of Education Review, 47, 1–16.

Hahn J., Todd P., Van der Klaauw W. (2001) 'Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design', Econometrica: Journal of the Econometric Society, 69 (1): 201–9.

Hahn, Y., Nuzhat, K., & Yang, H. S. (2018). The effect of female education on marital matches and child health in Bangladesh. Journal of Population Economics, 31, 915-936.

Hardee, K., Gay, J., Croce-Galis, M., & Peltz, A. (2014). Strengthening the enabling environment for women and girls: What is the evidence in social and structural approaches in the HIV response? Journal of the International AIDS Society, 17, 18619.

Hargreaves JR, Bonell CP, Boler T, et al. (2008). Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. AIDS; 22: 403–14.

Imbens G. W., Lemieux T. (2008) 'Regression Discontinuity Designs: A Guide to Practice', Journal of Econometrics, 142 (2008): 615–35.

Imbens, G. W. and Kalyanaraman, K. (2009). Optimal bandwidth choice for the regression discontinuity estimator. Unpublished manuscript, Department of Economics, Harvard University.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice, Journal of Econometrics 142(2): 615–635.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation, Journal of Economic Literature 47(1): 5–86.

Imbens, Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects, Econometrica 62(2): 467–475.

Iranzo, S., & Peri, G. (2009). Schooling externalities, technology, and productivity: theory and evidence from U.S. states. *Review of Economics and Statistics*, *91*(2), 420–431. https://doi.org/10.1162/rest.91.2.420

Jacob, R., Zhu, P., Somers, M.-A., and Bloom, H. (2012). A practical guide to regression discontinuity. Technical report, MDRC.

Joint United Nations Programme on HIV/AIDS. (2020). Global report: UNAIDS report on the global AIDS epidemic. Geneva, Switzerland, 2020.

Joint United Nations Programme on HIV/AIDS. (2023). Global report: UNAIDS report on the global AIDS epidemic 2023. Geneva, Switzerland, 2023.Central Statistical Office (CSO), Ministry of Health (MOH), Tropical Diseases Research Centre (TDRC), University of Zambia, and Macro International Inc. 2009. Zambia Demographic and Health Survey 2007. Calverton, Maryland, USA: CSO and Macro International Inc.

Keats, A. (2014). Women's Schooling, Fertility, and Child Health Outcomes: Evidence from Uganda's Free Primary Education Program. Unpublished, Wesleyan University Department of Economics

Keats, A. (2018). Women's schooling, fertility, and child health outcomes: Evidence from Uganda's free primary education program. Journal of Development Economics, 135, 142–159.

Kohler, H. P., & Thornton, R. L. (2011). Conditional cash transfers and HIV/AIDS prevention: Unconditionally promising? The World Bank Economic Review, 26(2), 165–190.

Larreguy, H., & Marshall, J. (2017). The Effect of Education on Civic and Political Engagement in Nonconsolidated Democracies: Evidence from Nigeria. Review of Economics and Statistics, 99(3), 387–401.

Lavy, V., & Zablotsky, A. (2015). Women's schooling and fertility under low female labor force participation: Evidence from mobility restrictions in Israel. *Journal of Public Economics*, *124*, 105-121

Le, K. and Nguyen, M. (2020). Shedding light on maternal education and child health in developing countries. World Development, 133, p.105005.

Lee D. S., Lemieux T. (2010) 'Regression Discontinuity Designs in Economics', Journal of Economic Literature, 48 (2): 281–355.

Lee, D. and Card, D. (2008). Regression discontinuity inference with specification error, Journal of Econometrics 142(2): 655–674.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics, Journal of Economic Literature 48(2): 281–355.

Lindeboom, M., Llena-Nozal, A., & van Der Klaauw, B. (2009). Parental education and child health: Evidence from a schooling reform. Journal of Health Economics, 28 (1), 109–131.

Lindskog, A. and Durevall, D. (2021). To educate a woman and to educate a man: Gender-specific sexual behavior and human immunodeficiency virus responses to an education reform in Botswana. Health economics, 30(3), pp.642-658.

Ludwig, J. and Miller, D. L. (2007). Does head start improve children's life chances: evidence from a regression discontinuity approach, Quarterly Journal of Economics 122(1): 159–208.

Makate, Marshall, and Clifton Makate. (2016). The Causal Effect of Increased Primary Schooling on Child Mortality in Malawi: Universal Primary Education as a Natural Experiment. Social Science & Medicine 168 (November): 72–83. https://doi.org/10.1016/j.socscimed.2016.09.003.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics, 107*(2), 407–437. https://doi.org/10.2307/2118477

Martin, Teresa Castro (1995). Women's Education and Fertility: Results from 26 Demographic and Health Surveys. Studies in Family Planning 26 (4): 187. https://doi.org/10.2307/2137845.

Masaiti, G., Mwelwa, K., and Mwale, N. (2016). Shifting from Government Bursaries to a Loan Scheme in Higher Education: Exploring the Zambian Experience in Student Financial Aid. Ministry of Higher Education (MOHE), Lusaka: University of Zambia.

Mason, K. O. (1986). The status of women: Conceptual and methodological issues in demographic studies. In Sociological Forum, Volume 1, pp. 284–300. Springer.

Masuda, K., & Yamauchi, C. (2017). The effects of female education on adolescent pregnancy and child health: evidence from Uganda's Universal Primary Education for fully treated cohorts. GRIPS Discussion Papers.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, *142*(2), 698-714.

McCrary, J., & Royer, H. (2011). The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth. American Economic Review, 101(1), 158–195.

Mensch, B. S., Singh, S., & Casterline, J. B. (2005). Trends in the timing of first marriage among men and women in the developing world. In C. B. Lloyd, J. R. Behrman, N. P.

Moav, O. (2005). Cheap children and the persistence of poverty. The Economic Journal 115 (500), 88–110.

Mocan, N. H. and C. Cannonier (2012). Empowering women through education: Evidence from Sierra Leone. Technical report, National Bureau of Economic Research.

Montez, J. K. and Friedman, E. M. (2015). Educational attainment and adult health: Under what conditions is the association causal? Social Science & Medicine, 127:1–7. Special Issue: Educational Attainment and Adult Health: Contextualizing Causality.

Mwansa, Audrey C., Alex Kaba, Lucas Zulu, John Kalokoni, and Griffin Nyirongo (2004). Free Basic Education (Grades 1-7) Policy Implementation Assessment. Ministry of Education. Lusaka, Zambia.

Nagashima, M., & Yamauchi, C. (2020). Female Education and Brideprice: Evidence from Primary Education Reform in Uganda. WIAS Discussion Paper No. 2020-004. Waseda Institute for Advanced Study, Waseda University.

Nepal, A. K. (2018). What matters more for child health: A father's education or mother's education? World Development Perspectives, 10, 24-33.

Nishimura, M. and Byamugisha, A. (2011). The Challenges of Universal Primary Education Policy in Sub-Saharan Africa, pages 225–245. Palgrave Macmillan US, New York.

Nunn, N. (2010). Religious conversion in colonial africa. American Economic Review, 100(2):147–152.

Odunowo, M. A. (2020). Essays in applied microeconomics. Texas A&M University.

Osili, U. O., & Long, B. T. (2008). Does female schooling reduce fertility? Evidence from Nigeria. Journal of Development Economics, 87(1), 57–75.

Osili, U. O., & Long, B. T. (2008). Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics*, *87*(1), 57–75. https://doi.org/10.1016/j.jdeveco.2007.10.003

Ozier, O., (2018). The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis. Journal of Human Resources 53 (1): 157–88. https://doi.org/10.3368/jhr.53.1.0915-7407R.

Pietro, G. D. (2012). The short-term effectiveness of a remedial mathematics course: Evidence from a uk university, The Manchester School forthcoming.

Psacharopoulos G (1985) Returns to education: a further international update and implications. J Hum Resour 20(4):583–604

Psacharopoulos G (1994) Returns to investment in education: a global update. World Dev 22(9):1325–1343.

Psaki, S. R., Chuang, E. K., Melnikas, A. J., Wilson, D. B., & Mensch, B. S. (2019). Causal effects of education on sexual and reproductive health in low and middle-income countries: A systematic review and meta-analysis. *SSM-Population Health*, *8*, 100386.

Republic of Zambia, (2017). Ministry of General Education, Directorate of Planning and Information, Educational Statistical Bulletin.

Rosenberg M, Pettifor A, Miller WC, et al. (2015). Relationship between school dropout and teen pregnancy among rural South African young women. International Journal of Epidemiology. DOI: 10.1093/ije/dyv007.

Rosenzweig, M. R. and T. P. Schultz (1989). Schooling, information and nonmarket productivity: contraceptive use and its effectiveness. International Economic Review, 457–477.

Schultz TP (1988) Education investments and returns. Handbook Dev Econ 1:543–630

Schultz, T. P. (1993). Returns to women's education. In King, E. M. and Hill, M. A., editors, Women's education in developing countries: Barriers, benefits, and policies. Baltimore, MD: Johns Hopkins University Press (for the World Bank).

Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, *74*(1), 199–250. https://doi.org/10.1016/j.jdeveco.2003.12.009

Schultz, T. P. (2007). Population policies, fertility, women's human capital, and child quality. Handbook of development economics, 4, 3249-3303.

Schultz, T. W. (1960). Capital Formation in Education. Journal of Political Economy; 68(4): 57 1-83.

Staiger, Douglas, and James H. Stock. (1997). 'Instrumental Variables Regression with Weak Instruments'. Econometrica 65 (3): 557–86. https://doi.org/10.2307/2171753.

Tequame, Miron, and Nyasha Tirivayi. 2015. 'Higher Education and Fertility: Evidence from a Natural Experiment in Ethiopia'. In UNU-MERIT Working Papers 2015-019. Maastricht: Maastricht Economic and social Research Institute on Innovation and Technology. https://doi.org/10.2139/ssrn.2637960.

Thomas D, Strauss J, Henriques MH (1991). How does mother's education affect child height? Journal of Human Resources 26(2):183–211.

United Nations Children's Fund (UNICEF), World Health Organization (WHO), International Bank for Reconstruction and Development/The World Bank. Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group Joint Child Malnutrition Estimates: Key findings of the 2023 edition. New York: UNICEF and WHO; 2023.

Van der Klauuw, W. (2002). A regression-discontinuity evaluation of the effects of financial aid offers on college enrollment, International Economic Review 43(4): 1249–1287.

Vandemoortele, J., & Delamonica, E. (2000). The "education vaccine" against HIV. Current issues in comparative education, 3(1).

Wang, T. (2024). The impact of female education on fertility: evidence from Malawi Universal Primary Education program. *Journal of Demographic Economics*, 1-29.

Wild, F., & Stadelmann, D. (2024). Heterogeneous Effects of Women's Schooling on Fertility, Literacy and Work: Evidence from Burundi's Free Primary Education Policy. *Journal of African Economies*, *33*(1), 67-91.

Willis R. J. (1973) A new approach to the economic theory of fertility behavior. Journal of Political Economy; 81: S14–64.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data, The MIT press.

World Bank Blogs. (2019). World Bank.

World Bank Group. (2018). The World Bank Group in Zambia – Report No. 128467-ZA. Lusaka: World Bank.

World Bank. (2015). Project Performance Assessment Report, Republic of Zambia, Technical Education and Entrepreneurship Training Development Support Program. Washington, DC.

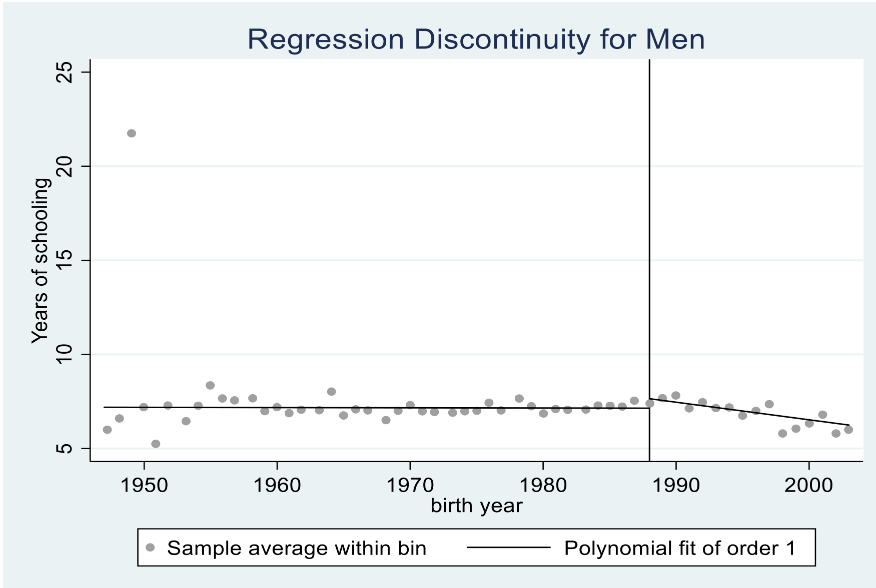
Zambia Statistics Agency, Ministry of Health (MOH) Zambia, and ICF. 2019. Zambia Demographic and Health Survey 2018. Lusaka, Zambia, and Rockville, Maryland, USA: Zambia Statistics Agency, Ministry of Health, and ICF.

Appendix

Table A1: Heterogeneous Effects of maternal education on Rural/Urban. Region and Religion

	HAZ	Stunting	WAZ	Underweight	WHZ	Wasting
	(1)	(2)	(3)	(4)	(5)	(6)
Rural	0.137***	0-.051***	0.099***	-0.033***	0.0159	-0.004
	(0.499)	(0.015)	(0.036)	(0.010)	(0.043)	(0.007)
Region Reference: Central						
	0.104	-0.033	-0.043	0.006	-0.1482**	0.006
Copperbelt	(0.076)	(0.022)	(0.057)	(0.015)	(0.068)	(0.009)
Eastern	0.153**	-0.047**	0.212***	-0.053***	0.1607**	0.001
	(0.0773)	(0.023)	(0.050)	(0.015)	(0.0642)	(0.009)
Luapula	0.108	0.002	-0.082	0.031	-0.2051***	0.051***
	(0.102)	(0.026)	(0.057)	(0.017)	(0.073)	(0.013)
Lusaka	0.151*	-0.016	0.069	-0.016	-0.02268	0.0132
	(0.081)	(0.022)	(0.052)	0.015	(0.065)	(0.009)
Muchinga	0.044	0.003	-0.096	0.024	-0.1647**	0.018**
	(0.078)	(0.024)	(0.056)	(0.018)	(0.064)	(0.01)
Northern	-0.188**	0.051**	-0.116**	0.023	-.0006	-0.010
	(0.078)	(0.022)	(0.053)	(0.018)	(0.657)	(0.008)
North-	0.143**	-0.055**	0.008	-0.024	-0.1283**	0.015
Western	(0.078)	(0.025)	(0.05)	(0.016)	(0.059)	(0.009)
Southern	0.137**	-0.058**	0.009	-0.005	-0.125**	0.001
	(0.073)	(0.023)	(0.059)	0.015	(0.063)	(0.008)
Western	0.312***	-0.142***	0.020	-0.019	-0.237***	0.004
	(0.079)	(0.026)	(0.054)	(0.018)	(0.064)	(0.010)
Religion Reference: Catholic						
Protestant	-0.002	0.002	0.014	-0.01	0.021	-0.008
	(0.049)	(0.014)	(0.036)	(0.010)	(0.041)	(0.007)
Muslim	0.469**	-0.095	0.47**	-0.029	0.299	0.011
	(0.236)	(0.064)	(0.169)	(0.042)	(0.228)	(0.032)

Figure A1: Men's Schooling by Birth Year Cohort



Notes: data from the 2007, 2013/2014, and 2018 Zambia DHS Male Surveys. The survey participants are limited to individuals who are 19 years of age or older.

Table A2: First stage results for Men

Dependent variable: Years of schooling for men	
1[Birth year ≥ 1988]	0.49***
	(0.131)
Mean of dep. var.	[7.19]
Interaction: 1*Year of Birth	-0.093***
	(0.029)
Intercept	7.15***
	(0.0355)
First Stage F-Statistic	13.78

This table presents first stage estimates for men. Standard errors have been clustered at the survey clusters and are presented in parentheses $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 13: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2018

	HIV test									
	DBS tested[13]		Refused to provide blood		Absent at the time of blood collection		Missing/Other[14]			
Residence	Interviewed	Not interviewed	Interviewed	Not interviewed	Interviewed	Not interviewed	Interviewed	Not interviewed	Total	Number
Urban	90.9	0.0	3.6	1.1	0.5	2.8	0.6	0.5	100.0	5766
Rural	93.9	0.0	2.7	0.5	0.2	2.0	0.1	0.5	100.0	8423
Province										
Central	96.8	0.0	1.4	0.2	0.2	0.6	0.2	0.5	100.0	1416
Copperbelt	89.2	0.0	3.9	1.4	0.4	3.0	1.8	0.4	100.0	1695
Eastern	91.3	0.1	3.5	1.1	0.1	3.1	0.1	0.7	100.0	1617
Luapula	86.3	0.0	8.5	1.2	0.8	2.2	0.3	0.7	100.0	1474
Lusaka	96.4	0.0	1.3	0.6	0.1	1.2	0.2	0.2	100.0	1811
Muchinga	96.4	0.1	1.4	0.4	0.0	1.2	0.0	0.5	100.0	1209
Nothern	94.1	0.1	0.9	0.7	0.2	3.8	0.0	0.2	100.0	1301
North-Western	95.8	0.0	2.3	0.5	0.2	1.1	0.0	0.2	100.0	1100
Southern	92.2	0.0	3.2	0.8	0.3	2.7	0.1	0.8	100.0	1407
Western	89.5	0.1	4.1	0.4	1.0	4.5	0.0	0.4	100.0	1159
Total	92.7	0.0	3.1	0.8	0.3	2.3	0.3	0.5	100.0	14189

Source: ZDHS 2018

[13] Includes all Dried Blood Spot (DBS) specimens tested at the lab and for which there is a final result, i.e., positive, negative, or indeterminate.
[14] Includes (1) other results of blood collection (e.g., technical problem in the field), (2) lost specimens, (3) noncorresponding bar codes, and (4) other lab results such as blood not tested for technical reason or not enough blood to complete the algorithm

Table A14: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2013-14

	HIV test							
	DBS tested[15]		Refused to provide blood		Missing/Other[16]			
Residence	Interviewed	Not interviewed	Interviewed	Not interviewed	Interviewed	Not interviewed	Total	Number
Urban	89.1	0.0	6.4	0.9	0.4	3.2	100.0	8212
Rural	91.7	0.0	4.4	0.4	0.3	3.1	100.0	8852
Province								
Central	89.0	0.0	6.3	0.5	0.2	4.1	100.0	1468
Copperbelt	85.7	0.0	9.6	1.4	0.3	2.9	100.0	1850
Eastern	91.6	0.0	4.8	0.1	0.5	3.0	100.0	2101
Luapula	92.2	0.0	4.1	0.3	0.5	2.9	100.0	1637
Lusaka	91.3	0.1	5.7	0.6	0.5	1.9	100.0	1962
Muchinga	89.6	0.0	4.8	0.7	0.2	4.7	100.0	1538
Nothern	93.6	0.1	2.9	0.7	0.2	2.5	100.0	1633
North-Western	93.1	0.0	4.5	0.7	0.2	1.5	100.0	1605
Southern	90.4	0.1	5.4	0.8	0.4	2.9	100.0	1799
Western	87.5	0.0	5.3	1.0	0.3	5.8	100.0	1471
Total	90.4	0.0	5.4	0.7	0.4	3.1	100.0	17064

Source: ZDHS 2013-14

[15] Includes all Dried Blood Spot (DBS) specimens tested at the lab and for which there is a final result, i.e., positive, negative, or indeterminate.
[16] Includes (1) other results of blood collection (e.g., technical problem in the field), (2) lost specimens, (3) noncorresponding bar codes, and (4) other lab results such as blood not tested for technical reason or not enough blood to complete the algorithm

Table 15: Percent distribution of women age 15-49 eligible for HIV testing by testing status, according to residence and province (unweighted), Zambia DHS 2007

	HIV test									
	DBS tested[17]		Refused to provide blood		Absent at the time of blood collection		Missing/Other[18]			
Residence	Interviewed	Not interviewed	Interviewed	Not interviewed	Interviewed	Not interviewed	Interviewed	Not interviewed	Total	Number
Urban	76.4	0.0	18.3	2.0	0.1	1.6	1.0	0.6	100.0	3320
Rural	77.7	0.0	18.5	1.0	0.0	1.2	0.9	0.7	100.0	4088
Province										
Central	70.5	0.0	22.6	1.7	0.1	1.4	2.4	1.1	100.0	702
Copperbelt	69.5	0.0	24.7	4.0	0.0	1.6	0.1	0.1	100.0	879
Eastern	84.6	0.0	12.6	0.4	0.1	0.3	0.9	1.1	100.0	958
Luapula	82.7	0.0	12.7	0.8	0.8	2.9	0.7	0.3	100.0	733
Lusaka	75.8	0.0	1.9.7	2.0	0.1	1.1	0.7	0.6	100.0	976
Nothern	78.9	0.0	17.8	1.0	0.2	0.7	0.7	0.9	100.0	804
North-Western	70.7	0.0	24.3	1.8	0.2	1.8	0.6	0.8	100.0	717
Southern	78.3	0.0	17.1	1.7	0.3	2.1	0.9	0.8	100.0	853
Western	81.9	0.0	14.8	0.6	0.1	0.8	1.4	0.4	100.0	786
Total	77.1	0.0	18.4	1.5	0.0	1.4	0.9	0.7	100.0	7408

Source: ZDHS 2007

[17] Includes all Dried Blood Spot (DBS) specimens tested at the lab and for which there is a final result, i.e., positive, negative, or indeterminate.
[18] Includes (1) other results of blood collection (e.g., technical problem in the field), (2) lost specimens, (3) noncorresponding bar codes, and (4) other lab results such as blood not tested for technical reason or not enough blood to complete the algorithm