Is schooling enough? Education and HIV: Evidence from Zambia *

Irène Dohouin†

Abstract

This paper studies how education affects women's HIV infection. By using an education reform that led to a sharp increase in women's education in Zambia, I estimate RDD, interacted with geographic differences in school capacity. I find that an increase in female education led to HIV higher rate. I find no evidence that education affected women's HIV knowledge and their risky behaviours. Instead, the results are driven by the increased urbanization of the better educated women. Although the results in the existing literature suggest that education enhance health behaviours, my findings are suggestive that those results could not be generalized.

*I am deeply indebted to my advisor, Raphaël Godefroy, for his constant support and help. I thank Joshua Lewis and Andriana Bellou for their comments

†Universite de Montreal. irene.dohouin@umontreal.ca

1 Introduction

The positive correlation between education and health is well documented in the literature (Grossman, 2017), and some studies have proven that this correlation reflects at least partly the causal effect of education on health (Lleras-Muney, 2005).

The aim of this paper is to estimate the causal impact of education on HIV infection. HIV has been one of the main causes of premature disability and death in sub-Saharan Africa in the past decades. Yet, the evidence of any impact of education on infection is unclear. A series of studies (Oster, 2005; Fortson, 2008; Paxson and Case, 2013; Duflo et al., 2015) have found little or no effect of education on the probability of being infected with HIV. HIV being relatively rare, though, the estimation of the effect of education on HIV may require relatively larger samples of observations than the ones used in those studies, or exogenous changes in education of a larger magnitude.

To address these problems, my paper focuses on Zambia. HIV infection is a distinctly important public health issue in Zambia. While the first case had been reported in 1988, 14% of total mortality was due to HIV disease in 1991, representing more than 15,000 deaths (Zambia National AIDS Program). In the 2018 Zambia Demographic Health Survey (DHS) reported that HIV prevalence among females aged 15-49 years was 14.2%, compared to 7.5% for males of the same age. These numbers indicate that Zambia is one of the countries where HIV is much more prevalent in sub-Saharan Africa.

To estimate the impact of education on HIV infection, I use a natural experiment that increased girls' education. Until the late 1990s, girls were much less likely to be enrolled in school than boys in general, and girls dropped out twice more than boys (Nkhata et al., 1998). In 2002, Zambia government announced and implemented free educational reform in the country. This reform abolished tuition fees and the mandate for primary pupils to wear uniform. Consequently, this policy reduces drastically the cost of education. As a result, the reform led to a large increase in women's education.

The data for the estimations are from the 2013 and the 2018 Zambia Demographic and

Health Surveys. I use parametric and non-parametric regression discontinuity designs on women born between 1982 and 1994 to estimate how the reform affects female education and HIV infection. By using women's birth year as an assignment variable, I estimate RDD while taking into account geographic differences in school provision across the country.

I find that the 2002 educational reform led to a sharp jump in female education between the first treated and last non-treated birth cohorts of women. Women in the treated cohorts are 12.4% more likely to complete primary school. The educational reform also increases women's total years of education by 10.2%. I attribute these changes to the reform alone, since no other change seems to have affected women born around the end of the 1980s and early 1990s.

Regarding HIV, I find that the reform increased the probability of being HIV positive. Numerically, there is an increase of 18% of the probability of being tested HIV positive for women who benefited from the reform.

These findings may hide heterogeneous effects due to the difference in education supply in terms of schools capacity across the country. Indeed, geographic differences in school availability could mitigate the effectiveness of the reform across the country (Duflo, 2001). To address this point, I construct for each province, an index of school supply that accounts for the number of schools built by the government between 1999 and 2005 per ten thousands children[1]. Using this index, I split the sample into high supply provinces and low supply provinces and I investigate the effect of the reform in the two groups [2].

I find that the reform has positive and significant effect on women's education and women's HIV infection in low supply provinces, and the patterns are robust alternative controls. Indeed, the reform is associated with a 18.7% increase in women's probability to achieve primary school and with 15.4% increase in women's total years of education. Also, the reform entails 29% increase in HIV infection among young women. As a result, I find that one additional year of education increases women's probability to contract HIV

[1]children under 15 years old

[2]I rank the provinces in a descending order of the index and I split the sample by choosing a threshold. High supply provinces are provinces with a high index value, and low supply provinces are provinces with a low index value relatively to the threshold

infection by 25.7%. I also find that the reform has no effect either on women's education or women's HIV infection in high supply provinces. There is no clear evidence on why I find no effect of the reform in high supply provinces. However, it could be that the reform occurred at a period of continuously increasing enrollment so that the slope of this trend is large enough to be much larger than any possible jump due to the reform.

I then investigate the mechanisms underlying the positive relationship between education and HIV infection. I find that the results cannot be attributed to the increase in the number of lifetime sexual partners of educated women, but that the increase in female education did not improve women behaviors about having safe sex, which is in contrast with De Walque (2007) yet consistent with Oster (2012).

I also find no change in women's bargaining power, so that Anderson (2018)'s finding that women who are less able to negotiate safe sex practices are more vulnerable to HIV than those who have more bargaining power to negotiate safe sex practices may not apply here.

Instead, what I do find is that more educated women are more likely to live in urban areas. Like Cutler and Miller (2005), I argue that my estimation of an effect of education on HIV may result from an impact of education on the probability of living in an urban area, where HIV prevalence is higher.

This paper contributes to three distinct strands of economic literature. First, it contributes to the literature that links education to health behaviors. Studies have generally found that education contributes to good health behaviors (Jensen and Lleras-Muney, 2012; Cutler and Lleras-Muney, 2010). However, this paper shows that the increase in female education does not change women's health behaviors about STIs. An important lesson from this study is that education, itself, is not sufficient to induce good behaviors at least in the case of HIV disease.

Second, it contributes to the literature on the determinant of HIV in Africa. Previous studies have documented a relationship between HIV infection and socioeconomic status. Several studies find that HIV is positively correlated with education (Over and Piot, 1991;

Gregson et al., 2001; Hargreaves and Glynn, 2002; Fortson, 2008; Paxson and Case, 2013) whereas others show that the correlation could be negative (Iorio and Santaeulàlia-Llopis, 2016). Certains papers underline that HIV infection is correlated with sex, age, income and place and residence (Magadi, 2017; Magadi and Desta, 2011). I find that education alone is not necessary a tool to fight against all kind of diseases, particularly HIV infection - as pointed out by Duflo et al. (2015), education alone did not reduce HIV infection. I find that female education can affect positively women's HIV infection when education does not change women's health behaviors in risky environment.

Third, this paper contributes to the literature on the links between urbanization and health. Indeed, the literature about the relationship between urbanization and health provides evidence that living in the city is dangerous for health (Cutler and Miller, 2005). This paper aligns with this finding. Urbanization sounds to be the channel through which education affected positively women's HIV infection.

The remainder of the paper proceeds as follows : section 2 provides some background on the context and the 2002 educational reform; section 3 presents the conceptual framework; section 4 describes the data; section 5 presents the empirical strategy; in section 6, section 7 and section 8 I present and discuss the results. section 9 concludes.

2 Background

Zambia is a country in southern Africa organized into 10 provinces [3]. Zambia population is concentrated mainly around Lusaka in the south and the Copperbelt province in the northwest. Zambia is constituted of many ethnic groups. However, the most representatives are Bemba and Tonga. Each ethnic group has particular ancestral norms. For example, Bemba people precluded premarital sexual behavior of girls by early marriage whereas Tonga people permitted it.[4]

Before the 21st century, literacy and school attendance in Zambia was low. In fact,

[3]The ten provinces are: Muchinga, Western, Copperbelt, North western, Northern, Southern, Eastern, Luapula, Central and Lusaka

[4]Ethnographic Atlas

during the first ten years of independence of Zambia, the primary and the secondary school enrollment have doubled. The enrollment has increased from 378,417 in 1964 to 858,191 in 1974 in primary school and from 13,871 in 1964 to 65,764 in 1974 in secondary school (Zambia ministry of education, 2015)[5]. After this expansion, the education system stagnated and was neglected for lack of means. As a result, enrollment rates in basic education have declined even though the school-going age population was growing quickly. Literacy rates have not improved but have tended to deteriorate and the gender gap in educational attainment persisted. Girls dropped out school earlier than boys in general and this situation was worse in the no denser area like the villages due to the lack of the means, the cost of the distance and the strong demand for child labour (Swainson, 1995).

In the late 1990s, the government of Zambia elaborated the Basic Education Sub-sector Plan (BESSIP) for 1999-2002 to improve access to quality of basic education. In February 2002, he announced the abolition of tuition fees and the non-obligation for students to wear school uniforms to encourage demand for education. This was applied immediately. So the reform has reduced the cost of education and it represents an incitation for educational demand for the poor. News schools were built gradually in each province to support the reform. Table 1 shows the distribution of the new schools by province from 1999 to 2005.

3 Conceptual Framework

There are many positive traits of having education. Indeed, education can help people to think, feel, and behave in a way that contributes to their well-being by improving, not only their personal utility, but also their community (Al-Shuaibi, 2014; Currie and Moretti, 2003).

In this paper, I take advantage from free primary education in Zambia - that increased significantly women education - to assess the effect of an exogenous increase in education on women HIV infection. I find that an increase in female education affects positively women's HIV infection through urbanization. In this section, I provide a theoretical

[5]Education for All 2015 National Review Report: Zambia

evidence that support this result.

Many factors could explain the relationship between education and HIV infection depending on whether this relationship is positive or negative. It could be marriage, believe about the disease, risky behavior, environmental risk etc. In a case of this paper, I develop a theoretical model of HIV infection and show under what assumptions education could affects positively HIV infection. I suppose that individual HIV status depend on the extent in which he involves in risky sexual behavior, what I call sexual risk (p); and the HIV prevalence in his living area, what I call environmental risk (y). This is reasonable assumption since HIV disease is a sexually transmitted disease. Furthermore, I allow p and y to depend on education(x), while keeping the other factors constant.

My empirical analysis consists by examining the reduced-form relationship between education (x) and HIV infection, ie $\frac{\partial HIV}{\partial x}$. This reduced form relationship between education and HIV can then be derived by using partial derivative:

$$\frac{\partial HIV}{\partial x} = \frac{\partial HIV}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial HIV}{\partial y}\frac{\partial y}{\partial x} \tag{1}$$

- $\frac{\partial HIV}{\partial p}$ stands for, ceteris paribus, the relationship between HIV infection and sexual risk. The risk of HIV infection is increasing in risky sexual behavior such as having multiple concurrent partners or unprotected sex. So, $\frac{\partial HIV}{\partial p} > 0$ (Oster, 2005; Stoneburner and Low-Beer, 2004; Potts et al., 2008). Besides, $\frac{\partial HIV}{\partial p}$ is increasing in the HIV prevalence in the individual living areas. Regions with high HIV prevalence will have a stronger relationship between sexual behavior and HIV infection than regions with low prevalence.

- Besides, $\frac{\partial HIV}{\partial y} > 0$, the risk of HIV infection increases with the prevalence of HIV in an area (y) conditional on the risky behavior of individuals.

- $\frac{\partial p}{\partial x}$ represents the relationship between sexual risk and education. Education can be an instrument that aids in the understanding of risks and the behaviors that reduce risks in order to improve health conditions. Thus, *this relationship* ($\frac{\partial p}{\partial x} < 0$) *will be* *negative*. But if education does not improve HIV risk perception, *this relationship will be positive,*($\frac{\partial p}{\partial x} \geq 0$).

- $\frac{\partial y}{\partial x}$ represents the relationship between education and the prevalence of HIV in an area. Apart from the elderly, educated people live mainly in urban area (or in developed regions) because of economic opportunities or people in urban (developed) area are more educated (resources are available for this purpose). In general, more developed regions have higher HIV prevalence than less developed regions. So, educated people are more likely to live in high HIV prevalence area, ($\frac{\partial y}{\partial x} \geq 0$).

In summary,

a. if education does not improve the risk perception, or the individual risky behavior does not change after receiving education, the relationship between HIV infection and education will be positive ($\frac{\partial HIV}{\partial x} \geq 0$).

b. if education improve the risk perception, or the individual risky behaviors change after receiving education, the relationship between HIV infection and education is ambiguous. It depend on the magnitude of the change and environmental risk.

4 Data

The data used in the frame of this study comes from two rounds of Demographic and Health Surveys (DHS) conducted in 2013 and 2018 in Zambia. The DHS are data collected by surveying a nationally representative sample of women (aged 15-49) and men (aged 15-59) in developing countries. The sample size varies across countries between 5,000 and 30,000 households. Initially the surveys consist in interviewing individuals on their socioeconomic, demographic and cultural characteristics. A blood test section has been adding, since 2001, to the verbal interview to test for various health conditions, including HIV status. Thus, there is no bias in HIV status since the blood test is anonymous, voluntary, and non-informative to respondents.

The sample that I use in this paper is made up of 11492 women born between 1982 and 1994. Those women are described by their sociodemographic characteristics (age, education, place of residence, matrimonial status), reproductive health characteristics (number of lifetime sexual partners, the contraceptive methods), their ethnicity and HIV status. Table 2 gives more details about these variables.

I use the women's birth year to define the treatment variable. In Zambia, the legal age to be enrolled in primary school is 7 years old, and primary education consists of seven grades (Zuilkowski et al., 2012). Then, I consider that a woman potentially benefits from the free education reform if she is less than 14 years in 2002. As a result, the beneficiaries are the 1989-1994 birth cohorts women and the non-beneficiaries, the 1982-1988 birth cohorts women.

5 Empirical strategy

The impact of the eligibility to 2002 Zambia educational reform on the probability to achieve primary school and on HIV infection is estimated by using non-parametric and parametric regression discontinuity Designs. The main equation for estimation is :

$$Y_{ic} = \alpha + \beta\, I_c^{89-94} + I_c^{89-94} \times f(dist_c) + I_c^{82-88} * f(dist_c) + \gamma\, X_i + \varepsilon_{ic} \qquad (2)$$

Where Y_{ic} denotes any outcome of interest for individual i who belongs to birth cohort c. Outcomes of interest are the following variables: HIV status, primary education, number of sexual partners, place of residence, use of contraceptive methods, woman's knowledge about HIV prevention[6], woman's bargaining power about safe sex.

In the baseline equation, I_c^{89-94} is a treatment variable. It is equal to 1 for individuals who are born between to the 1989-1994 birth cohorts. Those women are young enough in 2002 to benefit from the reform. In other words, the assignment variable is woman's birth year and the cutoff is 1989 since the reform was implemented in 2002 for the seven years

[6](Abstinence, always use condoms during sex; have one sex partner only, who has no other partners)

of primary school and the legal age for primary school entry is seven years old. The term X_{ic} is a vector of individuals controls including ethnicity. β is the coefficient of interest that captures the effect of the reform on our outcomes of interest.

To estimate the size of the discontinuity in outcomes and treatment, I follow standard methods of regression discontinuity analysis as in Imbens and Lemieux (2008) and Lee and Lemieux (2010). First, I restrict the data to a small window around the cutoff (1989) and I use a rectangular kernel so that the weight on each observation decays with the distance from the cutoff. I choose the bandwidths by a cross-validation procedure where the relationships between the main outcomes of interest and education is estimated with local linear regressions.

Second, within bandwidths, I estimate the baseline equation 2 by including separate age trend terms above and below the cutoff. The effect of the treatment is estimated by running a pooled regression on both sides of the cutoff point via OLS. To choose the form of the polynomial f (.) which fits the data well , I follow the procedure indicated in Lee and Lemieux (2010). I start with linear function. Thereafter, I add the set of bin dummies - used to graphically depict the data- to the regression and jointly test the significance of the bin dummies. After the test with linear regression, the bin dummies are not jointly significant for the difference outcomes. So, the polynomial f (.) is linear for the different estimations.

6 Results

6.1 Baseline results

I start by presenting graphical evidence of the effect of the 2002 Zambia educational reform on the main outcomes and the results of RDD estimation.

Figure 1 and Figure 2 plot respectively the mean of being HIV positive and the average of years of education by cohort in our sample. Figure 1 shows a large discontinuous jump in the average of years of the education and in the probability to achieve primary school among the first beneficiary cohort that benefit from 2002 educational reform. Among subsequent cohorts, the average of years of the education and the probability to achieve primary school continued to rise, coinciding with the continued expansion in access of education and with the extent to which they benefited from the reform. However, this improvement of women's education seems not to be returned into a clear drop in the HIV prevalence rate in this female population in general (Figure 2). The probability to be HIV positive is still decreasing, following the trend we observe prior the reform.

Following this graphical analysis, I run the RDD estimates of my baseline equation. Table 3 presents the results of the effect of education on HIV infection. I start by running a simple OLS estimates of the correlation between education and HIV while controlling for individuals ethnicity and the survey year fixed effects(cols 1-2). As we can observe, there is a strong and significant association between the total years of education completed by a woman and her probability to be HIV positive. This estimate is biased because of the endogeneity of education. The use of FPE and RDD estimates allows to correct the endogenous bias in the relation between education and HIV status. In cols.3-4, I present the RDD estimates of the effect of the reform on education. In line with graphical analysis, we can observe that the educational reform increases significantly the probability to achieve primary school. The estimates show that the reform increases the probability to achieve primary school by 12,3% [7]. Moreover, I present in the cols.5-6, the effect of the reform on women's probability to be HIV positive. The estimates show that women who benefit from the reform are more likely to be HIV positive[8]. Indeed, the reform is associated with 18% increases in women's probability to be HIV positive.

6.2 Geographic heterogeneous effects

The patterns that we observe in subsection 6.1 can dissimulate heterogeneous effects within the sample. Geographic differences in school supply could mitigate the effectiveness

[7]This relative effect is obtained by dividing the absolute effect of reform on the completion of primary education by its sample mean: $\frac{0.0785}{0.6355}$

[8]the non-parametric RDD confirm the results about Primary education and HIV in all sample see: Table 4

of the reform.To address this concern, I use official data on the increase in school building to assess the geographic differences in school supply. Table 1 presents the number of new schools per 10000 children aged 0-15 in each province between 1999 and 2005, meaning 3 years before and 3 years after the implementation of the reform. In Muchinga province, about 91 new basic schools were built per 10000 children aged 0-15 between 1999 and 2005 whereas about 64 new basic schools were built in Central. I use this index to split the sample into two groups of provinces, depending on whether the relative increase in school construction between 1999 and 2005 is below or above 77.1 schools per 10000 children[9]. The provinces with high supply of new schools are the provinces whose index is greater than or equal to 77.1. These are provinces that have had more new constructed schools per 10000 children between 1999 and 2005 to fill the classroom gap. The other provinces, ie those with an index less than 77.1, are considered to be provinces with low supply. In order to distinguish the effect of the reform in the two groups, I estimate a variant of our baseline equation where I account for the possible heterogeneity in the response within the two groups. The estimated equation is as follows:

$$
\begin{aligned}
Y_{ic} = \alpha + \lambda\, I_i^{low} + \beta_1\, I_c^{89-94} * I_i^{High} + \beta_2\, I_c^{89-94} * I_i^{Low} \\
+[I_c^{89-94} * f(dist_c) + I_c^{82-88} * f(dist_c)] * I_i^{High} \\
+[I_c^{89-94} * g(dist_c) + I_c^{82-88} * g(dist_c)] * I_i^{Low} \\
+\gamma\, X_{ic} + \varepsilon_{ic}
\end{aligned}
\tag{3}
$$

I_i^{High} and I_i^{Low} are dummy variables equal 1 if individual i is from a high supply of new schools provinces and low supply of new schools provinces respectively. These dummy variables represent the provinces groups. f(.) and g(.) are polynomial functions that control smoothly for the relative distance from birth cohort to the cutoff. All other variables are defined as in equation 2. I run OLS estimates and I use the approach of Cameron et al. (2011) to address the spatial correlation. I cluster the standard errors along ethnicity dimension.

[9] I discuss this threshold and use other thresholds in the robustness checks of section 8

The figures 3 show the effect of the reform in provinces with low and high supply provinces. As we can see, there is a sharp increase in female education in provinces with low supply. In opposite, there is no clear jump in female education in provinces with high supply. These results could be due to many reasons. It could be the result of a continuously increasing enrollment in certain provinces so that the slope of this trend is large enough to be much larger than any possible jump due to the reform. In other words, the reform was adopted spontaneously in provinces with low supply whereas it was adopted gradually in the provinces with high supply. The figures 4 plot the average (without trend) of HIV infection by cohort in the two groups of provinces. We can see that there is a positive change in the level of HIV infection in provinces with low supply whereas any change cannot be observed in provinces with high supply.

I run the equation 3 and the results are presented in Table 5. The first row reports the effect of the reform on various outcomes in provinces with low supply whereas the second row reports the results for provinces with high supply. As we can observe in cols.1-4, the reform has positive effect on both women's probability to completion primary education and women's total year of education in provinces with low supply. The effects are large in magnitude and statistically significant. In fact, young women who benefit from the reform are 11.4 percentage points more likely to complete primary education, which represents 18.7% of the sample mean. Relatedly, the reform increases women's total year of education by 15.4%. Moreover, the reform is associated with a large and significant increase in HIV infection in provinces with low supply. As we can observe in cols.5-6, women who benefit from the reform are 4.2 percentage points more likely to be HIV positive. This effect corresponds roughly to 29% of the average HIV positive rate in the sample. Consequently, the sharp increase in women's education induced by the free schooling reform has a positive and significant effect on HIV infection among women. In fact, one additional year increase in women's total years of education is associated with 25.7% in HIV rate among women.

In opposite to the provinces with low supply, the reform has no effect in the provinces with high supply, either on women's education, or on women's HIV infection. In point of fact, as we can see in Table 5, the reform has positive but non significant effect on education (cols 1-4). Also cols. 5-6 show that there is no effect on women's HIV infection in provinces with high supply.

7 Exploring the potential channels

In this section, I explore the potential channels underlying the positive effect of education on HIV among women. Women could contract HIV through different manners, starting from individual risky behavior towards HIV infection to the environmental or neighbourhood risk associated with the disease.

7.1 Exploring the role of risky behavior of educated women

I leverage the richness of DHS data to investigate in which extent the reform is associated with women's reproductive health and women's risky behavior towards HIV. In DHS surveys, women have been asked questions about their total number of lifetime sexual partners, their knowledge about contraceptive methods that help to avoid HIV. They have also been asked questions about which contraceptive methods that they use and their opinion on whether a husband can beat his wife if she refuses sex to him. Based on those information, I define four measures of women's reproductive health and women's risky behavior towards HIV:

Sexual_partner: is a dummy variable equal 1 if the respondent has had at least two sexual partners in her lifetime.

Avoid_HIV: is a dummy variable defined for single women and takes 1 if a woman reports never have sex or use condoms as contraceptive methods.

ABC: is a dummy variable that takes 1 if a woman reports to know the ABC of HIV prevention (Abstinence, Being faithful, Condom use).

"Beat wife who refuses sex to husband": is a dummy variable that takes 1 if a woman agrees that a man can beat her wife if she refuses sex to him

I assess the effect of the reform on each of these variables (in the group of provinces with low supply in new schools) by running a variant of the baseline equation. The estimates are presented in cols 3-6 of table 6. These results show that the reform has no association with women's sexual risk attitudes or women's reproductive health behaviors. The effect of the reform is not significant for any of the four outcomes. This results are suggestive that there is no difference in health attitudes between women who benefit from the reform and those who do not. As a result, the educational difference in HIV infection could not be explained by differences in health behavior between beneficiaries and the non-beneficiaries.

7.2 Exploring the role of urbanization

In the subsection 7.1, I argue that the positive effect of education on HIV infection cannot be attribute to any change in the risky behavior of educated women. The increase in HIV infection that we observe among educated women could be explained by the environmental risk of HIV infection faced by educated women. In this section, I explore whether the effect is due to urban residency of educated women. The intuition behind this channel is guided by the fact that, according to the DHS data, HIV infection rate among men is 15.22% in urban areas and 8.31% in rural (Table 2). In other words, women living in urban face a higher risk of engaging with an infected man relatively to those living in rural areas. If the urbanization mechanism is at work, it will return into two implications. First, we should expect the reform to be associated with a high urban residency, ie cohorts that benefit from the free education policy will be more likely to reside in urban areas. Second, the reform should be associated with a positive and significant effect on both education and HIV infection in urban areas, but should affect only education in rural areas. I test these predictions and present the results in Table 7. Col.1 presents the results where I run the baseline equation by using urban residency as dependent variable. The estimates show that cohorts that benefited from the reform are 5.6% more likely to live in urban areas, representing 12% of urban residency in the sample. Consequently, one additional year of female education increases urban residency by 16%. Furthermore, cols 2-4 show that the reform has positive et significant effect on both education and HIV infection in urban areas. However, cols 5-7 show that the reform has positive effect on education (cols. 5-6), but no impact on HIV infection in rural areas (col7). Also, we can observe that the effect on primary education is higher in rural areas than in urban areas(9.5% vs 23%). Then, we can emphasis that despite of its positive impact on education, the reform is not associated with a positive effect on HIV infection in rural areas. As a result, urbanization of educated women seems to be a mechanism at work.

8 Robustness

In this part, I perform a couple of robustness check with the purpose of coming up with some potential issues that could drive my main results.

8.1 Migration

To investigate the geographic heterogeneity in the effect of the reform on education and HIV infection, one must know in which province each women was living at the time of primary education. Since such information is not in DHS, I assign each woman to her province of residence at the time of the survey, assuming that women do not move from their initial province of residence after primary education. The measurement error that stems from this assignment could bias the main results if the reform induces inter provincial migration among women. To address this issue, I rely on information provided by Zambia DHS on migration. Women have been asked how long they have been living at their current place of residence. The answer is "always" if they always live at that place. If not, they report the total years spent at the current place of residence. Unfortunately, only the 2018 DHS data reports the previous provinces of residence. Then, I use these information to construct a variable that accounts for inter provincial migration after primary education. And I run the baseline estimates on this variable to investigate whether the reform is associated with migration.

Table 8 presents the results. First, Only 16% of the sample migrate between provinces after primary education. Second, the estimates show that the reform has no effect on migration between provinces. This results suggest that our findings is not bias by inter-provincial migration.

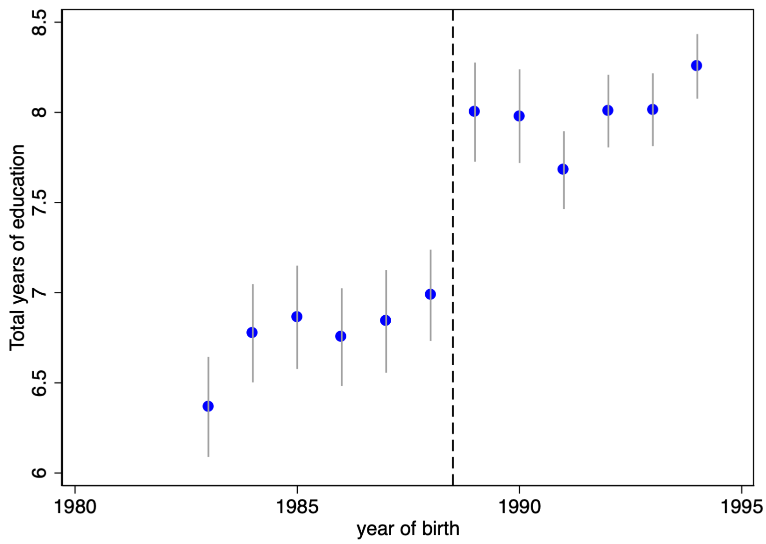
8.2 Sensitivity to the threshold

To analyze the heterogeneous effect of the reform, I split the sample by using an index that accounts for the number of schools built in each provinces. By using a threshold of 77.1 schools per 10000 children, I identify four (04) provinces with high supply in new basic schools and six (06) provinces with low supply in new basic schools. In this section, I investigate how the estimates are sensitive to the threshold. To do so, I change the threshold to split the sample into three (03) provinces with high supply and seven (07) provinces with low supply and I estimate the baseline equation in each group giving this subdivision. The results are presented Table 9. The results are quite similar to the patterns that we observe in subsection 6.2. They are consistent with the main findings of this paper.

9 Conclusion

It is believed that increasing female education is an important step in improving maternal health in developing countries. However, the direction of the effect of education specifically on HIV disease is very controversial in the literature of health economics. This paper assesses the effet of female education on the HIV infection by using 2002 free primary education reform in Zambia. I find, through RDD, that women' access to education contributed to a significant increase in women's HIV infection. The effect appear to have been driven by urbanization and unprotected sex. These results underline the necessity to support women's education with education on safe and secure sex.

Figure 1: Education, DHS 2014, 2018



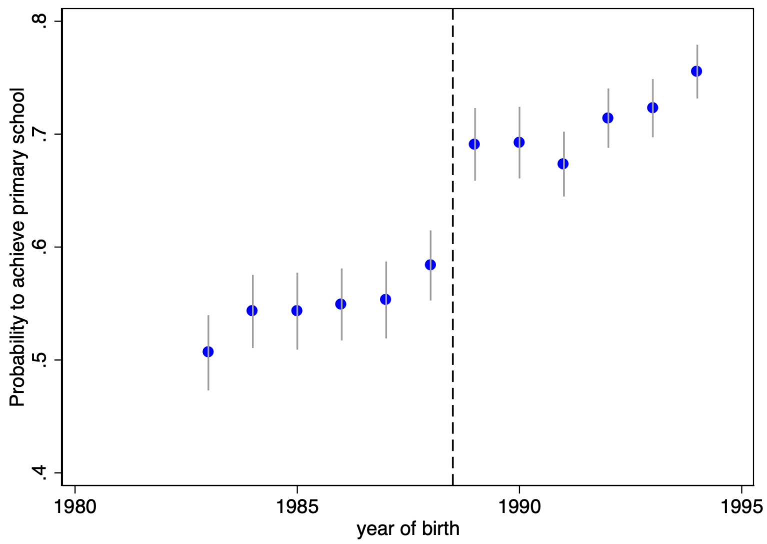
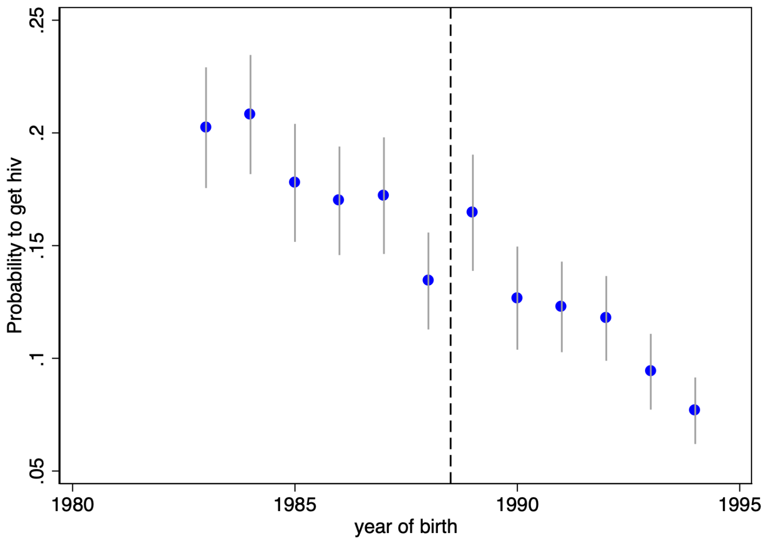
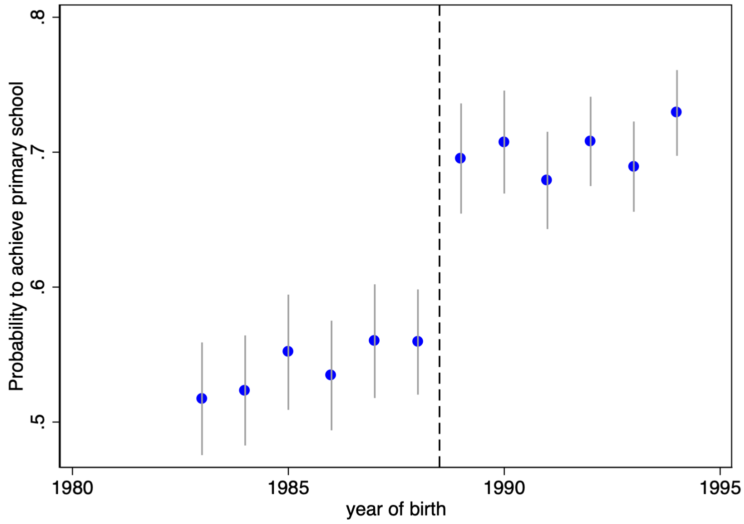


Figure 2: Women probability to get HIV, DHS 2014-2018

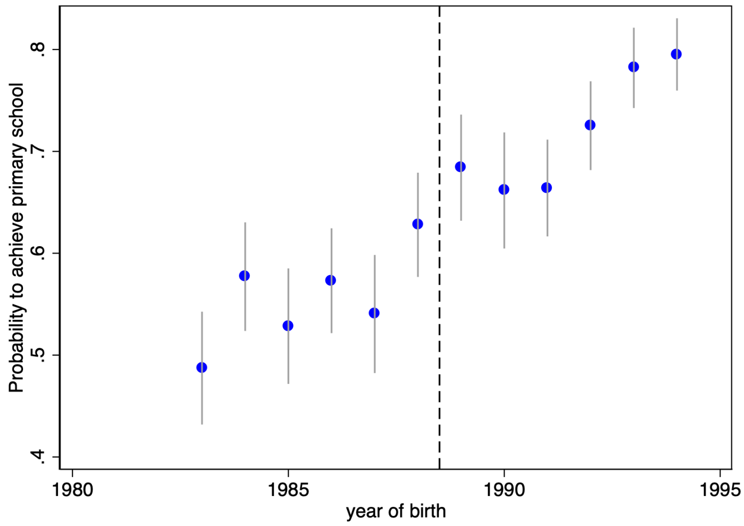


Notes: Each point represents the average of outcome variable by cohort. Each full line represents the interval of confidence of the point that it contains. The vertical dash separates the last non beneficiary cohort (1988) and the first beneficiary cohort (1989) of the reform.

Figure 3: Primary education by group of province, DHS 2014, 2018

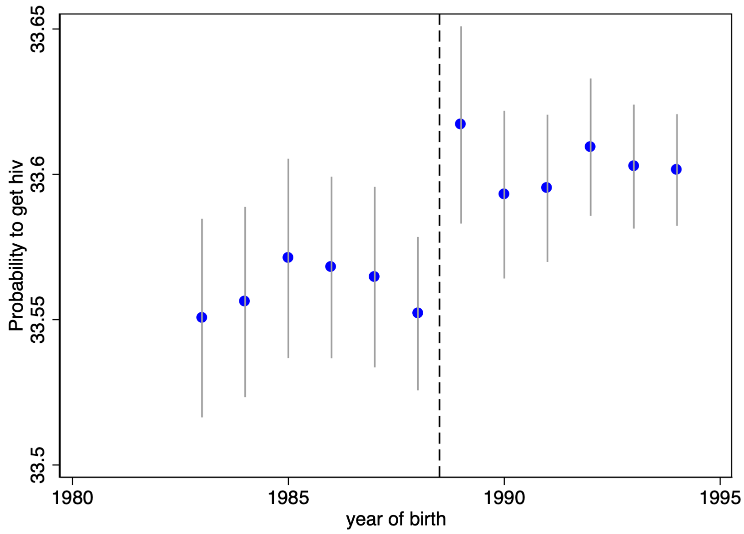


(a) Provinces with a **low** supply in new basic schools

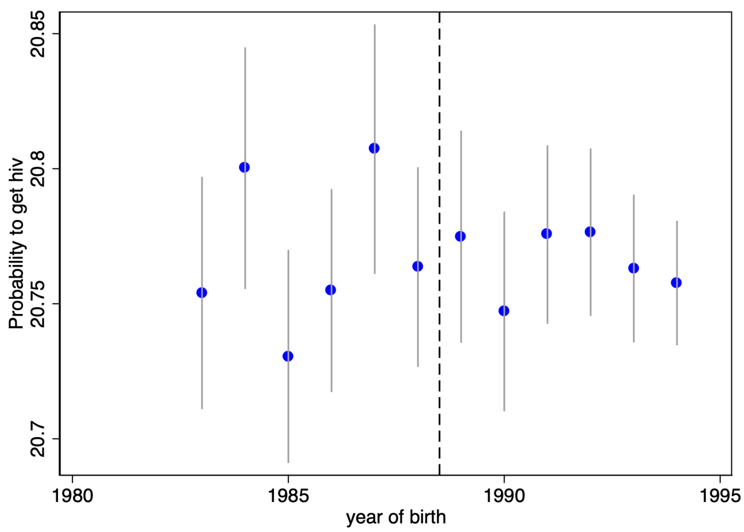


(b) Provinces with a **high** supply in new basic schools

Figure 4: Women probability to get HIV by group of province , DHS 2014, 2018



(a) Provinces with a **low** supply in new basic schools



(b) Provinces with a **high** supply in new basic schools

Notes (Figure 4a and Figure 4b): Each point represents the average of outcome variable by cohort. Each full line represents the interval of confidence of the point that it contains. The vertical dash separates the last non beneficiary cohort (1988) and the first beneficiary cohort (1989) of the reform. **Figure 4b and Figure 4a are without trend.**

Table 1: Construction of schools in Zambia

	Provinces	Number of school built between 1998 and 2005	Number of children under 15 years old	Index
1	muchinga	296	32577	90.86
2	western	314	38253	82.09
3	copperbelt	604	78239	77.20
4	north western	252	32654	77.17
5	northern	390	50610	77.06
6	southern	527	71048	74.18
7	eastern	486	70950	68.50
8	luapula	292	44360	65.83
9	central	366	56964	64.25
10	lusaka	447	84622	52.82

Notes: Index column contains the number of new basic schools built between 1998-2005 per 10000 children aged 0-15. Source: Ministry of Education, Government of Zambia; census 2010.

Table 2: some statistics

		Obs.	Mean	Confidence interval (5%)
	HIV	11,492	0.1425	(0.1361, 0.1489)
	Total years of education	11,486	7.4294	(7.3582, 7.5006)
	Primary education	11,489	0.6354	(0.6266, 0.6442)
	Urban residency	11,492	0.4755	(0.4664, 0.4847)
	Sexual partner	11,492	0.2615	(0.2535, 0.2695)
	knowledge of ABCs	11,326	0.7480	(0.7400, 0.7560)
	Use methods to avoid HIV	1,554	0.1197	(0.1035, 0.1358)
	Beat wife who refuses sex to husband	11,238	0.3248	(0.3161, 0.3335)
	Movement between provinces	4,654	0.1667	(0.1560, 0.1775)
	Rural exodus	5,935	0.1503	(0.1412, 0.1594)
	Reform	11,492	0.5373	(0.5282, 0.5464)
	Group	11,492	0.6339	(0.6251, 0.6427)
	Men's HIV rate	21,643	0.1116	(0.1074, 0.1158)
Rural	HIV	6,027	0.0881	(0.0809, 0.0953)
	Total years of education	6,024	5.9421	(5.8511, 6.0330)
	Primary education	6,025	0.4822	(0.4695, 0.4948)
	Sexual partner	6,027	0.2456	(0.2347, 0.2564)
	knowledge of ABCs	5,918	0.7349	(0.7236, 0.7461)
	Use methods to avoid HIV	2,184	0.0989	(0.0864, 0.1114)
	Beat wife who refuses sex to husband	5,895	0.3997	(0.3872, 0.4122)
	Movement between provinces	2,614	0.1186	(0.1062, 0.1310)
	Reform	6,027	0.5291	(0.5165, 0.5417)
	Group	6,027	0.6531	(0.6410, 0.6651)
	Men's HIV rate	12,722	0.0831	(0.0783, 0.0879)
Urban	HIV	5,465	0.2026	(0.1919, 0.2132)
	Total years of education	5,462	9.0698	(8.9763, 9.1632)
	Primary education	5,464	0.8044	(0.7938, 0.8149)
	Sexual partner	5,465	0.2790	(0.267, 0.2909)
	knowledge of ABCs	5,408	0.7624	(0.7510, 0.7737)
	Use methods to avoid HIV	2,308	0.2175	(0.2007, 0.2343)
	Beat wife who refuses sex to husband	5,343	0.2422	(0.2307, 0.2537)
	Movement between provinces	2,040	0.2284	(0.2102, 0.2467)
	Reform	5,465	0.5464	(0.5332, 0.5596)
	Group	5,465	0.6128	(0.5999, 0.6257)
	Men's HIV rate	8,921	0.1522	(0.1448, 0.1597)
High new offer	HIV	4,207	0.1362	(0.1258, 0.1466)
	Total years of education	4,207	7.6202	(7.5018, 7.7385)
	Primary education	4,207	0.6499	(0.6354, 0.6643)
	Urban residency	4,207	0.5030	(0.4879, 0.5181)
	Sexual partner	4,207	0.3002	(0.2864, 0.3141)
	knowledge of ABCs	4,172	0.7845	(0.7720, 0.7970)
	Use methods to avoid HIV	1,605	0.1564	(0.1386, 0.1742)
	Beat wife who refuses sex to husband	4,106	0.3422	(0.3277, 0.3567)
	Movement between provinces	1,659	0.1627	(0.1450, 0.1805)
	Rural exodus	2,098	0.1606	(0.1449, 0.1764)
	Reform	4,207	0.5415	(0.5264, 0.5565)
	Men's HIV rate	7,523	0.1082	(0.1012, 0.1152)
Low new offer	HIV	7,285	0.1462	(0.1381, 0.1543)
	Total years of education	7,279	7.3191	(7.2301, 7.4081)
	Primary education	7,282	0.6270	(0.6159 , 0.6381)
	Urban residency	7,285	0.4597	(0.4483 , 0.4712)
	Sexual partner	7,285	0.2391	(0.2293 , 0.2489)
	knowledge of ABCs	7,154	0.7267	(0.7164, 0.7371)
	Use methods to avoid HIV	2,257	0.1936	(0.1773 , 0.2099)
	Beat wife who refuses sex to husband	7,132	0.3148	(0.3040 , 0.3256)
	Movement between provinces	2,995	0.1689	(0.1555 , 0.1824)
	Rural exodus	3,837	0.1446	(0.1335, 0.1558)
	Reform	7,285	0.5349	(0.5235, 0.5464)
	Men's HIV rate	14,120	0.1134	(0.1082, 0.1186)

Table 3: Education and HIV (Full sample)

	(1) HIV	(2) HIV	(3) Total years of education	(4) primary	(5) HIV	(6) HIV
Total years of education	0.0044***	0.0041***				
	(0.0011)	(0.0011)				
Reform			0.7642***	0.0785***	0.0268**	0.0257**
			(0.2194)	(0.0194)	(0.0124)	(0.0123)
Birth cohort polynomial			Linear	Linear	Linear	Linear
Ethnicity FE	No	Yes	Yes	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,486	11,486	11,486	11,489	11,492	11,492
Mean	0.1426	0.1426	7.4294	0.6354	0.1425	0.1425

Notes: Columns 1 and 2 report the results of OLS estimate. Columns 3, 4, 5 and 6 report the results of estimate of different regression based on Equation 2 and include a linear polynomial control for year of birth. Standard standard errors are reported in parentheses. ***,**, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 4: Education and HIV (Non-parametric RDD)

	All provinces		Low supply in new basic schools #		high supply in new basic schools	
	Primary education	HIV	Primary education	HIV	Primary education	HIV
	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth below	5.16	4.15	5.22	2.04	5.02	4.08
Bandwidth above	2.01	5.40	5.28	5.68	5.57	5.02
difference (jump)	0.1120	0.026	0.1331	0.0652	0.0843	0.0073
	(0.0298)	(0.0162)	(0.0258)	(0.0250)	0.0354	(0.0271)

Notes: # Provinces with a low supply in new basic schools are Northern, Eastern, Lusaka, Southern, Central and Luapula. Each column report the result of Non-parametric RDD. All include a linear polynomial control for year of birth and the bandwidths are chosen by a cross-validation procedure.

Table 5: Education and HIV (RDD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary education		Total years of education		HIV	
$I_c^{89-94} * I_i^{low}$	0.1241***	0.1178***	1.1710***	1.1278***	0.0423**	0.0423**
	(0.0225)	(0.0187)	(0.2387)	(0.2247)	(0.0187)	(0.0184)
$I_c^{89-94} * I_i^{high}$	0.0132	0.0120	0.1471	0.1337	-0.0005	-0.0031
	(0.0162)	(0.0156)	(0.1774)	(0.1729)	(0.0058)	(0.0061)
Ethnicity FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort polynomial	Linear	Linear	Linear	Linear	Linear	Linear
Observations	11,489	11,489	11,486	11,486	11,492	11,492
Mean of dep. var. in G	0.627	0.627	7.319	7.319	0.146	0.146
Mean of dep. var. in (1-G)	0.6499	0.6499	7.620	0.136	0.136	

Notes: Each column reports the result of estimate from a different regression based on Equation 3. All include a linear polynomial control for year of birth. The first line represents the provinces of low supply in new basic schools and the second line is the provinces of high supply in new basic schools. Standard errors are reported in parentheses. ***,**, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 6: Potential channels: Effect on reproductive health behaviors

	(1) Primary education	(2) HIV	(3) Sexual partner	(4) Use methods to avoid HIV	(5) ABC	(6) Beat wife who refuses sex to husband
Reform	0.1157***	0.0422**	0.0070	-0.0276	0.0212	-0.0200
	(0.0188)	(0.0184)	(0.0128)	(0.0336)	(0.0176)	(0.0147)
Birth cohort polynomial	Linear	Linear	Linear	Linear	Linear	Linear
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,282	7,285	7,285	2,257	7,154	7,132
Mean	0.6270	0.1462	0.2391	0.1936	0.7267	0.3148

Notes: Each column report the result of estimate from a different regression based on Equation 2. All include a linear polynomial control for year of birth. Standard errors are reported in parentheses. ***,**, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 7: Potential channels: Urbanization

		Urban			Rural		
	Urban residency	Primary education	Total years of education	HIV	Primary education	Total years of education	HIV
Reform	0.0559***	0.0760***	0.9213***	0.0778**	0.1128***	0.9479***	-0.0019
	(0.0135)	(0.0230)	(0.2517)	(0.0383)	(0.0203)	(0.2579)	(0.0143)
Birth cohort polynomial	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,285	3,348	3,346	3,349	3,934	3,933	3,936
Mean	0.4597	0.7948	8.8918	0.2063	0.4842	5.9812	0.0950

Notes: Each column report the result of estimate from a different regression based on Equation 2. All include a linear polynomial control for year of birth. Standard errors are reported in parentheses. ***,**, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 8: Movement between provinces after primary education

	(1) Movement between provinces	(2) Total years of education	(3) Primary education
Reform	0.0043	0.4247***	0.0638***
	(0.0118)	(0.1326)	(0.0124)
Ethnicity FE	Yes	Yes	Yes
Birth cohort polynomial	Linear	Linear	Linear
Observations	4,654	4,688	4,688
Mean of dep. var.	0.1667	7.3471	0.6143

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Education and HIV (RDD), provinces with **first seven low** supply in new basic schools

	(1) Primary education	(2) Primary education	(3) Total years of education	(4) Total years of education	(5) HIV	(6) HIV
$I_c^{89-94} * I_i^{Low}$	0.1074***	0.1028***	1.0386***	1.0086***	0.0353**	0.0350**
	(0.0209)	(0.0177)	(0.2421)	(0.2276)	(0.0163)	(0.0159)
$I_c^{89-94} * I_i^{High}$	0.0128	0.0110	0.0900	0.0722	0.0018	-0.0011
	(0.0248)	(0.0241)	(0.1747)	(0.1689)	(0.0117)	(0.0124)
	(0.0260)	(0.0379)	(0.2203)	(0.3467)	(0.0362)	(0.0179)
Ethnicity FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort polynomial	Linear	Linear	Linear	Linear	Linear	Linear
Observations	11,489	11,489	11,486	11,486	11,492	11,492

Notes: Each column report the result of estimate from a different regression based on Equation 1. All include a linear polynomial control for year of birth. The provinces with supply in new basic schools are: Muchinga, western and Copperbelt. The line represents the group of low supply in new basic schools and the second line is the group of high supply in new basic schools. Standard standard errors are reported in parentheses. ***,**, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 10: Contraception method in baseline group

	Current marital status	Abstinence	No contraception method	Condom	Other method different from condom	Total	Total
	Without partner	14.62	60.79	4.74	19.85	100	30.38
High	With partner/husband	0	48.87	3.2	47.93	100	69.02
	Total						100
	Without partner	12.71	60.62	2.93	23.74	100	38.15
Low	With partner/husband	0	50.54	2.92	46.54	100	61.85
	Total						100

Notes: Each cell contains the proportion of women using a category of contraceptive method according to having a given marital status.

References

Abdulghani Al-Shuaibi. The importance of education. 01 2014.

Siwan Anderson. Legal origins and female hiv. *American Economic Review*, 108(6): 1407–39, 2018.

Manuela Angelucci and Daniel Bennett. Assortative matching under asymmetric information: Evidence from malawi. *American Economic Review*, 107(5):154–57, 2017.

A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011.

Pierre-André Chiappori, Sonia Oreffice, and Climent Quintana-Domeque. Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, 120(4):659–695, 2012.

Janet Currie and Enrico Moretti. Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly journal of economics*, 118(4):1495–1532, 2003.

David Cutler and Grant Miller. The role of public health improvements in health advances: the twentieth-century united states. *Demography*, 42(1):1–22, 2005.

David M Cutler and Adriana Lleras-Muney. Understanding differences in health behaviors by education. *Journal of health economics*, 29(1):1–28, 2010.

Damien De Walque. How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. *Journal of Development Economics*, 84(2):686–714, 2007.

Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4): 795–813, 2001.

Esther Duflo, Pascaline Dupas, and Michael Kremer. Education, hiv, and early fertility: Experimental evidence from kenya. *American Economic Review*, 105(9):2757–97, 2015.

Pascaline Dupas. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 3(1):1–34, January 2011. doi: 10.1257/app.3.1.1. URL https://www.aeaweb.org/articles?id=10.1257/app.3.1.1.

Jane G Fortson. The gradient in sub-saharan africa: socioeconomic status and hiv/aids. *Demography*, 45(2):303–322, 2008.

Susan Godlonton and Rebecca L Thornton. Learning from others' hiv testing: Updating beliefs and responding to risk. *American Economic Review*, 103(3):439–44, 2013.

Simon Gregson, Heather Waddell, and Stephen Chandiwana. School education and hiv control in sub-saharan africa: from discord to harmony? *Journal of International Development: The Journal of the Development Studies Association*, 13(4):467–485, 2001.

Michael Grossman. 1. on the concept of health capital and the demand for health. In *Determinants of Health*, pages 6–41. Columbia University Press, 2017.

James R Hargreaves and Judith R Glynn. Educational attainment and hiv-1 infection in developing countries: a systematic review. *Tropical Medicine & International Health*, 7 (6):489–498, 2002.

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635, 2008.

Daniela Iorio and Raül Santaeulàlia-Llopis. Education, hiv status and risky sexual behavior: how much does the stage of the hiv epidemic matter? *Available at SSRN 2820991*, 2016.

Robert Jensen and Adriana Lleras-Muney. Does staying in school (and not working) prevent teen smoking and drinking? *Journal of Health Economics*, 31(4):644–657, 2012.

David S Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355, 2010.

Adriana Lleras-Muney. The relationship between education and adult mortality in the united states. *The Review of Economic Studies*, 72(1):189–221, 2005.

Monica Magadi and Muluye Desta. A multilevel analysis of the determinants and cross-national variations of hiv seropositivity in sub-saharan africa: evidence from the dhs. *Health & place*, 17(5):1067–1083, 2011.

Monica A Magadi. Understanding the urban–rural disparity in hiv and poverty nexus: the case of kenya. *Journal of Public Health*, 39(3):e63–e72, 2017.

Bentry Nkhata, Mumba Elizabeth, Chikalanga Israel, Anne Sikwibele, et al. Advancing girl's education in zambia. Technical report, Ministry of Education, 1998.

Emily Oster. Sexually transmitted infections, sexual behavior, and the hiv/aids epidemic. *The Quarterly Journal of Economics*, 120(2):467–515, 2005.

Emily Oster. Hiv and sexual behavior change: Why not africa? *Journal of health economics*, 31(1):35–49, 2012.

Mead Over and Peter Piot. *HIV infection and sexually transmitted diseases.* Population, Health and Nutrition Division, Population and Human Resources, 1991.

Christina Paxson and Anne Case. Hiv risk and adolescent behaviors in africa. *American Economic Review*, 103(3):433–438, 2013.

Roland Pongou and Roberto Serrano. Fidelity networks and long-run trends in hiv/aids gender gaps. *American Economic Review*, 103(3):298–302, 2013.

Malcolm Potts, Daniel T Halperin, Douglas Kirby, Ann Swidler, Elliot Marseille, Jeffrey D Klausner, Norman Hearst, Richard G Wamai, James G Kahn, and Julia Walsh. Reassessing hiv prevention. *Science*, 320(5877):749–750, 2008.

Rand L Stoneburner and Daniel Low-Beer. Population-level hiv declines and behavioral risk avoidance in uganda. *Science*, 304(5671):714–718, 2004.

Nicola Swainson. Redressing gender inequalities in education. a review of constraints and priorities in malawi, zambia, and zimbabwe. 1995.

Stephanie Simmons Zuilkowski, Gunther Fink, Corrina Moucheraud, and Beatrice Matafwali. Early childhood education, child development and school readiness: Evidence from zambia. *South African Journal of Childhood Education*, 2(2):20, 2012.