From Predictive to Protective? The Changing Relationship between HIV and Education

[DRAFT: PLEASE DO NOT CITE]

Elizabeth Gummerson

4/1/2011

*Contact: Elizabeth Gummerson, (lizbeth@princeton.edu), 219 Wallace Hall, Office of Population Research, Princeton University, Princeton, NJ 08544

Abstract:

This study uses longitudinal HIV prevalence data from Mali, Tanzania Kenya and Zambia to examine whether the positive relationship between educational attainment and HIV is reversing. My findings support previous scholars' hypotheses that the relationship between HIV and education has begun to reverse in Africa. Although I find that, at the regional level, the relationship is still positive across age groups, it is much weaker for the youngest generation. Furthermore, when I examine the relationship at the individual level, using region and country fixed effects, I find no positive association for the youngest cohort. Secondarily, I evaluate evidence for two commonly hypothesized explanations for such a change- the erosion of educational infrastructure in high prevalence areas and the adoption of protective knowledge among the educated. I find evidence consistent with the hypothesis that education is becoming protective in the present state of the epidemic; regions with higher levels of adult education at baseline experience larger drops in HIV prevalence over the time period. I do not find evidence that educational attainment is eroding over time in higher prevalence areas. In fact, there is a consistent trend towards increased years of schooling for children in both high and low prevalence regions.

Mark Twain once famously quipped “Be careful about reading health books. You may die of a misprint.” Despite his skepticism, there is a preponderance of evidence that in fact, people who read books (health or otherwise) actually enjoy longer life and better health. There is a robust body of research demonstrating that being well educated is associated with better outcomes across a broad range of health measures. Moreover, scholars have documented this relationship repeatedly across a variety of contexts and time periods. In part, the positive association between education and health may be explained by access to information. People with more education may know more about health risks and consequently exhibit healthier behaviors. Education may also affect health indirectly through its association with wealth. More educated people tend to have more income, and higher incomes can confer a host of health benefits such as access to better care, less physically demanding work, better nutrition and lower levels of psychological stress (Cutler & Lleras-Muney 2006; Ross & Wu 1995). Consequently, having more years of education is generally viewed as protective against a host of disease risks.

One very notable exception to this pattern has been HIV in sub-Saharan Africa. Data collected from Africa in the early decades of the epidemic demonstrated that there was a robust *positive* association between education and HIV infection (Gregson et al. 2001; Hargreaves & Glynn 2002; Fortson 2010). This association is likely not because of ‘misprints’ or other failings of education, but rather because many of the correlates of education are also risk factors for HIV. Areas with high numbers of educated people also tend to be urban, have more developed economies and transport networks, and larger migrant labor populations- all of which are associated with the spread of HIV. In addition, at the individual level, people who are more educated tend to marry later and may have more sexual partners before marriage, which can potentially increase exposure to HIV risk (Singh & Samara 1996; Gregson et al. 2001; Fortson 2009). Consequently, having more years of education has become largely accepted in the literature as a risk factor for HIV in sub-Saharan Africa.

However, there are reasons to think that this positive relationship between education and HIV risk may be changing as the epidemic matures. A variety of scholars in the field have hypothesized that the present positive association may ultimately be replaced by a negative one (Gregson et al. 2001; Hargreaves & Glynn 2002; Fortson 2008). The literature implies at least two pathways through which this reversal could take place. The first is through a general societal deterioration resulting from high HIV rates. For some time, observers and scholars have warned that the particular characteristics of the HIV epidemic– that it targets working age adults, that it clusters in households, that the rates of infection are higher among the educated – mean that the effects of the disease will go far beyond the afflicted households and result in widespread societal deterioration. In this scenario, schools, hospitals and civil institutions deteriorate as a result of the deaths of a large swath of the educated laboring population. In addition, parents in affected communities, who perceive a high risk of early mortality, may reduce investment in their children's education because they do not expect to live long enough to reap the benefits of the investments.

Recent research lends some support to these bleak projections. Fortson (2011) uses Demographic and Health Surveys (DHS) in sub-Saharan Africa to show that areas with higher HIV prevalence have larger schooling declines among those who were of school going age when the epidemic hit. Case and Paxson (2009) document deterioration of health infrastructure in areas of high HIV prevalence. Behrman et al (2010) show, in a high prevalence area of Malawi, that mothers who perceive a higher risk of HIV reduce investment in children's education. In addition, Akbulut-Yuksel and Turan (2010) use the DHS surveys to document that children in high HIV areas accumulate comparatively less education than their parents, even within households without HIV infected members. Collectively, this research suggests that HIV is associated with losses in education and health that go beyond the individuals, or even the generation, directly infected. Over time, this process (which I refer to subsequently as *educational erosion*) could plausibly work to reverse the positive relationship between HIV and education. If enough erosion of infrastructure and reduced investment in human capital were to occur, high HIV prevalence regions could eventually become associated with lower levels of education rather than higher.

The second path through which the relationship between education and HIV might reverse is simply through the dissemination of information about the disease and its pathways. Historically, as disease mechanisms become better understood, the more highly educated have adopted protective behaviors earlier and an education gradient in mortality and morbidity has subsequently emerged. For example, in the 1960s, there was little difference in mortality for older American men across education levels, but once the risks of smoking become understood (beginning with the release of the Surgeon General's report on smoking in 1964), better educated men's mortality dropped throughout the 70s and 80s, becoming significantly below that of men with less education. (Cutler & Lleras-Muney 2006; Crimmins & Saito 2001; Preston & Elo 1995; Feldman et al. 1989).

In the case of HIV in Africa, it is plausible that this process is only now taking place. Transmission pathways were imperfectly understood until the late 1980s and early 1990s. Even once transmission pathways were identified in the scientific community, disinformation and political resistance to the new knowledge slowed its impact in sub-Saharan Africa. Influential figures such as the then President of South Africa publicly questioned the scientific consensus, effectively obfuscating the public narrative about how HIV was spread and how it might be prevented. By the 2000s, however, the public skepticism about the pathways to HIV infection had been largely overcome, and there now seems to be widespread acceptance in both the scientific and popular press on the causes of HIV. Now that the knowledge of how HIV spreads and can be prevented has become better accepted in sub-Saharan Africa, we might expect to see the more highly educated adopting protective behaviors first, leading the relationship between education and HIV eventually to develop the more familiar inverse association.

Recent work by Case and Paxson (2011) lends support to this theory (which I refer to henceforth as the '*protective knowledge*' pathway). The paper examines behavioral changes among younger women and finds that young women in high prevalence areas respond by decreasing teen sexual activity outside of marriage and by entering marriage earlier. More importantly for this analysis, they find that the behavioral effect on early marriage is magnified for highly educated girls. These findings suggest both that younger cohorts of Africans may be changing their behaviors in response to the epidemic, and that there may be some differential in response by education level. If the more highly educated adopt protective behaviors earlier and/or at higher rates, we may ultimately see a reversal in the positive correlation between HIV and education.

The remainder of this paper examines these questions by using the most recently released DHS data from the four countries where there have been multiple rounds of HIV testing- Kenya, Mali, Tanzania and Zambia. I take advantage of the newly released longitudinal data to look at the evolution of HIV prevalence and education levels in different regions and ask two questions- 1) *Is there evidence that the positive correlation between education and HIV is changing?* and 2) *What is the available evidence for either the 'educational erosion' or the 'protective knowledge' pathways?* I find that for the youngest cohort in the study, the relationship between HIV and education is significantly weaker than for older cohorts, although still slightly positive at the regional level. The association between education and HIV does not appear to strengthen as the cohort ages, nor does the difference appear to be driven by differential mortality. I also find that among a subsample of individuals in the younger cohort, HIV prevalence increases over time for the poorly educated, but not for the highly educated. I do not find strong support for the educational erosion pathway, and the trend toward increased schooling for children is similar in both high and low prevalence areas. I do, however, find evidence consistent with the protective knowledge pathway; regions with higher educational levels at baseline experience larger drops in HIV prevalence over the period in question, which is consistent with the hypothesis that education may be becoming protective in the present stage of the epidemic.

Data and Measures

The data for this analysis is drawn from the Demographic and Health Surveys (DHS). The DHS are large nationally representative household surveys that cover household members' education, nutrition, fertility, and health. More recently, some African DHS have asked a subsample of adult respondents to provide blood samples for HIV testing[1]. The four countries examined here - Kenya, Mali, Tanzania and Zambia- are, to date, the only ones for which two rounds of HIV testing data are available. For Mali and Tanzania, the two waves of HIV and household data were collected approximately five years apart. In Mali, HIV tests were collected in 2001/02 and 2006. In Tanzania they were collected during 2003/4 and again in 2008. For Kenya and Zambia, the two waves were collected six years apart. In Kenya, HIV prevalence data was collected in 2003 and again in 2009. For Zambia, it was collected in 2001 and 2007. The analysis here rests upon measures of HIV prevalence and educational attainment calculated at the regional level at the two points in time. Regions are identified within each country by using the geographic regions as specified in the DHS and further dividing them by urban and rural sectors where applicable, resulting in 91 distinct regions. I calculate regional HIV prevalence using information on 39,913 adults aged 15-49 in the HIV tested subsample.[2]

I also use two measures of regional educational attainment, which are calculated using reported education of the 149,116 adults aged 15-49. The first measure is mean years of completed education among adults in the region. However, because some of the respondents are

[1]Response rates were high and analysis on a subset of countries has shown that bias from non-response is small l(Mishra et al. 2006).
[2] DHS surveys provide HIV sample weights to adjust the tested subsample to be representative of the sample as a whole, and I use them here to compute regional HIV prevalence.

still quite young and could still be completing education, I also use percent of adults aged 15-49 who have completed primary schooling as an additional measure of regional educational levels. Both of the education measures come from the household questionnaires. In the section testing educational erosion, I examine changes in children's schooling over time. To this end, I use the responses to the household survey to calculate regional attendance rates and mean years of completed education based on 82,304 children who are aged 7-14 in the sample households.

Table 1 presents summary statistics of regional HIV prevalence and regional educational outcomes by country and wave. Kenya, Tanzania and Zambia are fairly typical of the East and Central African region in that they all have relatively high levels of adult education and correspondingly high levels of HIV. Mali is the lone West African country and has a different pattern with lower levels of adult education and lower levels of HIV. Over the course of the two waves, there is a general trend of a slight increase in mean years of education completed and relatively little change in percentage of adults with primary education. There are also modest drops in mean regional HIV prevalence across all four sample countries over time. As mentioned above, the analysis relies on regional aggregations of individual level data, and Table 2 contains basic descriptive information on the individuals from which the regional means are constructed. Within both the adult and child samples, the basic age and gender distributions are comparable across both countries and waves. As noted at the regional level, there is an upward trend overall in educational attainment and Mali has lower measures of education for both adults and children than the other countries.

Evidence for Change

I approach the question of whether the relationship between education and HIV is reversing by splitting the sample into three different age cohorts and looking at the relationship in successively younger cohorts. I divide the adults into a 'youngest' cohort who are aged 15 to 24 in the first wave, a 'middle' cohort who are aged 25-34 in the first wave, and an 'oldest' cohort who are aged 35 to 44 in the first wave. I am then able to follow the cohorts as they age in the second wave, which is either five or six years later, depending upon the country. The youngest cohort is of particular interest because they would have been born between 1976 and 1987 and entered adolescence during the 1990s and early 2000s. In effect, this group is the first that could have begun making decisions about sexual behaviors in a period when the transmission pathways of HIV were widely accepted and the means of prevention were known and publicized. To begin examining the question of whether the relationship between HIV prevalence and education levels differs by cohort, I estimate a simple least squares model of the relationship between regional HIV prevalence and education for each cohort as below:

$$(1) \qquad HIV_{rcw} = \beta_0 + \beta_1 Ed_{rcw} + \theta X_{rcw} + \varepsilon$$

where the dependent variable HIV_{rcw} is the regional HIV prevalence for adults in region *r*, within age cohort *c* at in wave *w*. Independent variable Ed_{rcw} is the regional mean for the educational outcome (either 'years of education' or 'primary school completion rate') for adults in age cohort *c* and region *r* at wave *w*. X is a vector of control variables for the same region, cohort and wave, including year of the survey measurement and an additional control for the duration of the epidemic.[3] This is included to account for the possibility that areas with many highly educated adults may also be those with the longest running epidemics, and therefore changes in prevalence in those areas could be related to duration of the epidemic rather than differences in education levels in the populace.

Table 3 shows the coefficients of ordinary least squares regressions of regional HIV prevalence on the mean regional education level for each cohort in the first wave (model (1)). Figures 1a and 1b show the same regression lines in graphical form. For both measures of education, the relationship is positive and statistically significant across age cohorts. However, the coefficient for the youngest cohort is less than half those for the older two cohorts. For adults between 25 and 34, one additional year in the mean education for adults in the region is associated with a correspondingly higher level of HIV prevalence of about 2.5 percentage points, whereas for 15-24 yr olds, the same difference in mean years of education is associated with only about 1.2 percentage point greater HIV prevalence. Further, this difference between cohorts is statistically significant (pval<0.01).

A similar pattern exists for primary school completion. Whereas 25-34 year olds see an approximately 0.29 percentage point higher HIV prevalence associated with a one percentage point increase in the share of adults who completed primary school, the younger cohort has a much weaker association, with a 0.11 percentage point greater HIV prevalence associated with the each one percentage point increase in the share of adults who completed primary school.

---

[3] Because all of the countries included are generalized heterosexual epidemics, the general trajectory of the disease (as modeled by UNAIDS) will be similar, but the tempo of when prevalence peaks and naturally levels off might differ based on the initial timing of the epidemic.

This weaker relationship between HIV and education for the younger age cohort is consistent with the idea that the strong association between education and HIV is beginning to change. If there is consistency in education levels within regions across cohorts, then young people in highly educated areas are potentially still exposed to more risk, in part due to the residual higher prevalence in their communities. However, the attenuation of the education-HIV association suggests that they may be also altering behaviors more at the individual level to protect themselves. Conversely, the pattern of weakening of the association depicted in Figures 1a and 1b could also be consistent with an association that simply strengthens as people age. As time goes on, and the young cohort experiences a longer period of sexual activity, those in highly educated regions may contract the diseases at higher rates and ultimately the relationship will look exactly as it does for the older cohorts. Because I have two rounds of HIV testing data, I am able to examine this possibility by looking at the progression of the association for all the cohorts over a 5-6 year period.

Figures 2a and 2b overlay the regression lines for the same three age cohorts when education and HIV prevalence is measured again in second wave. The solid lines are the same regression lines from Figures 1a and 1b. The dotted line represents model (1) repeated in wave 2 for the youngest cohort, now approximately aged 20-31. The dashed lines similarly represent the relationship for the older cohorts as they are measured in wave 2, either five or six years later.[4] These coefficients are also presented in Table 3. If the relationship between HIV and education strengthens as the cohort ages, I would expect to see the slope of the dotted line begin to rotate up –approaching the slopes observed for the older cohorts -and the OLS coefficient on the

[4] Part of the oldest age group ages out of measurement over the course of the two waves (i.e. they pass 49 yrs of age and are no longer asked the individual questionnaires), As such, the wave 2 measurement is based on only the younger members of this cohort.

education measure to increase. In fact, it does not. HIV prevalence goes up somewhat among the younger group at all education levels, but the slope of the line does not change in any substantive way. The coefficient estimate on ‘years of education’ which is 0.012 in the first wave actually decreases to 0.009 in the second wave, although the difference between the two is non-significant. The pattern for the middle cohorts is similar, with a small statistically non-significant change in the coefficient on years of education or primary school completion rate. The oldest cohort actually demonstrates the most change in the coefficient between waves. The slope of the regression line rotates considerably downwards as the group ages, and the coefficient on years of education drops from 0.032 to 0.020.

This pattern could indicate several things about the oldest age group. It is possible that less educated people in the older group are contracting HIV at a much higher rate than the more educated. But it is also possible that selective morality by education level is driving the observed change. While I test for selective mortality by education for all three age cohorts below, I am most concerned about it for this age group- who are the ones who are more likely to have contracted HIV early in the epidemic and therefore the most likely to have had the disease long enough to be experiencing high levels of AIDS mortality. Although it is not statistically significant, I do find a slight drop in mean education levels of this cohort over time. In addition, a small part of this cohort ages out of the survey between wave one and two. Because of these two factors, I treat the estimates from this cohort with caution, and the remaining analyses presented are conducted both including this age cohort and excluding it, to ensure the results are robust to their exclusion. Ultimately, the inclusion or exclusion of this cohort does not change the substantive results of the analysis, but I focus the remaining sections on the younger two cohorts, for whom I have complete information in both waves, and for whom there is no evidence of selective mortality by education.

Collectively, I interpret the differences between cohorts illustrated in this section as early evidence that the relationship between HIV and education at the regional level is beginning to change. Moreover, this change is concentrated among younger people, who both have lower rates of HIV prevalence and a weaker association between regional education rates and regional HIV prevalence. In an additional analysis, which is not shown, I repeated the above separately for men and women but did not find substantively different results by gender.

The observation that the positive relationship between education and HIV is weakening across age cohorts, although interesting in itself, does not account for the causes or pathways for change. If the observed changes are due to the deterioration of educational infrastructure in previously highly educated areas, then it is hardly good news. Conversely, if the change is reflecting behavioral reactions among the educated resulting from the dissemination of prevention knowledge, then it’s a hopeful development in the efforts to curtail the epidemic. Although their implications differ, there is no reason that these two mechanisms are mutually exclusive. Both could be happening simultaneously. The remaining sections of the paper begin to explore the evidence for the possible pathways for this change, looking first at the ‘educational erosion’ pathway and secondly at the ‘protective knowledge’ pathway.

Educational Erosion

As noted above, much recent work on HIV in sub-Saharan Africa has documented an association between high prevalence and deterioration in human capital. To my knowledge, this data offers the first opportunity to examine both the changes in HIV and education longitudinally in several nationally representative datasets. If high HIV prevalence is causing a general deterioration in educational infrastructure, we might expect to see education levels to fall over time in high prevalence areas. (Note that high prevalence areas are generally more educated to begin with, so the overall levels of both adults and children will be higher there.)

In order to assess the possibility that HIV prevalence negatively impacts provision or access to education, I compare changes in educational levels over time across regions with varying levels of prevalence. Specifically, I look at school going children ages 7-14 in both waves, and test whether the education levels of children in high HIV prevalence areas fall (or rise less) than those in lower prevalence areas. I concentrate on children because of the potential bi-directionality of the HIV/education relationship among adults and because the educational attainment of adults is largely completed and unlikely to change much over the observed time period. Once again, I use two different measures of regional education levels. The first is a measure of share of children aged 7-14 who are attending school. However, the corrosive effect of HIV might not be seen simply in attendance rates, but also in children's overall completion or progress through school. As such, I also use average years of completed education for the 7-14 age group. This could be a problematic outcome variable if there were large differences in age distribution across the samples. For example, if one country or region has many more very young children, then the average number of years of schooling completed might be much lower, but there wouldn't necessarily be reason for concern. However, the age distributions, as shown in Table 2, are almost identical across countries and waves. For both educational outcomes, I estimate the least squares regression for the model,

(2) $$\Delta Ed_{rw2-rw1} = \beta_0 + \beta_1 HIV_{rw1} + \theta X_{rw1} + \varepsilon$$

in which $\Delta Ed_{rw2-rw1}$ is the change in the regional mean for the educational outcome (either 'average years of education completed' or 'school attendance rate') for children in region *r* between wave 1 and wave 2. Independent variable HIV_{rw1} is the *adult* regional HIV prevalence for region *r* in wave 1, and X_{rw1} is the same vector of regional control variables from model (1). If high HIV prevalence is leading to a general erosion of educational opportunity, I would expect to see a drop in children's educational attainment between the waves (or perhaps smaller rises) in places with high adult HIV prevalence at baseline. Figures 3a and 3b show the regional changes in children's educational outcomes in relation to the adult HIV prevalence in wave 1. Table 4 shows the regression coefficients which correspond to the regression lines depicted in the figures. As you can see in the figures, the general educational trend is positive. Educational outcomes for children in most areas changed for the better over the period covered by the two waves. I do not find any statistically significant evidence that the positive trend in educational attainment is slowed or reversed in regions with higher HIV prevalence. Although the coefficients on the OLS regressions are negative, they are not significant either in aggregate or separated by gender.

I also find that separating the analysis by *adult* gender does not change the finding. Changes in children's educational outcomes are related similarly to the adult male HIV rate and the adult female HIV rate at baseline. Finally, I tested whether the change in adult regional HIV prevalence over time (rather than the initial level) was associated with changes in children's educational trends in the region during the same period. However, I found that an increase in adult HIV rate was not significantly associated with a simultaneous change in children's average educational outcomes for either measure.

In summary, contrary to previous work, I don't find evidence for educational erosion in high HIV areas. I interpret these results with caution, however, since the sample is limited to 91 regions. It is possible that, as more longitudinal data becomes available, there will be stronger evidence either for or against the educational erosion pathway.

Protective Knowledge

To examine the second pathway through which the relationship between education and HIV might change, I ask whether areas with higher education levels experience greater decreases in HIV prevalence. If education has become protective as a result of dissemination of knowledge, we might expect to see more highly educated people adopt protective behaviors first and, consequently, see HIV prevalence fall or rise more slowly in more highly educated areas. To examine this possibility, I look at the change in the adult HIV prevalence between the waves in relation to the average adult education levels in the region at baseline. To do this, I use least squares regression to estimate the equation,

(3) $$\Delta HIV_{rcw2-rcw1} = \beta_0 + \beta_1 Ed_{rcw1} + \theta X_{rcw1} + \varepsilon$$

where $\Delta HIV_{rcw2-rcw1}$ is the change in HIV prevalence between wave 1 and wave 2 for adults in region *r*, within age cohort c . Ed_{rcw1} is the regional mean for the educational outcome (either 'years of education' or 'primary school completion rate') for adults in age cohort *c* and region *r* at wave 1. X is the vector of regional control variables from model (1). Table 5 reports coefficient β_1 for model 3 for both education measures.. Results are presented for the population as a whole and independently for the younger and middle age cohorts. There is a statistically significant association between the education level at baseline and the trajectory of HIV prevalence for both measures of education. One additional year of mean adult education is associated with -0.7 percentage point difference in the change in regional HIV prevalence between the two waves. In other words, regions with higher adult education experience larger falls (or smaller rises) in HIV prevalence (as shown graphically in Figures 4a and 4b) and this pattern appears consistent across age cohorts and in both high and low prevalence countries.[5] The finding is also consistent for both men and women when model (3) is run separately by gender.

[5] The coefficient on regional mean adult years of education and mean primary completion for Mali (the lone low prevalence country) are not statistically significant on their own as there are only 16 regions in Mali (and therefore very low statistical power) but the pattern is the same. i.e. the coefficients are negative and only slightly smaller than those for the high prevalence countries.

While this pattern is supportive of the theory that education has begun to become protective, it is equally consistent with selective mortality by education level. If, as we know, highly educated people also experienced higher HIV rates early in the epidemic, then it is possible that there is differential mortality by education. That is, those more educated people that first contracted the disease are dying at a higher rate, bringing the prevalence down among the educated via mortality rather than via averted cases. Unfortunately, I cannot distinguish HIV mortality from other types of mortality with this data, so it is difficult to examine this directly. However, I can test for differential mortality by education by looking at whether education levels drop among the same age group over time. Since respondents don't normally lose years of education once they have accumulated them, the average education levels of a cohort should only be able to rise or stay the same as time goes on. If average education drops, then it is very likely that more educated people are dying at a higher rate. As mentioned above, we might expect to find this effect most among the older cohorts since the youngest are less likely to have begun experiencing high mortality in the relatively short time periods covered by the two waves.

For all three cohorts, I calculate the mean years of education and primary school completion at baseline, and then again in the second wave, and test to see if there has been a fall in mean education levels. I find that, for all three cohorts, there is no statistically significant change in years of completed education or primary school completion rate over the period in question. For the youngest and middle cohorts, the mean years of education and primary school completion increase slightly, although the change is not statistically significant. For the oldest cohort, the mean regional education values decrease (although, again, this decrease is not statistically significant). I take this as evidence that differential mortality by education is unlikely to be driving the observed changes in the education-HIV relationship. However, I also calculate model (3) for just the younger two cohorts, to account for the possibility that there is differential mortality by education in the oldest cohort which is not captured adequately by the test here. The findings are robust to the exclusion of the oldest group and the correlation between high adult education levels and reduction in regional HIV prevalence is statistically significant and of similar magnitude when only the younger cohorts are analyzed.

Testing the Relationships at the Individual Level

As a final test of the protective knowledge pathway, I test whether the HIV rate changes more over the period for relatively more educated individuals versus relatively less educated individuals. In order to do this, I move away from the regional level analyses performed so far and focus on the relationship between HIV status and completed education at the individual level. While using individual level data to test a pathway that hinges on individual behavior change is logical step, I am somewhat limited by the structure of the available data. Because the initial HIV surveys in Mali and Zambia were not linkable to the household survey, I cannot link individuals to their HIV results. Therefore, for those two countries, I cannot test variation in HIV prevalence as it relates to an individual’s characteristics, such as education, age or gender. I am limited to the adult samples from Tanzania and Kenya for this part of the analysis. In total, I am able to link HIV test data to complete individual data for 24,943 individuals in the youngest and middle cohorts.

I begin by testing the base premise that adults in the sample with more education actually are more knowledgeable about HIV/AIDS prevention. I use the same two education measures at the individual level (years of schooling completed and a binary measure of whether the individual completed primary school) and I find that both education measures are positively and statistically significantly associated with HIV knowledge.[6] More educated people are more likely to know that a healthy looking person can be infected with HIV and that HIV risk can be minimized by regular condom use, abstinence, and/or minimizing concurrent sexual partners. Education is also negatively associated with incorrect beliefs that HIV can be transmitted through mosquito bites or sharing food with an infected person. This association between education and HIV/AIDS knowledge is consistent across cohorts and across both waves.[7]

Given that adults with more education do demonstrate more accurate knowledge of how to prevent the transmission of HIV/AIDS, I then test whether education translates to lower prevalence among the educated. As a first step, I compare the probability of testing HIV positive at each year of educational attainment for the two cohorts in each wave. Figure 5 shows this relationship for both cohorts in the first and second wave. Several things are evident from this simple cross sectional relationship. The first is that the youngest cohort is at lower risk across all education levels in both waves. The second is that for the younger people at the lower end of the educational scale, the prevalence of HIV increases considerably more between waves. It appears that for individuals in the youngest cohort, the relationship between education and HIV may actually be reversing and becoming negative. I test this relationship more directly by splitting the

[6] As only 4% of the sample has more than 12 years of education, I top code years of education at 12 yrs, but results are robust to using 16 years of education as well.
[7] This analysis is not shown here, but is available on request. I run probit regressions of an individual's probability of correctly answering questions about HIV on the education measures. I control for individuals age, age squared, urban residence, country, region, gender and the year of measurement. All results reported are statistically significant at 5% level or higher.

sample into adults with primary education or more, and those with less than a primary education. I then run a difference of differences model by estimating,

$$(4) \quad HIV_{icw} = \beta_0 + \beta_1 Primary_{icw} + \beta_2 W2 + \beta_3 Primary_{icw} * W2 + \theta X_{icw} + \varepsilon$$

where HIV_{icw} is the probability that individual *i* from cohort *c* in wave *w* tests positive for HIV, and $Primary_{icw}$ is a binary indicator of whether that individual reported having completed primary school or more. $Wave2$ is an indicator for whether the observation is during the second wave of the survey, and X_{icw} is a vector of controls including age, gender, urban residence and region fixed effects. . The coefficient $\beta_3$represents the difference-in-difference estimator. In other words, β_3 is an estimate of how much the change in HIV prevalence between waves for people with primary education differs from the change in HIV for people with less than a primary education. I find that for the youngest cohort, β_3 is statistically significant and equal to -0.016 (pval<.05). In other words, the probability of being HIV positive for an individual with less than a primary education increases by 1.6 percentage points more than it does for someone with a primary education or more. [8] For the middle cohort, the coefficient estimate of β_3 is -0.009 and is non-significant.

Finally, I estimate the above model for each year of completed education rather than simply splitting the respondents by completion of primary school. Figures 6a and 6b show the estimates of β_3 and their confidence intervals for each year of completed education, first for the youngest cohort (6a) and then for the middle cohort (6b). While most of the estimates are non-

[8] Estimates are based on a sample of 14,323 HIV tested individuals in the youngest cohort, and 10,620 HIV tested individuals in the middle cohort. Pvalues based on bootstrapped standard errors.

significant, a pattern is detectible for the youngest cohort, with lower years of education being associated with positive changes and higher years of education being associated with no change. For the middle cohort, there is not as discernable a pattern. I interpret this as evidence that, at the individual level as well as the regional level, the relationship between education and HIV has begun to change for the youngest generation.[9] Whereas for the older generations, education was a risk factor for HIV infection, it now appears to be comparatively protective, at least at the primary school level and above.

Conclusion

The analysis above takes advantage of some of the first available nationally representative and longitudinal evidence on both HIV prevalence and educational trends. My findings support the hypotheses of previous scholars that the relationship between HIV and education has begun to reverse in sub-Saharan Africa. Although I find that, at the regional level, the relationship is still positive across age groups, it is much weaker for the youngest generation.

In my examination of potential pathways for this change, I find support at both the regional and individual level for the idea that education is becoming more protective, perhaps as knowledge about the disease disseminates. Regions with the highest levels of adult education saw the largest drops in HIV prevalence over time and younger people with a primary education or higher experienced less increase in HIV prevalence than those in the same age cohort with less education. I do not, however, find evidence of general erosion in education in high prevalence regions. Although the evidence on the pathways is not conclusive, it does support a more hopeful

[9] In an additional analysis, I performed the same differences-in-differences estimation at the regional level (comparing regional changes in HIV prevalence across waves for primary educated and non-primary educated) and results were equivalent to the individual level analysis.

interpretation of the changes in the education HIV relationship. Rather than showing evidence of general societal deterioration as a result of AIDS, the changes may reflect some success for the educational campaigns of the last two decades.

References

Behrman, J., Kohler, H.-P. & Castro, R., 2010. Perception of HIV risk and the quantity and quality of children: The case of rural Malawi. *PSC Working Paper Series*. Available at: http://repository.upenn.edu/psc_working_papers/23.

Case, A. & Paxson, C., 2009. The Impact of the AIDS Pandemic on Health Services in Africa: Evidence from Demographic and Health Surveys. *National Bureau of Economic Research Working Paper Series*, No. 15000. Available at: http://www.nber.org/papers/w15000 [Accessed April 26, 2011].

Crimmins, E.M. & Saito, Y., 2001. Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. *Social Science & Medicine*, 52(11), pp.1629-1641.

Cutler, D.M. & Lleras-Muney, A., 2006. Education and Health: Evaluating Theories and Evidence. *National Bureau of Economic Research Working Paper Series*, No. 12352. Available at: http://www.nber.org/papers/w12352 [Accessed March 31, 2011].

Feldman, J.J. et al., 1989. NATIONAL TRENDS IN EDUCATIONAL DIFFERENTIALS IN MORTALITY. *American Journal of Epidemiology*, 129(5), pp.919 -933.

Fortson, J., 2009. HIV/AIDS and fertility. *American Economic Journal: Applied Economics*, 1(3), p.170–194.

Fortson, J.G., 2011. Mortality risk and human capital investment: The impact of HIV/AIDS in sub-Saharan Africa. *The Review of Economics and Statistics*, (0).

Fortson, J., 2008. The gradient in sub-Saharan Africa: socioeconomic status and HIV/AIDS. *Demography*, 45(2), p.303–322.

Gregson, S., Waddell, H. & Chandiwana, S., 2001. School education and HIV control in sub-Saharan Africa: from discord to harmony? *Journal of International Development*, 13(4), p.467–485.

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

Mishra, V. et al., 2006. HIV Testing in National Population Based Surveys: Experience from the Demographic and Health Surveys. *Bulletin of the World Health Organization*, 84(7), pp.537-545.

Preston, S.H. & Elo, I.T., 1995. Are Educational Differentials in Adult Mortality Increasing in the United States? *Journal of Aging and Health*, 7(4), pp.476 -496.

Ross, C.E. & Wu, C.-ling, 1995. The Links Between Education and Health. *American Sociological Review*, 60(5), pp.719-745.

Singh, S. & Samara, R., 1996. Early Marriage Among Women in Developing Countries. *International Family Planning Perspectives*, 22(4), pp.148-175.

Table 1: Regional HIV and Educational Outcomes by Country and Wave

		Kenya (2003)	Kenya (2009)	Mali (2001)	Mali (2006)	Tanzania (2003)	Tanzania (2008)	Zambia (2001)	Zambia (2007)	Total (Wave 1)	Total (Wave 2)
HIV Rates	Mean	0.076	0.063	0.015	0.012	0.081	0.069	0.173	0.154	0.087	0.075
	Standard Deviation	0.056	0.038	0.014	0.008	0.062	0.056	0.075	0.058	0.076	0.066
	25th Percentile	0.041	0.033	0.004	0.007	0.041	0.031	0.102	0.117	0.035	0.021
	75th Percentile	0.114	0.079	0.021	0.017	0.107	0.097	0.224	0.199	0.117	0.116
Adult Years of Completed Education	Mean	7.28	8.14	2.15	2.25	6.08	6.08	6.90	7.24	5.75	5.97
	Standard Deviation	2.53	2.26	1.45	1.42	1.15	1.21	1.62	1.58	2.34	2.42
	25th Percentile	6.45	7.24	0.86	1.04	4.86	5.09	5.39	5.57	4.68	5.01
	75th Percentile	9.53	9.67	3.00	3.25	7.13	7.24	8.32	8.59	7.29	7.56
Percent Adults with Primary Education	Mean	0.57	0.65	0.19	0.19	0.69	0.68	0.61	0.64	0.57	0.58
	Standard Deviation	0.23	0.21	0.14	0.13	0.13	0.14	0.19	0.18	0.24	0.24
	25th Percentile	0.44	0.52	0.06	0.07	0.55	0.57	0.43	0.43	0.43	0.43
	75th Percentile	0.76	0.81	0.28	0.29	0.80	0.79	0.79	0.79	0.77	0.79
	Number of Regions	15	15	16	16	42	42	18	18	91	91

All regional means calculated based on men and women aged 15-49, HIV prevalence calculated using DHS supplied HIV weights, Educational Means Calculated using DHS sample weights

Table 2: Sample Demographics by Country and Wave

	Kenya (2003)	Kenya (2009)	Mali (2001)	Mali (2006)	Tanzania (2003)	Tanzania (2008)	Zambia (2001)	Zambia (2007)	Total
Adults Aged 15-49									
Mean Age	27.9	28.3	29.0	28.5	28.5	28.1	27.5	28.1	28.3
	9.4	*9.6*	*9.7*	*9.8*	*9.7*	*9.3*	*9.4*	*9.3*	*9.6*
Mean Yrs Education	7.45	7.85	1.80	2.16	5.77	5.73	6.32	7.08	4.99
	4.15	*4.09*	*3.59*	*3.90*	*3.42*	*3.32*	*3.62*	*3.65*	*4.45*
Percent Female	0.51	0.53	0.54	0.53	0.53	0.53	0.51	0.52	0.53
	0.50	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*
N	17,464	16,937	25,925	29,442	13,742	14,167	16,469	14,970	149,116
Percent of Total Sample	0.12	0.11	0.17	0.20	0.09	0.10	0.11	0.10	1.00
Children Aged 7-14									
Mean Age	10.5	10.4	10.3	10.2	10.3	10.4	10.3	10.4	10.3
	2.3	*2.3*	*2.3*	*2.3*	*2.3*	*2.3*	*2.3*	*2.3*	*2.3*
Percent Attending School	0.72	0.83	0.35	0.42	0.76	0.74	0.67	0.76	0.60
	0.45	*0.38*	*0.48*	*0.49*	*0.43*	*0.44*	*0.47*	*0.43*	*0.49*
Percent Female	0.48	0.50	0.51	0.51	0.50	0.49	0.50	0.51	0.50
	0.50	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*	*0.50*
N	8,313	8,608	15,645	17,087	7,369	7,656	8,951	8,675	82,304
Percent of Total Sample	0.10	0.10	0.19	0.21	0.09	0.09	0.11	0.11	1.00

Standard Deviaitons in Italics, Means Calculated using DHS sample weights

Table 3: OLS of Regional HIV Prevalence on Regional Education by Cohort and Wave

	Regional HIV Prevalence Youngest Cohort		HIV Prevalence Middle Cohort		HIV Prevalence Oldest Cohort	
	Wave 1 (15-24)	*Wave 2 (20-30)*	*Wave 1 (25-34)*	*Wave 2 (30-40)*	*Wave 1 (35-44)*	*Wave 2 (40-49)*
Completed Years Education for Youngest Cohort in Region	0.012 ***	0.009 **				
	(0.003)	(0.003)				
Completed Years Education for Middle Cohort in Region			0.025 ***	0.018 **		
			(0.004)	(0.005)		
Completed Years Educationfor Oldest Cohort in Region					0.032 ***	0.020 ***
					(0.006)	(0.007)
Percent Youngest Cohort with Primary Education	0.113 ***	0.142 ***				
	(0.035)	(0.034)				
Percent Middle Cohort with Primary Education			0.289 ***	0.259 ***		
			(0.037)	(0.052)		
Percent Oldest Cohort with Primary Education					0.349 ***	0.262 ***
					(0.056)	(0.084)
N	91	91	91	91	86	86

Based on adults aged 15-44, regional HIV prevalence calculated using DHS hiv weights, education means calculated using provided sample weights, robust standard errors are in parentheses and an additional control is included for year of measurement and duration of epidemic in country.

* p<0.05, **p<0.01, ***p<0.001

Table 4: OLS of Change in Regional Educational Outcomes of Children 7-14 on Regional HIV Prevalence of Adults

	Change in Percent Attending School in Region		
	All Children	Girls	Boys
HIV Prevalence among 15-49 Year Olds in Region in Wave 1	-0.194	-0.161	-0.185
	(0.114)	(0.161)	(0.142)
Number of Regions	91	91	91
	Change in Average Years of Completed Education in Region		
	All Children	Girls	Boys
HIV Prevalence among 15-49 Year Olds in Region in Wave 1	-0.907	-0.152	-1.550
	(0.584)	(0.817)	(1.270)
Number of Regions	91	91	91

Educational outcomes of children aged 7-14 taken from DHS household quesitonnaire and calculated using provided sample weights, Regional HIV prevalence calculated based on adults 15-49 and using HIV survey weights, robust standard errors in parentheses and an additional control for year of measurement and duration of epedemic is included but not shown; * p<0.05, **p<0.01, ***p<0.001

Table 5: OLS of Change in Regional HIV Prevalence on Regional Education Levels

	Regional Change in HIV Prevalence: *All Cohorts* (15-44 in wave1)	Regional Change in HIV Prevalence: *Youngest Cohort* (15-24 in wave1)	Regional Change in HIV Prevalence: *Middle Cohort* (25-34 in wave1)
Completed Years Education of Adults 15-44 in First Wave	-0.007 *** (0.002)		
Completed Years Education of Adults 15-24 in First Wave		-0.006 * (0.003)	
Completed Years Education of Adults 25-34 in First Wave			-0.008 * (0.003)
Percent of 15-44 yr olds with Primary Education in First Wave	-0.061 ** (0.023)		
Percent of 15-24 yr olds with Primary Education in First Wave		-0.062 * (0.031)	
Percent of 25-34 yr olds with Primary Education in First Wave			-0.018 (0.030)
N	91	91	91

Based on adults aged 15-49, regional HIV prevalence calculated using DHS provided HIV weights, education means calculated using provided sample weights, robust standard errors in parentheses and an additional control is included for year of measurement and duration of epedemic in country.

'+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

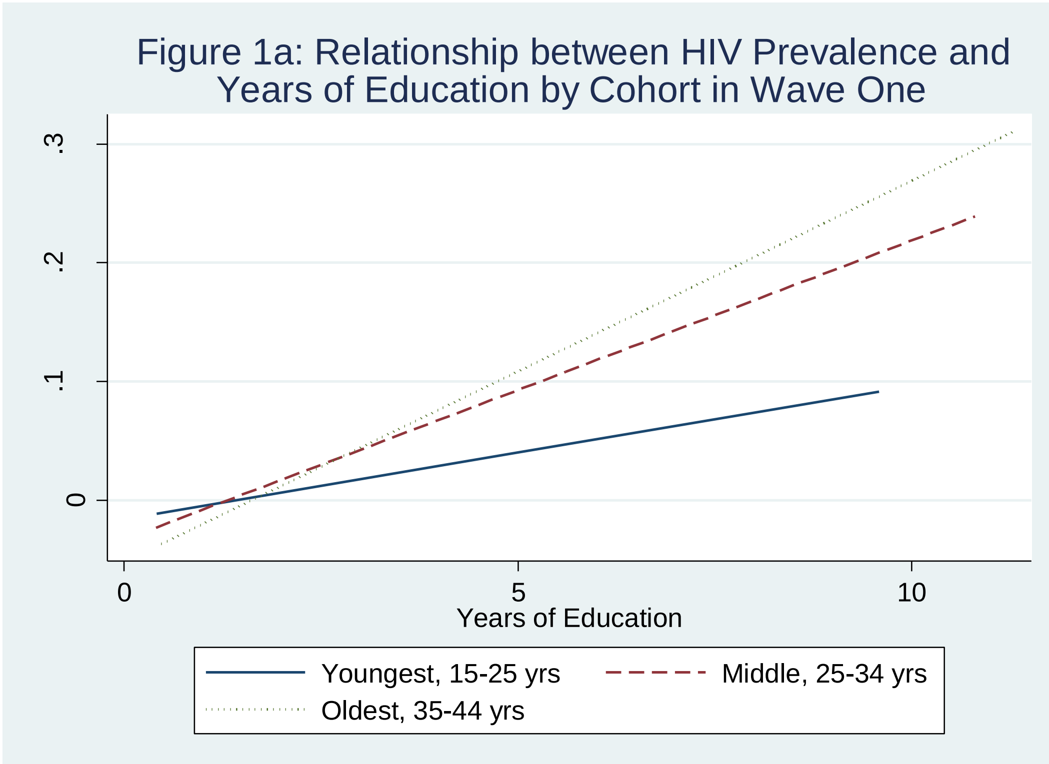


Figure 1a: Relationship between HIV Prevalence and Years of Education by Cohort in Wave One
.3
.2
.1
0
0
5
10
Years of Education
Youngest, 15-25 yrs
Middle, 25-34 yrs
Oldest, 35-44 yrs

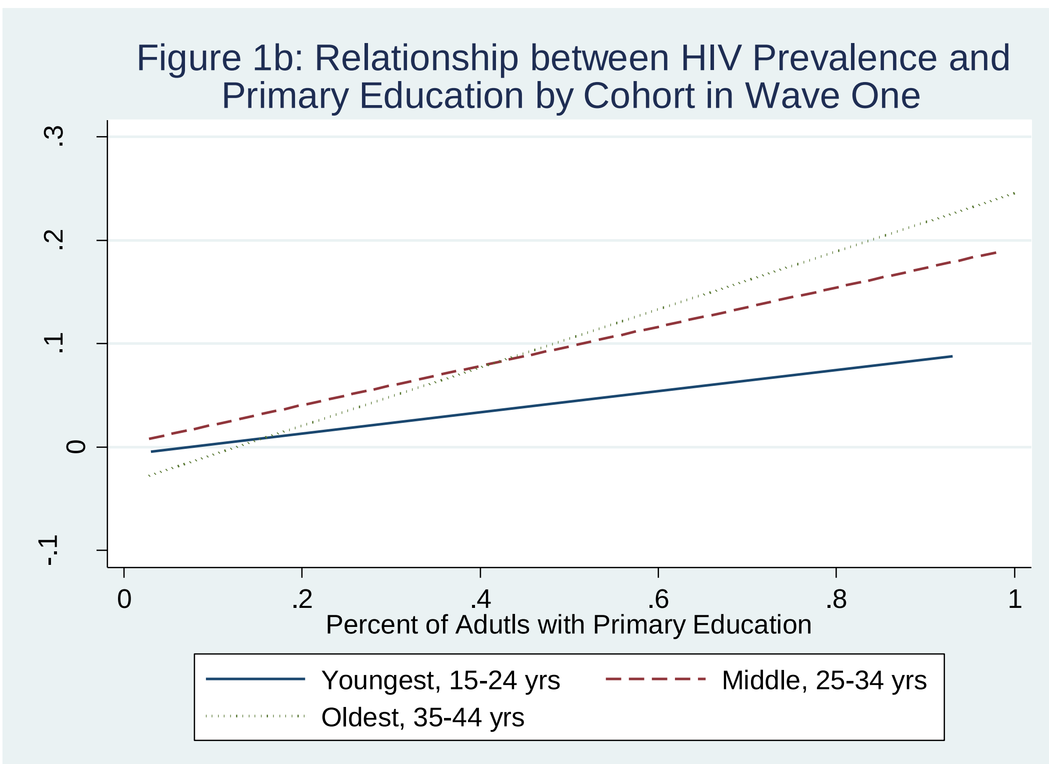


Figure 1b: Relationship between HIV Prevalence and Primary Education by Cohort in Wave One
.3
.2
.1
0
-.1
0
.2
.4
.6
.8
1
Percent of Adutls with Primary Education
Youngest, 15-24 yrs
Middle, 25-34 yrs
Oldest, 35-44 yrs

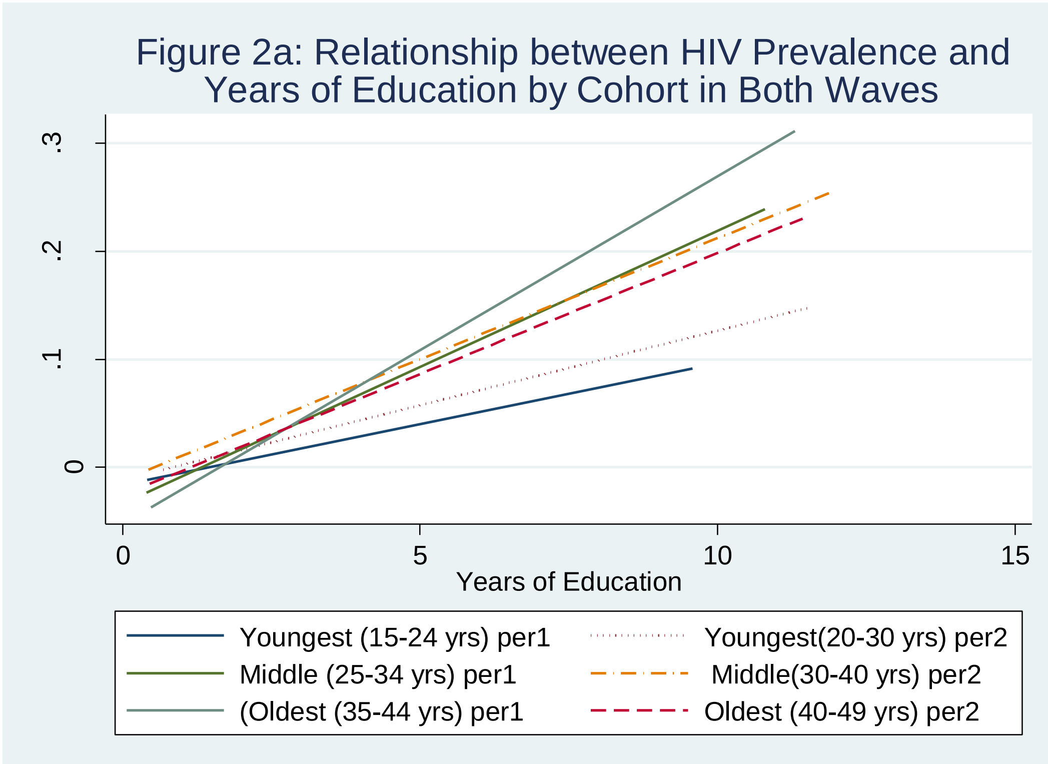


Figure 2a: Relationship between HIV Prevalence and Years of Education by Cohort in Both Waves
0
.1
.2
.3
0
5
10
15
Years of Education
Youngest (15-24 yrs) per1
Middle (25-34 yrs) per1
(Oldest (35-44 yrs) per1
Youngest(20-30 yrs) per2
Middle(30-40 yrs) per2
Oldest (40-49 yrs) per2

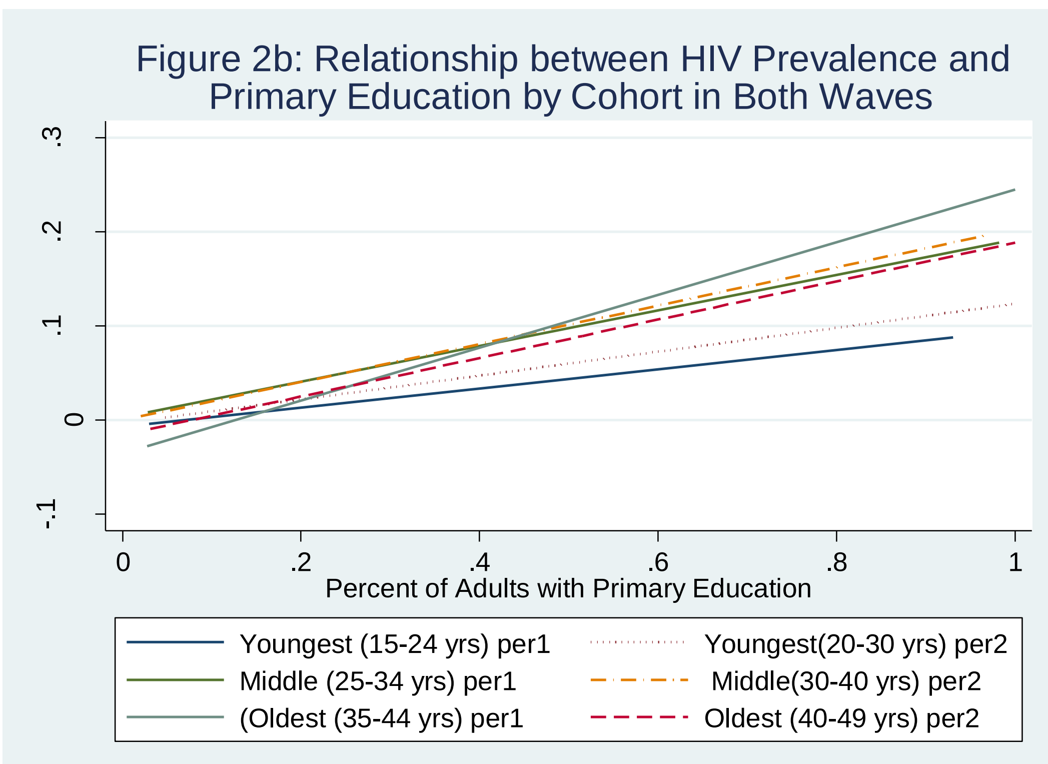


Figure 2b: Relationship between HIV Prevalence and Primary Education by Cohort in Both Waves
-.1
0
.1
.2
.3
0
.2
.4
.6
.8
1
Percent of Adults with Primary Education
Youngest (15-24 yrs) per1
Middle (25-34 yrs) per1
(Oldest (35-44 yrs) per1
Youngest(20-30 yrs) per2
Middle(30-40 yrs) per2
Oldest (40-49 yrs) per2

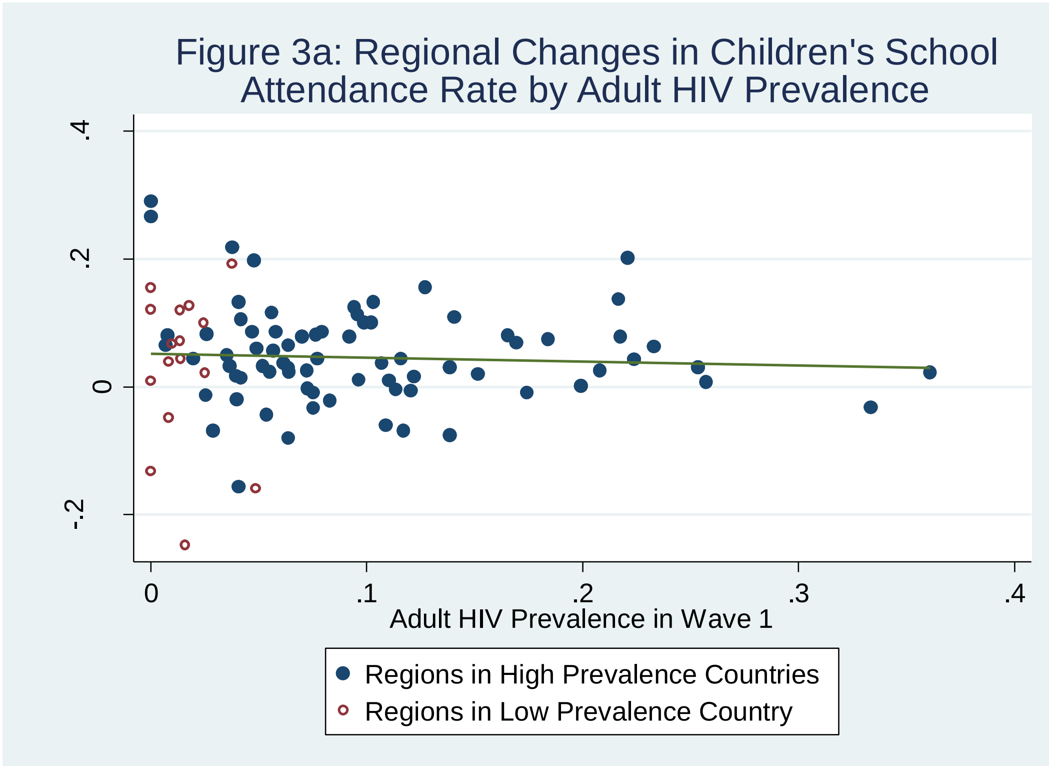


Figure 3a: Regional Changes in Children's School Attendance Rate by Adult HIV Prevalence
-.2
0
.2
.4
0
.1
.2
.3
.4
Adult HIV Prevalence in Wave 1
Regions in High Prevalence Countries
Regions in Low Prevalence Country

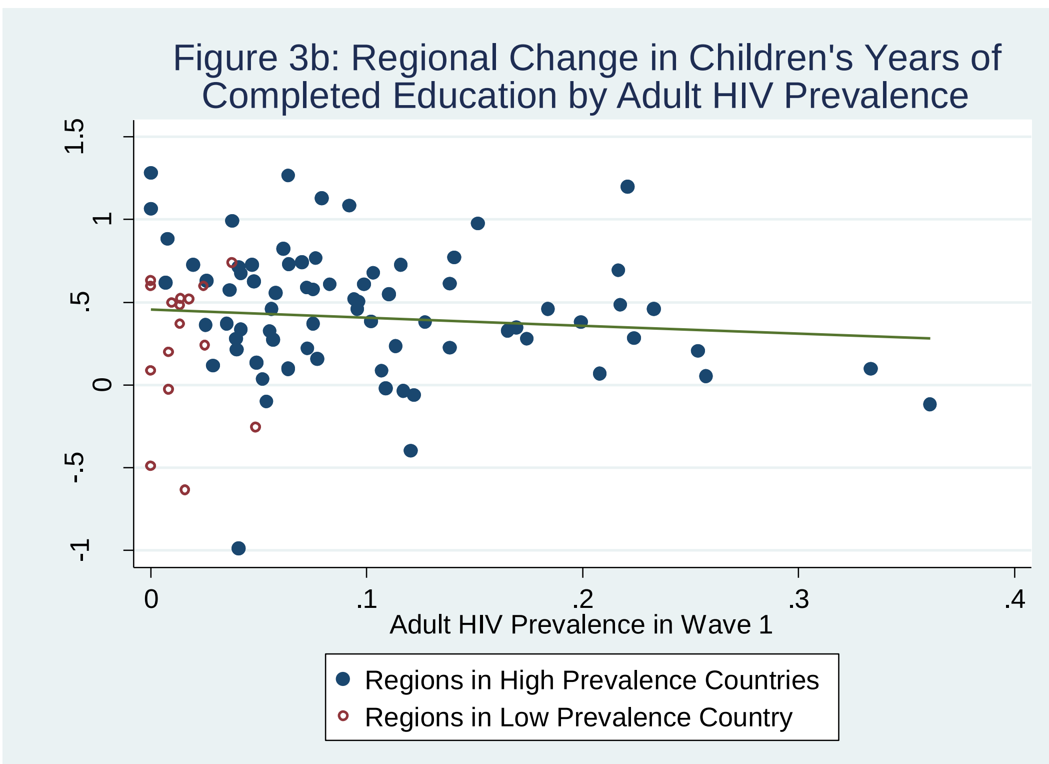


Figure 3b: Regional Change in Children's Years of Completed Education by Adult HIV Prevalence
-1
-.5
0
.5
1
1.5
0
.1
.2
.3
.4
Adult HIV Prevalence in Wave 1
Regions in High Prevalence Countries
Regions in Low Prevalence Country

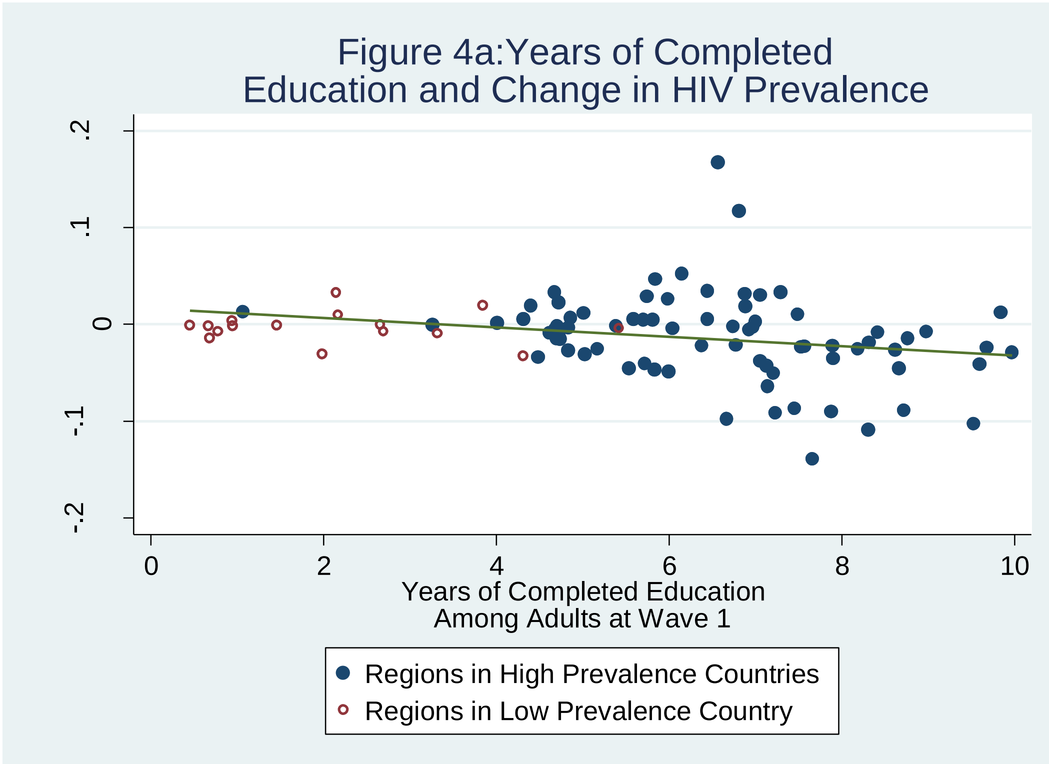


Figure 4a:Years of Completed
Education and Change in HIV Prevalence
-.2
-.1
0
.1
.2
0
2
4
6
8
10
Years of Completed Education
Among Adults at Wave 1
Regions in High Prevalence Countries
Regions in Low Prevalence Country

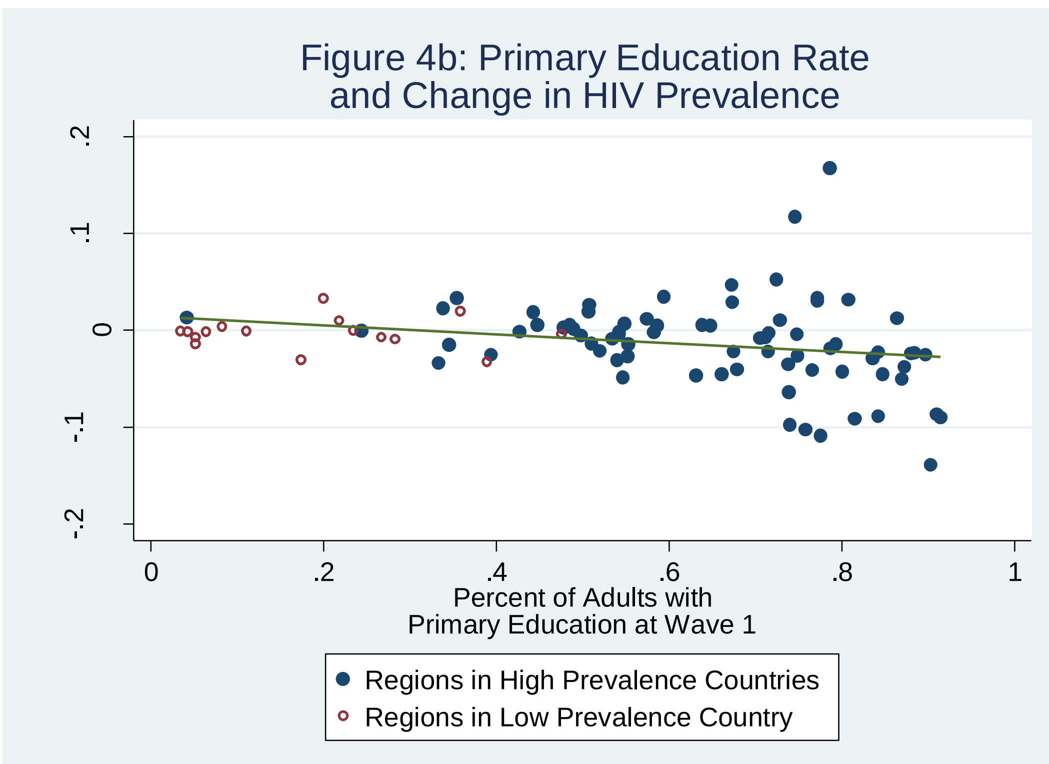
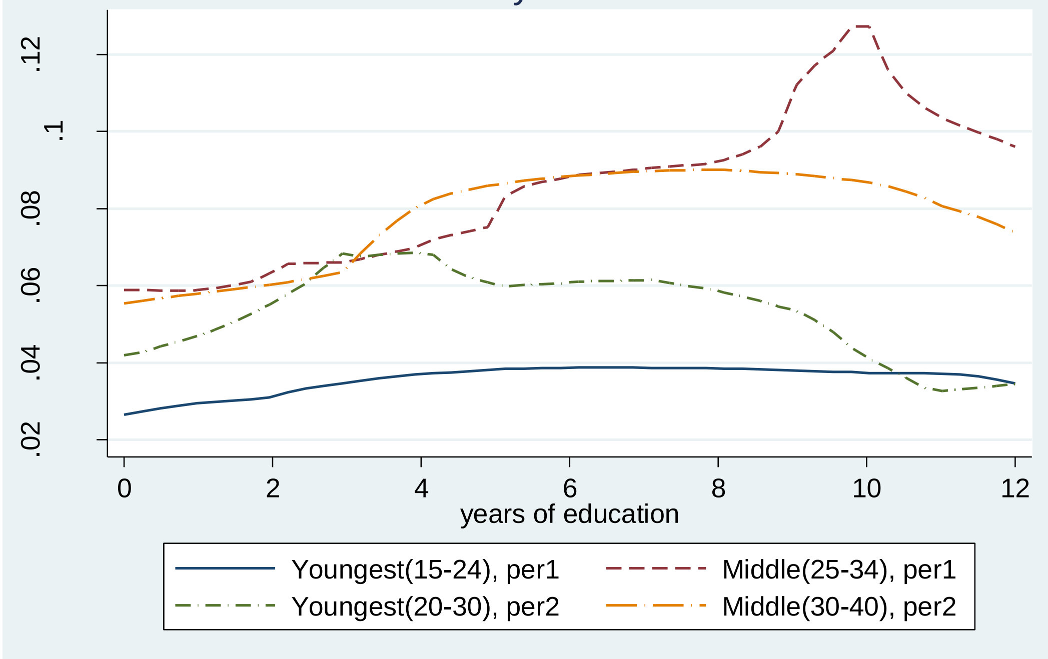


Figure 4b: Primary Education Rate
and Change in HIV Prevalence
-.2
-.1
0
.1
.2
0
.2
.4
.6
.8
1
Percent of Adults with
Primary Education at Wave 1
Regions in High Prevalence Countries
Regions in Low Prevalence Country

Figure 5: Education and HIV Risk
By Cohort



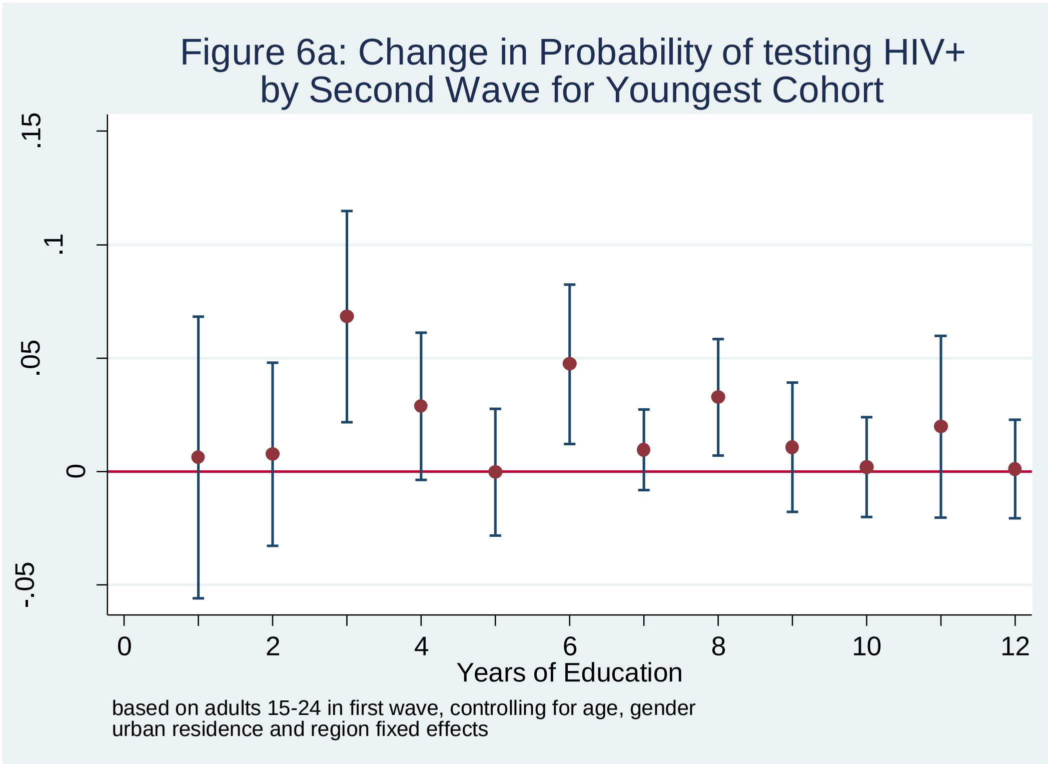


Figure 6a: Change in Probability of testing HIV+
by Second Wave for Youngest Cohort
-.05
0
.05
.1
.15
0
2
4
6
8
10
12
Years of Education
based on adults 15-24 in first wave, controlling for age, gender
urban residence and region fixed effects

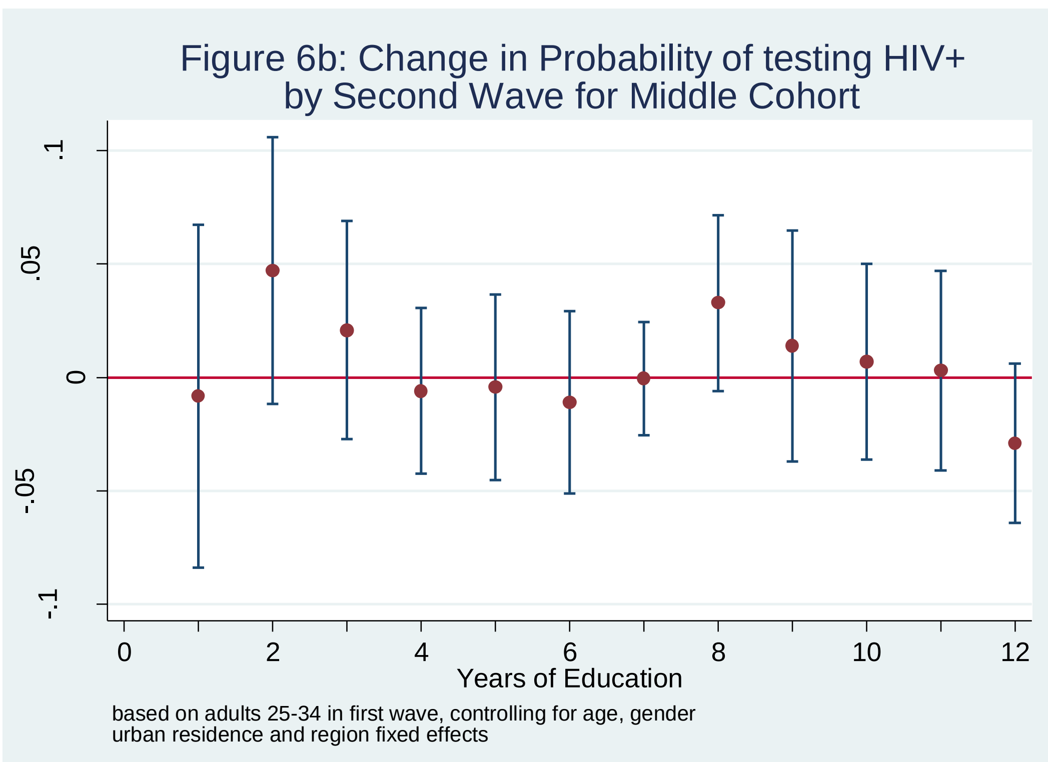


Figure 6b: Change in Probability of testing HIV+
by Second Wave for Middle Cohort
-.1
-.05
0
.05
.1
0
2
4
6
8
10
12
Years of Education
based on adults 25-34 in first wave, controlling for age, gender
urban residence and region fixed effects