The Changing Role of Education in the AIDS Crisis

Anne Case
Christina Paxson

Center for Health and Wellbeing
Princeton University

August 2010

We thank Kimberly Bryan and Thu Vu for expert data assistance, and Emily Oster, participants in the University of Chicago Demography seminar, and participants at the 2009 AIID conference in Amsterdam for useful comments. We acknowledge support from The Demography of Aging Center at Princeton University, funded under the National Institute of Aging grant P30 AG024361.

Abstract

We use multiple rounds of Demographic and Health Surveys conducted in sub-Saharan Africa to examine the relationship between education, sexual activity and HIV prevalence, and to investigate whether, over time, the role of education in the AIDS pandemic has changed. The oldest cohort of women in our study came of age before HIV had been identified, and we find that educated women in this cohort were significantly less likely to be married before age 20 and more likely to be sexually active outside of marriage as teens. Moreover, it is in regions where women began reporting markedly lower rates of teen marriage and higher rates of teen sexual activity outside of marriage, before the onset of the epidemic, that HIV prevalence rates are highest today. In high prevalence regions, we find evidence consistent with a behavioral response among younger women. The youngest women in our study respond to higher HIV rates among older women by reducing teen sexual activity outside of marriage and by increasing their rates of teen marriage. The effect of high HIV rates on teen marriage is magnified for young women with at least some formal education.

Introduction

The AIDS pandemic has taken the lives of 25 million people globally, and it is estimated that over 33 million people are currently living with HIV (UNAIDS 2010). Although the disease burden will remain high, there are beginning to be promising signs of change in the epidemic. Ahead of the XVIII International AIDS Conference in Vienna (July 2010), UNAIDS reported that HIV prevalence rates among the youngest cohorts of adult women were falling in many of the hardest hit countries in sub-Saharan Africa, with gains due to identifiable changes in sexual behavior—age at sexual debut, number of partners, and condom use. The behavioral response appears to be larger for educated Africans, who were initially the group hardest hit by HIV.

The AIDS pandemic presents an opportunity to study the role of education over the life of an epidemic. In this paper, we examine the changing relationship between education, marital and sexual behavior, and disease risk in sub-Saharan Africa. We focus on two questions: how did behaviors, in part driven by education, shape the spread of the disease before HIV was well understood? And how has the role of education changed with the arrival of information about prevention?

The growing field of economic epidemiology incorporates health behaviors into models predicting how disease spreads.[1] Individuals respond to the information they have and the incentives they face, and differences in education across people can result in heterogeneity in individual responses. Research in a number of fields documents that education is more protective against health risks, the better the risks are understood. Preston (1996), for example, notes that in 1900, the mortality risk for a school teacher's child in the US was no better than that of the average child. However, by 1925, once the germ theory of disease was well accepted, teachers'

[1] Philipson (2000) provides a review of the older literature and Toxværd (2010), a review of more recent work.

children's survival rates were dramatically better than those for other children. Feldman et al. (1989) document that in 1960 there was little difference in middle-age or old-age mortality rates for men by educational attainment. Twenty years later, once the risks associated with smoking were understood, better educated men's mortality risks were significantly and substantially lower than those of less well educated men.

After the onset of the HIV epidemic, generally dated at 1981, knowledge about HIV transmission in sub-Saharan Africa grew at uneven speed. While the HIV virus was isolated in 1984, there was uncertainty even in the scientific community as to whether HIV could be spread by heterosexual contact. One of the earliest studies of HIV/AIDS in a sub-Saharan African country (then Zaire), published in 1984, states that patients in Zaire appear to differ from those of European and American origin in that equal numbers of men and women have the disease. The study found that, in two married couples observed, both partners had the disease. On the basis of this evidence, the authors stated that there was a "strong indication" of heterosexual transmission, and recommended that additional studies be conducted, "further clarifying this new epidemiological pattern" (Piot et al., 1984). Clarification, at least in the scientific community, came quickly. Another article, published in *Science* four years later, reviewed a number of then-recent studies that led to the confident assertion that HIV-1 infection in Africa is "mainly heterosexually acquired" (Piot et al., 1988). Knowledge about HIV transmission and prevention began to spread in African communities in the late 1980s and early 1990s.

Although some researchers argued early on that education would protect women from HIV (Summers 1994), others argued that education put women's health at risk through its effect on partnership patterns. Bledsoe (1990) noted that a woman's education reduces her pool of potential marriage partners. This, together with economic opportunities that accompany education, led young women to forgo participation in polygamous marriages (and instead take on the role of "outside wife" or mistress). Singh and Samara (1996) argued that the effect of education in extending the time from menarche to marriage—due to the time cost of attaining education and the limiting effects education has on acceptable marriage partners—may put women at risk for sexually transmitted diseases, including HIV, through its anticipated impact on premarital sex.

As data became available with which to examine the associations between education and HIV risk in sub-Saharan Africa, it became clear that education was positively correlated with HIV (Gregson et al. 2001, Hargreaves and Glynn 2002). Fortson (2008), using data from several Demographic and Health Surveys (DHS) conducted in sub-Saharan Africa after 2000, found a robust positive association between education and HIV infection. In exploring mechanisms that may explain this correlation, she found that education was positively and significantly associated with premarital sex.[2] Fortson (2008), Gregson et al. (2001), and Over and Piot (1993, p. 464) all argue that, although HIV may be more common in better educated people, the pattern could change over time as those with more education alter their marital and sexual behaviors. Our focus is to document whether, and to what extent, such behavioral change has occurred.

Examining the links between marriage, sexual activity, and HIV infection is complicated by the fact that decisions about sexual activity and marriage affect HIV prevalence and, at the same time, such decisions are influenced by HIV rates. Additionally, while education may change behaviors that affect HIV risk, local HIV prevalence rates may affect an individual's or family's ability or willingness to invest in education. Our strategy for handling these identification issues relies on the fact that decisions on schooling, sexual activity and marriage that were made before

[2] Increased rates of premarital sex among educated women may have led to higher rates of sexually transmitted infections, an important risk factor for HIV (Oster 2005).

or early after the onset of the HIV epidemic could not have been influenced by HIV risk. Specifically, we use the fact that the oldest cohort of women for whom we have information—those born between 1958 and 1965—reached adulthood before anything was known about HIV. Their decisions about marriage and sexual activity during adolescence, which play a large role in their current HIV status, could not have been influenced by knowledge about HIV. In addition, we use the fact that initial school enrollment decisions for the youngest cohort of women for whom we have information—those born between 1975 and 1983—would have been made before information was available on the causes and consequences of HIV infection, to eliminate feedback from HIV rates to enrollment decisions.[3]

We use multiple rounds of DHS surveys conducted in sub-Saharan Africa to examine the relationship between education, sexual activity and HIV prevalence and to investigate whether, over time, the role of education in the AIDS pandemic has changed. For the oldest cohort of women in our study, we find that educated women were significantly and substantially less likely to be married before age 20 and more likely to be sexually active outside of marriage before age 20. Moreover, it is in regions where women began reporting markedly lower rates of teen marriage and higher rates of teen sexual activity outside of marriage before HIV had been identified that HIV prevalence rates are now the highest. In high prevalence regions today, we find evidence consistent with a behavioral response among younger women, who respond to higher HIV prevalence rates among older women in their region by reducing teen sexual activity outside of marriage and by increasing their rates of teen marriage. In addition, we find that the effect on teen marriage is magnified for young women with at least some formal education.

[3] Fortson (2010) uses recent DHS surveys to examine the impact of current HIV rates on parents' decisions to send children to school, and finds that areas with higher HIV rates have experienced larger schooling declines. Because we focus on decisions about schooling made in the 1980s, we avoid causal effects of HIV on enrollment decisions.

Data

Demographic and Health Surveys are large, nationally representative household-based surveys that collect information on population, health and nutrition. DHS surveys have been conducted in low-income and middle-income countries since 1986 and, once started in a country, are generally repeated at five year intervals. The standard DHS questionnaire collects information from women between the ages of 15 and 49, covering topics such as age at first marriage, age at sexual debut, and educational attainment. Recent surveys have tested respondents for HIV, making it possible to measure HIV prevalence within countries and, using information on the region within countries in which respondents live, regional HIV prevalence. In addition, most DHS surveys collect information on the gender, birth year, vital status and, if deceased, year of death for each of the respondents' siblings. We use information on the mortality of siblings to examine whether survivorship bias affects our results.

Our research uses data from 45 DHS surveys conducted between 1988 and 2006 in eighteen countries in sub-Saharan Africa, listed in Table 1. With two exceptions, we use data from all sub-Saharan African countries for which HIV data are currently available, either through the DHS or an Aids Indicator Survey (AIS).[4] One exception is Uganda: access to the HIV data collected in the 2004-05 AIS is restricted. The other is Rwanda, which experienced massive mortality in the 1994 genocide. It is likely that the behaviors we are studying were affected by

[4] In two cases, the HIV data are not drawn from the standard DHS, but from AIDS Indicator Surveys (AIS). All data for Cote d'Ivoire in 2005 are from an AIS. The data used to construct regional HIV prevalence for Tanzania are from the 2003/04 Tanzanian AIS. See http://www.measuredhs.com/aboutsurveys/ for a description of the surveys. Below, we do not distinguish between AIS and DHS, but refer to both as "DHS."

the genocide, and that survivorship bias would affect our measures of age at first marriage and first sex.[5]

We use the region of residence information provided by the DHS to define regions that are consistently identified over time within countries. In some countries, we must aggregate sub-regions up into larger regions in order to obtain regional definitions that are consistent across survey waves. In Ghana and Tanzania, the regional definitions changed so much that we must exclude their two earliest surveys—the 1988 survey for Ghana and the 1991-92 survey for Tanzania—from our analyses. Table 1 shows the survey waves we use and the number of regions for each country, ordered by the countries' HIV prevalence rates. The number of regions varies considerably across countries: Malawi and Cameroon have only three regions, whereas Tanzania has 20. On average, there are 2,394 women between the ages of 15 and 49 per region.

Regional HIV prevalence rates are calculated using HIV testing data for men and women ages 15-49, from the most recent DHS available.[6] Earlier research on a subset of countries we use concludes that response rates for HIV testing are high, and that bias from non-response and the exclusion of non-household population groups is small (Mishra, Barrere, Hong and Khan, 2008.) Table 1 provides statistics on HIV prevalence for each country overall, and the minimum and maximum regional prevalence. The statistics show the familiar pattern of low HIV rates in West Africa and high HIV rates in East and Southern Africa, with considerable variation across regions within countries.

---

[5] Other countries in our sample have also experienced wars and political upheaval during the time period we examine. We have checked that our results are robust to the exclusion of any single country.

[6] The surveys provide sampling weights to compute HIV prevalence.

Our research examines marriage and sexual activity prior to the age of 20, using a sample of women who were born between 1958 and 1983, and who were ages 20 and older at the time of the survey. In some of our analyses, we group women into three cohorts: those born between 1958 and 1965; those born between 1966 and 1974; and those born between 1975 and 1983. The women in the earliest-born cohort turned 20 in 1985 or earlier, when the epidemic was young and (as noted above) before it was widely known, even in the scientific community, that HIV could be spread through heterosexual activity. The middle cohort turned 20 between 1986 and 1994. These women came of age in a period when AIDS knowledge was still evolving. They may or may not have had knowledge of AIDS when making decisions as teenagers about sexual activity and marriage.

Members of the youngest cohort turned 6 between 1981 and 1989. It is unlikely that parental decisions about whether to enroll these girls for at least one year of formal schooling were influenced by HIV. However, as members of this cohort moved into adolescence, knowledge of AIDS was becoming increasingly widespread. DHS evidence we present below shows substantial variation across groups of women in knowledge of HIV transmission and prevention, and we do not assume that all women in the latest-born cohort had full knowledge of how to avoid infection. Below, we examine whether women with any formal schooling in high-AIDS regions are more likely than women with no formal schooling to change their patterns of marriage and sexual activity in response to the arrival of information on AIDS prevention.

Descriptive statistics for the key variables in our analyses are in Table 2. Our main outcome variables are measures of teen marriage and sexual activity. “Married by 20” is an indicator variable that equals 1 if the respondent indicates that she had married or lived with a partner before the age of 20.[7] "Sexually active and unmarried by age 20" is an indicator variable that equals 1 if the respondent does not report having been married before age 20, but reports age at first sex is less than 20 years. "Married by 20" and "sexually active and unmarried by 20" are mutually exclusive. The remaining category is "never married and a virgin at age 20." Table 2 indicates that teen marriage rates have fallen over time across cohorts and teen non-marital sexual activity has risen. Over these periods, there has been a slight decrease in the fraction of women who are unmarried and not sexually active by age 20.

The table also contains descriptive statistics from all of our DHS surveys for age, urban status and education. We use two education measures: an indicator for whether the respondent has had "any education," defined as at least one year of formal schooling, and an indicator that the respondent is literate, defined as being able to read (easily or with difficulty). Although the DHS surveys also provide information on completed years of education, we do not use this variable in our analysis because, for the youngest cohort in our sample, completed years of education could have been affected by the HIV epidemic. Both of our measures of education—exposure to at least one year of schooling and literacy—increase across cohorts. The correlation between reporting any education and reporting an ability to read is very high (0.79), and results presented below are qualitatively the same if we use literacy or an indicator for any education.

For women in our oldest cohort—who as teenagers were making decisions on sexual activity and marriage before information on HIV transmission and prevention was available—those who are literate were 14.4 percentage points less likely to be married by age 20, on average, and 7.3 percentage points more likely to be sexually active and unmarried by age 20. (These results are from regressions of each behavior on an indicator for literacy, with a complete

[7] Throughout "marriage" refers to legal marriage or cohabitation.

set of controls for year of birth, region, and urban status. Similar results obtain if we regress these behaviors on an indicator of having any education. Results were estimated, but are not reported in Table 2.)

HIV status is explored in Table 3, where a respondent's HIV status is regressed on indicators that she reported teen marriage, teen non-marital sexual activity, and literacy. Results show that all three are significantly associated with HIV infection, with teen non-marital sexual activity more strongly associated with infection than teen marriage. Specifically, controlling for current age using a complete set of age indicators, we find that women who were married by age 20 are 1.3 percentage points more likely to be HIV positive on average than are women of the same age who were virgins at age 20 (regardless of their current marital status). Women who were sexually active and unmarried by age 20 are 7.9 percentage points more likely to be HIV positive than women of the same age who were virgins at age 20. Results in column two are from a regression that includes a complete set of country indicator variables (country fixed effects), and in column three, a complete set of country-region indicators.

With the inclusion of region fixed effects, the coefficients answer the question of whether, *within the same region*, women who were married by age 20 or were unmarried and sexually active by age 20 have higher risk of HIV than do women who were virgins at age 20. That the estimated associations between HIV, teen marriage and teen sexual activity change little when country or country-region fixed effects are included suggests that the within-region associations and between-region associations are very similar. Literacy is positively and significantly related to HIV. However, the coefficient on literacy declines 10-fold when country fixed effects are included, indicating that countries with high HIV prevalence tend to have better educational outcomes for women, and those with low prevalence have poorer educational outcomes.

The associations between current relationship status and HIV infection are shown in the last column of Table 3. Although interesting, these do not provide information on whether relationship status influences the risk of HIV infection, or whether HIV infection influences relationship status. The associations indicate that the highest rates of HIV are among widowed, divorced and separated women, followed by unmarried sexually active women, followed by married women. (The omitted category here is women who have never had sex.) The high infection rate among widows is likely due to HIV infection of a spouse resulting in widowhood and own HIV infection. Similarly, HIV infection of one partner in a union may lead to divorce or separation.

It is because we cannot use results like those presented in Table 3 to discuss the impact of HIV prevalence rates on behaviors that we identify an 'oldest' cohort, whose teen behaviors were set prior to knowledge being available on HIV, and a 'youngest' cohort, some of whom would have been armed with information. We examine this in the next section.

Results

Marriage, sexual activity and education

The patterns of marriage and sexual activity in place just prior to the introduction of HIV in sub-Saharan Africa may have played a critical role in determining where HIV would spread most rapidly. Higher rates of non-marital sexual activity among teens in some regions, during the period when the dangers of HIV were unknown, are likely to have made those regions more vulnerable to infection. In this section, we document the associations between teen marriage, teen sexual activity, and current HIV prevalence at the country level and, within countries, at the regional level. We trace patterns in these variables from the period before to the period after information on HIV transmission was available. Because women's education is a primary driver of age at marriage, we also show how trends in girls' schooling and literacy among older and younger cohorts differ across countries and regions with higher and lower current HIV rates.

Table 4 presents the associations between current country-level HIV prevalence and teen marriage, sexual activity and educational attainment for the cohort of women born between 1958 and 1965. The first column repeats the information on HIV prevalence for each country shown in Table 1. The second column shows the fraction of women in the 1958-1965 cohort who were married by age 20. In general, teen marriage was less prevalent in countries with currently high HIV rates: the Pearson correlation coefficient between the country-level HIV rate and the teen marriage rate is –0.581. This correlation is driven in part by the exceptionally high teen marriage rate of Niger (93.1 percent), which has the second-lowest HIV rate, and the exceptionally low teen marriage rate of Swaziland (36.8 percent), which has the highest HIV rate. However, the Spearman rank correlation coefficient, which is robust to outliers, is also high (–0.437). The reverse pattern is seen for non-marital sexual activity among teens: countries with currently high HIV rates had greater rates of non-marital sexual activity among teens making decisions in the pre-HIV period.

Countries that had high rates of teen marriage and low rates of teen non-marital sexual activity, in the 1958-1965 cohort, had larger fractions of women who never attended school and who were illiterate. The correlation between education among girls born in this period and the rate of HIV today is striking. Among the six highest-HIV countries, all but Malawi had female literacy rates that exceeded 65 percent, whereas among the ten lowest-HIV countries, all had female literacy rates below 50 percent.

The last two rows of Table 4 show similar patterns for the correlation coefficients between the country-region level HIV rate and the country-region rates of teen marriage, teen sexual activity, and education: in country-regions with currently higher HIV rates, women in the older cohort were less likely to be married as teenagers, more likely to engage in non-marital teen sexual activity, and more likely to have attended school and become literate.

We next consider how the relationships between HIV rates and teen marriage, teen sexual activity and education at the country-region level have changed over time. We do this by estimating simple non-parametric regressions of the following form:

$$(1) \quad y_{irt} = \theta + \{\sum_{r=1}^{R} \sum_{t=1958}^{1983} \delta_r + \gamma_t + \beta_t H_r\} + \alpha X_{irt} + \varepsilon_{irt}.$$

In (1), y_{irt} is an outcome (such as teen marriage, teen sexual activity, or education) for woman i living in country-region r, who was born in year t, where t ranges from 1958 to 1983. The equation includes a set of region fixed effects (δ_r), which capture time invariant features of the region, such as economic or cultural factors that influence teen behavior or the provision of education to girls. We also include a set of birth year effects (γ_t), which capture changes over time that are common to all regions, and a set of controls for characteristics of the woman (X_{irt}), which are her age in years at the time of the survey and an indicator of whether she lives in an urban area. The variable H_r measures HIV prevalence in region r in the most recent survey year. Although H_r is fixed over time within regions, the coefficients β_t are permitted to vary over time. Because an intercept is included, identification requires that one year effect, one region effect, and one of the values of β_t be normalized to zero.

The main parameters of interest are the coefficients β_t, which measure how the time effects vary across regions that have higher and lower HIV prevalence rates. Ignoring the control variables X_{irt}, the difference in the predicted value of the outcome in region r across two time periods, 1 and 2, will be:

$$(2) \quad \hat{y}_{r2} - \hat{y}_{r1} = (\hat{\gamma}_2 - \hat{\gamma}_1) + (\hat{\beta}_2 - \hat{\beta}_1)H_r.$$

Equation (2) indicates that the predicted year-to-year change in the outcome contains two components. The first term, $(\hat{\gamma}_2 - \hat{\gamma}_1)$, is common to all regions. The second, $(\hat{\beta}_2 - \hat{\beta}_1)H_r$, shows the additional change that is scaled by the region's HIV prevalence rate in the most recent year.

The regression results are displayed graphically in Figures 1 to 4. Each figure shows predicted values of the outcome for each of the birth years, at four values for the HIV prevalence H_r: 0 percent, 5 percent, 10 percent and 20 percent. To compute predicted values, we set the variables in X_{irt} to their sample means. We also normalize all of the predicted values by subtracting the predicted values for the cohort born in 1958. This normalization makes it easier to see how the outcomes vary over time across higher and lower HIV regions. However, it is important to bear in mind that, because we have subtracted off the predicted values for the 1958 cohort for each prevalence class, the graphs tell us *relative to the 1958 cohort* in this prevalence class, what we would predict for each birth year, for each outcome, at different prevalence rates. Vertical dashed lines distinguish the three birth cohorts.

Several features of the results in Figure 1, for teen marriage, are noteworthy. The first is that teen marriage rates have declined in both higher and lower HIV regions. At an HIV prevalence rate of zero, the teen marriage rate among women who were born in 1983 is predicted to be approximately 10 percentage points lower than the rate among women who were born in

1958. In addition, the largest predicted declines in teen marriage in the higher-HIV regions occurred *before* the HIV epidemic began: in the highest HIV regions, the predicted fraction of women who were married by age 20 declines a full 10 percentage points from the 1958 birth cohort to the 1965 birth cohort. A decline of this size took 25 years to accomplish in those regions that currently have the lowest HIV prevalence. The results of Table 4 indicate that teen marriage was less common among the oldest cohort in higher-HIV countries (and country-regions) than in lower-HIV countries (and country-regions). The additional evidence in Figure 1 shows that this difference was becoming more pronounced over time *in advance of* the HIV crisis.

The results for teen non-marital sexual activity, shown in Figure 2, are consistent with the evidence for teen marriage. In currently low-HIV regions, non-marital sexual activity was flat or trending up very slightly in advance of the epidemic. In contrast, in regions with currently high-HIV rates, non-marital teen sexual activity increased most sharply among women born between 1958 and 1965, who would have reached age 20 before HIV transmission was understood. Predicted values of teen non-marital sexual activity either flatten (in high HIV countries) or decline (in low-HIV countries) among those in the latest-born cohort, who entered adolescence in the late 1980s. Although this pattern is consistent with a casual effect that runs from HIV to non-marital sexual activity, it may also be that regions in which women forgo non-marital sexual activity are those that have become lower-HIV regions. Below, we use the timing of women's decisions to identify whether regional HIV rates influenced the behaviors of those in the youngest cohort.

Figures 3 and 4 present regional patterns in our measures of educational attainment. Both figures show that educational attainment has increased over time across cohorts, with increases somewhat larger in high-HIV regions. Furthermore, as shown in Table 3, higher-HIV regions started with higher levels of educational attainment among the oldest cohorts. Results in Figures 3 and 4 indicate that this educational advantage was maintained or slightly increased over time across cohorts.

Mortality-driven selection

One concern with the results presented above is that they do not account for women who died of AIDS or other causes. This could bias our results. As knowledge of AIDS became more widespread, the gap in mortality between women with high and low education levels may have widened in high-HIV regions relative to low-HIV regions. Deaths of less-educated women would increase average education levels in the (living) sample, more so in high-HIV than in low-HIV regions—consistent with results in Figures 3 and 4. Given the high mortality produced by the AIDS epidemic, it is important to examine whether adjustments for mortality-driven selection are capable of producing large changes in our results.

In most DHS survey waves, respondents were asked the sex, year of birth and, if deceased, year of death of each of their siblings. We use this to construct a data file of all deceased female siblings who were born between 1958 and 1983 and would have been between the ages of 20 and 49 had they survived. These selection criteria ensure that these women, had they not died, would have been eligible to be in our analysis sample. Because we are looking at marriage and sexual activity by age 20, we select sisters who survived to that age. We do not have information on the marital status, sexual activity or educational status of the deceased siblings, so we cannot simply re-estimate our models including this group. Instead, we re-estimate the models under different scenarios: first, that the deceased women were all married by age 20, or were all sexually active and unmarried by age 20, or were all literate, or all had some formal education; second, that the deceased women all had values of 0 for these dependent variables; and third, that they all had the same values of the dependent variables as the living sisters who reported their deaths. When re-estimating, deceased siblings are given a sample weight equal to the inverse of one plus the number of living female siblings ages 20 to 49 that each respondent has. This prevents deceased women from large sibships from being over-weighted. These corrections do not account for deceased women who have no living siblings eligible to be respondents in the DHS, so we are still undercounting deceased women. The results of these robustness checks, available in (what will be) a web appendix (attached here for referees) indicate that these mortality adjustments do not overturn our findings.

Overall, our results indicate that regions that had more educated women, and (perhaps consequently) lower teen marriage rates, and higher rates of teen non-marital sexual activity prior to the arrival of HIV, were more likely to have high HIV prevalence in the early 21st century. With the data that are available, it is not possible to pinpoint the circumstances that led some countries and regions within countries to place more emphasis on girls' education and, perhaps as a consequence, to become more tolerant of later marriage and non-marital sexual activity. Whatever the impetus, these changes appear to have made some regions more susceptible to the spread of HIV.

Education and behavior change

Although members of the oldest cohort in our sample could not have known how their decisions as teenagers would affect their risk of HIV infection, this was not the case for members of the youngest cohort. Younger women living in high prevalence areas would have seen older women and men become sick and die of AIDS. Increasingly widespread knowledge about the causes of and prevention methods for AIDS may have led them to adopt behaviors that reduced their risk of HIV infection. In addition, if knowledge about AIDS spread the most quickly among educated women, one would expect that educated women would have the largest behavioral responses.

To examine these issues, we estimate regressions that relate the changes in teen marriage and sexual activity between the middle and youngest cohorts to the HIV prevalence of the oldest cohort. Specifically, we estimate models of the following form:

$$(3)\ y_{ir3} = \delta_c + \beta_1 \overline{y_{ir2}} + \beta_2 \overline{H_{ir1}} + \varepsilon_{ir3}$$

where y_{ir3} is an outcome for woman i living in country-region r who is a member of the third cohort of women—those born between 1975 to 1983; $\overline{y_{ir2}}$ is the average value of the outcome for women in region r who are members of the second cohort, born between 1966 and 1974; and $\overline{H_{ir1}}$ is the HIV prevalence rate among women in region r who are members of the oldest cohort, that is, those born between 1958 and 1965.[8] The intercept δ_c is a country-specific intercept. If we restrict β_1 to equal 1, which we do in some specifications, the parameter β_2 indicates how HIV rates among members of the oldest cohort are related to changes in the outcome of interest between the middle and youngest cohorts. Our hypothesis is that high HIV rates among the members of the oldest cohort produce declines in risky behavior—that is, larger increases (or, smaller declines) in teen marriage rates, and reductions (or, smaller increases) in non-marital sexual activity.

[8] We also constructed a crude measure of adult mortality among women from the earliest-born cohort, using the sibling mortality data, and use it in place of the HIV prevalence measure. The results are qualitatively similar to what we find using the HIV prevalence measure.

We also estimate variants of (3) that include a measure of education and an interaction between the individual's education (E_{ir3}) and the HIV rate for the oldest cohort:

$$(4)\ y_{ir3} = \delta_c + \beta_1 \overline{y_{ir2}} + \beta_2 \overline{H_{ir1}} + \beta_3 E_{ir3} + \beta_4 (E_{ir3} \times \overline{H_{ir1}}) + \varepsilon_{ir3}.$$

If education increases the behavioral response to HIV among the youngest cohort, we expect the coefficient β_4 to be positive if y_{ir3} is protective against HIV (e.g. teen marriage) and negative if it increases HIV risk (e.g. teen non-marital sexual activity.)

The estimates of (3) are shown in Table 5, and support our hypothesis. Controlling for the average teen marriage rate of members of the middle cohort, women in the youngest cohort are more likely to get married as teenagers, the higher is the HIV rate among members of the oldest cohort. Similarly, high HIV rates among the oldest cohort predict reductions in non-marital sexual activity. The effects, although statistically significant, are modest in size. The results imply that an increase in the HIV rate of the oldest cohort from 0 to 10 percent is associated with a 2.7 percentage point increase in the probability a women in the youngest cohort is married by age 20, and a 1.2 percentage point decline in the probability that she is sexually active and unmarried by age 20.

We find evidence that education increases the responsiveness of teen marriage to HIV. The first 4 columns in Table 6 indicate that women who are literate or have formal schooling are less likely to be married by age 20. However, the interaction between the HIV rate of the oldest cohort and education is positive, indicating that the negative association between education and teen marriage is smaller when HIV rates are higher. For women who are not literate, an increase in HIV among the oldest cohort from 0 to 10 percent increases the probability of teen marriage by 1.1 percentage points. However, for women who are literate, the same increase in HIV increases the probability of teen marriage by nearly 4 percentage points (1.13+2.84). Similar results are obtained when we use the indicator of “any education” rather than the indicator of literacy.

The hypothesis that education increases the response of teen non-marital sexual activity to HIV is not supported. Higher HIV prevalence of the oldest cohort is negatively related to non-marital sexual activity. However, the coefficients on the interactions between the HIV rate of the oldest cohort and our two education measures are not statistically different from zero. It may be that educated women in high HIV regions adopted other risk prevention strategies as teens, such as having less frequent sex, having fewer sexual partners, or using condoms.

AIDS knowledge

We can assess how knowledge about AIDS and strategies to prevent it vary by education. Table 7 shows regression results of indicators of AIDS knowledge on measures of education, with controls for age, urban status, birth year, and country-region indicators. Whether education is measured as years of schooling, reporting any education, or being literate, women with more education are more likely to know about AIDS and AIDS prevention. Women who are literate are 4.7 percentage points more likely to have heard of AIDS and 11.6 percentage points more likely to believe it can be avoided. They are more likely than illiterate women to believe that using condoms, being faithful and practicing abstinence can prevent AIDS. Further, among those who believe AIDS can be avoided, literate women are 9.4 percentage points less likely than illiterate women to think AIDS can be spread by mosquito bites, and 9.9 percentage points less likely to think it can be spread by sharing food.

Our findings are consistent with younger women—both educated and uneducated—changing their behaviors in high HIV prevalence areas. With respect to choosing marriage by age 20, there is an additional effect of being educated in a high HIV area: we find a modest but significant additional increase in the probability that educated women in the youngest cohort are married before age 20, the higher is the HIV prevalence among the oldest women in their region.

Conclusions

Our results, based on DHS data, provide support for the 2010 UNAIDS findings that, in areas hardest hit by AIDS, behavioral responses of younger adults may be beginning to provide some protection against HIV. Our results suggest modest increases in teen marriage, and reductions in teen non-marital sexual activity, in response to high HIV prevalence. The responsiveness of teen marriage is larger among women who have at least some formal schooling. These results, coupled with evidence that women without education are less likely to know that AIDS can be prevented and more likely to have misinformation about prevention strategies, suggests that it may be a priority to tailor prevention campaigns to women with low levels of education.

Our results also suggest that the association between education and HIV has changed over time as knowledge about AIDS has spread. Change in the role of education over the course of the AIDS epidemic mimics the progression observed for many diseases: better educated people may be at equal or even higher risk before mechanisms for transmission are well understood but, over time, they may be less likely to succumb to diseases that they can protect against. In this way, the education-health gradient may rotate from one in which education is positively related to disease, to one in which education is protective against it.

References

Bledsoe, C. 1990. "Transformations in Sub-Saharan African Marriage and Fertility," *Annals of the American Academy of Political and Social Science* 510: 115-125.

Feldman, J.J., D.M. Makuc, J.C. Kleinman, and J. Cornoni-Huntley. 1989. "National Trends in Educational Differences in Mortality." *American Journal of Epidemiology* 129:919-1033.

Fortson, J. G. 2008. "The Gradient in sub-Saharan Africa: Socioeconomic Status and HIV/AIDS," *Demography* 45(2): 303-22.

Fortson, J.G. 2010. "Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa." Forthcoming Review of Economics and Statistics. Posted online on 9 July 2010: http://www.mitpressjournals.org .

Mishra, V. B. Barrere, R. Hong, and S. Khan, S. 2008. "Evaluation of Bias in HIV Seroprevalence Estimates From National Household Surveys," *Sexually Transmitted Infections* 84: i63-i70.

Oster, Emily. 2005. "Sexually Transmitted Infections, Sexual Behavior, and the HIV/AIDS Epidemic," *Quarterly Journal of Economics* 120(2): 467-515.

Over, M. and P. Piot. 1993. "HIV Infection and Sexually Transmitted Diseases." Pp. 455-527 in *Disease Control Priorities in Developing Countries*, edited by D.T. Jamison, W.H. Mosley, A.R. Measham, and J.L. Bobadilla. New York: Oxford University Press.

Philipson, T .J. 2000. "Economic Epidemiology and Infectious Disease" in A. Culyer and J. Newhouse (eds.) *Handbook of Health Economics* 1B, Part 8, Elsevier: North Holland.

Piot, P. et al. 1984. "Acquired Immunodeficiency Syndrome in a Heterosexual Population in Zaire," *The Lancet* 324(8394): 65-69.

Piot, P. et al. 1988. "AIDS : An International Perspective," *Science* 239(4840): 573-579.

Preston, S. H. 1996. "American Longevity: Past, Present, and Future," Policy Brief No. 7, Syracuse University Maxwell School of Citizenship and Public Affairs/Center for Policy Research.

Singh, S. and R. Samara. 1996. "Early Marriage Among Women in Developing Countries." *International Family Planning Perspectives* 22(4): 148-57.

Summers, L. H. 1994. "Investing in All the People: Educating Women in Developing Countries." EDI Seminar Paper No. 45, The World Bank.

Toxværd, F. 2010. "Recurrent Infection and Externalities in Prevention." Mimeo, University of Cambridge.

Table 1: Descriptive Statistics for countries

Country	Survey years	# regions	Sample HIV rate (%)	Lowest regional HIV rate (%)	Highest regional HIV rate (%)
Senegal	1992-93, 1997, 2005	4	0.70	0.49	2.06
Niger	1992, 1998, 2006	6	0.73	0.34	1.30
Dem. Rep. of Congo	2007	11	1.30	0.20	3.72
Mali	1995-96, 2001, 2006	9	1.33	0.63	2.09
Ethiopia	2000, 2005	11	1.47	0.31	5.84
Guinea	1999, 2005	5	1.55	0.94	2.40
Liberia	2006-07	7	1.60	0.13	2.45
Burkina Faso	1992-93, 1998-99, 2003	13	1.84	0.13	4.02
Ghana	1993-94. 1998-99, 2003	10	2.17	0.95	3.72
Cote d'Ivoire	1994, 1998-99, 2005	9	4.72	1.71	5.84
Cameroon	1991, 1998, 2004	3	5.50	2.64	8.39
Tanzania	1999, 2004-05, 2007-08	20	5.73	1.63	16.39
Kenya	1993, 1998, 2003	7	6.76	4.21	14.99
Malawi	1992, 2000, 2004-05	3	11.79	6.44	17.56
Zambia	1992, 1997-98, 2001-02, 2007	9	14.27	6.79	20.75
Zimbabwe	1994, 1999, 2005-06	10	18.12	15.15	20.78
Lesotho	2004-05	10	25.35	17.93	29.70
Swaziland	2006-07	4	25.95	23.15	28.82

Table 2: Descriptive Statistics for women

	Cohort born 1958-1965	Cohort born 1966-1974	Cohort born 1975-1983
Married by age 20	0.743	0.692	0.655
Sexually active and unmarried by age 20	0.138	0.163	0.173
Lives in high-HIV country	0.413	0.427	0.452
Lives in urban area	0.286	0.312	0.331
Age	37.52	29.68	24.50
Literate	0.363	0.431	0.473
Has any education	0.438	0.522	0.580
Observations	62,657	90,626	76,788
		Surveys with HIV testing	
HIV positive at time of survey	0.085	0.118	0.102
Observations	11,868	18,667	27,043

Notes: The sample is for women ages 20 to 49 who were born between 1958 and 1983. All means are calculated using the relevant sample weights. HIV rates are based on the most recent year of survey data for each country, for all men and women.

Table 3: Relationship status and HIV infection. Dependent variable: Whether HIV positive.

	(1)	(2)	(3)	(4)
Married or cohabiting at age 20	0.013 (0.007)	0.027 (0.005)	0.025 (0.005)	0.017 (0.005)
Not married and sexually active by age 20	0.079 (0.020)	0.076 (0.010)	0.073 (0.010)	0.059 (0.008)
Literate	0.133 (0.013)	0.013 (0.005)	0.012 (0.005)	0.011 (0.005)
Never married, sexually active				0.080 (0.016)
Married or cohabiting				0.037 (0.009)
Widowed				0.259 (0.024)
Divorced				0.142 (0.017)
Separated				0.113 (0.015)
Country fixed effects		X		
Country-region fixed effects			X	X

Notes: Ordinary Least Squares. Observations=46,862. The sample consists of women ages 20 to 49, born between 1958 and 1983, who received an HIV test. All regressions control for a complete set of age indicators and an urban status indicator. Standard errors are clustered at the country/region level.

Table 4: Marital status, teen sexual behavior and education among women born from 1958-1965

Country	Current HIV rate (%)	Women born between 1958-65			
		Married by age 20 (%)	Unmarried & sexually active by age 20 (%)	Has any formal education (%)	Literate (%)
Senegal	0.70	72.34	5.13	2.75	25.29
Niger	0.73	93.07	1.00	9.20	6.68
Dem. Rep. of Congo	1.30	64.33	23.33	6.64	48.55
Mali	1.33	81.84	9.03	16.32	12.38
Ethiopia	1.47	84.39	3.22	11.18	14.48
Guinea	1.55	81.94	7.67	13.59	9.96
Liberia	1.60	66.00	28.03	30.54	21.54
Burkina Faso	1.84	85.28	6.41	10.82	8.51
Ghana	2.17	62.28	22.09	62.96	41.79
Cote d'Ivoire	4.72	64.57	28.02	32.63	30.09
Cameroon	5.50	72.65	21.30	64.12	54.52
Tanzania	5.73	65.68	19.91	61.67	56.15
Kenya	6.76	60.24	26.93	81.56	72.88
Malawi	11.79	75.67	12.47	55.05	44.36
Zambia	14.27	74.10	15.18	84.27	66.86
Zimbabwe	18.12	65.33	16.30	81.35	75.01
Lesotho	25.35	67.61	10.13	96.19	92.06
Swaziland	25.95	36.79	48.91	80.82	83.71
Pearson's ρ with HIV rate, country level		–0.581	0.424	0.751	0.839
Spearman's ρ with HIV rate, country level		–0.437	0.461	0.752	0.818
Pearson's ρ with HIV rate, country-region level		–0.357	0.241	0.664	0.741
Spearman's ρ with HIV rate, country-region level		–0.347	0.332	0.712	0.774

Notes: Marriage, sexual activity and literacy are measured using a sample of women ages 20 to 49 who were born between 1958 and 1965. The relevant sample weights are used for all calculations. HIV rates are based on the most recent year of survey data for each country, for all men and women.

Table 5: Marriage and sexual activity by age 20, cohort 3 born 1975-1983

	married by age 20	Δmarried by age 20	Sexually active and unmarried by age 20	ΔSexually active and unmarried by age 20
Mean of dependent variable among cohort 2, born 1966-1974	0.782 (0.038)		0.904 (0.056)	
Mean of HIV+, cohort 1, born 1958-1965	0.271 (0.098)	0.359 (0.108)	–0.115 (0.092)	–0.158 (0.085)
F-Test: coefficient on mean of dependent variable among cohort 2 equal to 1	33.00 (0.000)		2.95 (0.088)	
Observations	76,788	76,788	73,070	73,070

Notes: The sample consists of women ages 20 to 49 who were born between 1975 and 1983. Their outcomes are regressed on the mean of the dependent variable among women in the same country and region who were born between 1966 and 1975, and the HIV rate of women in the same country and region who were born between 1958 and 1965. Controls are also included for age, age squared, and indicators for year of birth, country and urban status. Standard errors are clustered at the country/region level.

Table 6: Marriage and sexual activity by age 20 and education, cohort 3, born 1975-1983

	married by age 20				Sexually active and unmarried by age 20			
Mean of dependent variable among cohort 2, born 1966-1974	0.661 (0.037)	0.649 (0.038)	0.674 (0.037)	0.669 (0.037)	0.815 (0.064)	0.819 (0.064)	0.834 (0.060)	0.838 (0.060)
Mean of HIV+, cohort 1, born 1958-1965	0.306 (0.090)	0.113 (0.119)	0.309 (0.085)	0.133 (0.112)	–0.126 (0.098)	–0.169 (0.114)	–0.121 (0.090)	–0.221 (0.106)
Literate	–0.191 (0.011)	–0.212 (0.012)			0.082 (0.008)	0.078 (0.012)		
Literate x (mean of HIV+, cohort 1)		0.284 (0.089)				0.062 (0.080)		
Any education			–0.167 (0.010)	–0.180 (0.011)			0.078 (0.009)	0.070 (0.011)
(Any education) x (mean of HIV+, cohort 1)				0.220 (0.092)				0.123 (0.080)
Observations	71,497		76,733		68,037		73,019	

Notes: The sample consists of women ages 20 to 49 who were born between 1975 and 1983. Their outcomes are regressed on the mean of the dependent variable among women in the same country and region who were born between 1966 and 1975, and the HIV rate of women in the same country and region who were born between 1958 and 1965, without and with interactions with education measures. Controls are also included for age, age squared, and indicators for year of birth, country and urban status. Standard errors are clustered at the country/region level.

Table 7: Associations between education and AIDS knowledge

Dependent variable	Heard of HIV/AIDS	Believes AIDS can be avoided	Believes that AIDS can be avoided by…			If believes AIDS can be avoided, believes AIDS can be spread by…	
			Condom use	Being faithful	Abstinence	Mosquitoes	Sharing food
Mean of dependent variable	0.936	0.834	0.613	0.739	0.708	0.310	0.198
Model 1							
Years of education	0.0047	0.0130	0.0221	0.0156	0.0134	–0.0157	–0.0125
	(0.0005)	(0.0010)	(0.0015)	(0.0011)	(0.0010)	(0.0015)	(0.0012)
Model 2							
Indicator: has any education	0.057	0.123	0.184	0.130	0.110	–0.075	–0.085
	(0.006)	(0.011)	(0.010)	(0.010)	(0.009)	(0.013)	(0.008)
Model 3							
Indicator: Literate	0.047	0.116	0.182	0.126	0.104	–0.094	–0.099
	(0.005)	(0.010)	(0.012)	(0.010)	(0.009)	(0.012)	(0.008)
Observations*	221,500	162,487	93,516	93,569	93,446	48,261	48,217

Notes: All regressions include controls for age, age squared, an indicator for urban status, indicators for whether the respondent's birth year, and a set of country/region dummies. Standard errors are clustered at the country/region level. Not all AIDS knowledge questions are asked in all countries in all years.
*The number of observations listed is for the sample used in Models 1 and 2. The number of observations is slightly lower in Model 3, due to missing values in the measure of literacy.

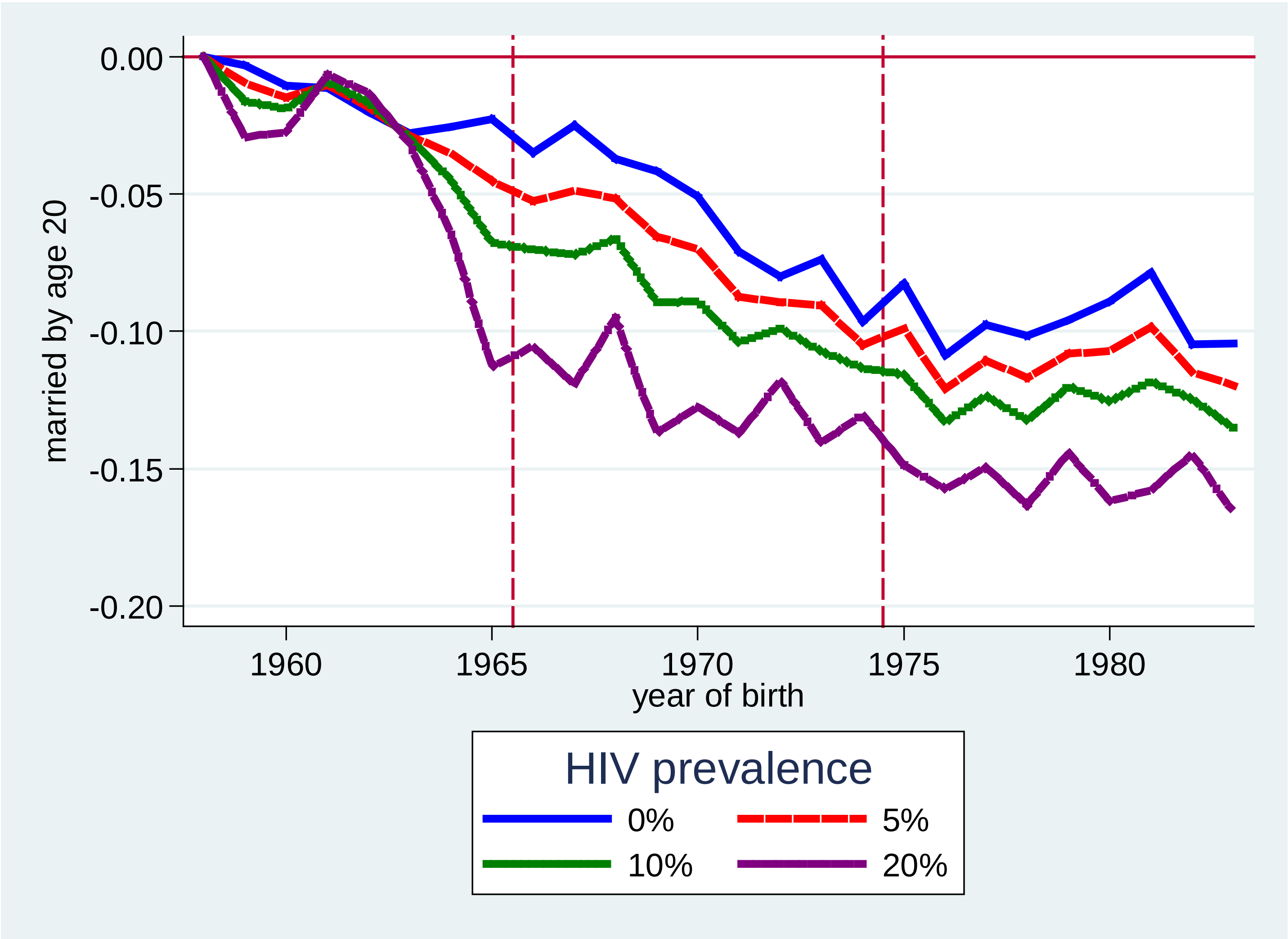

Figure 1: Teen marriage by birth cohort, across higher and lower HIV regions

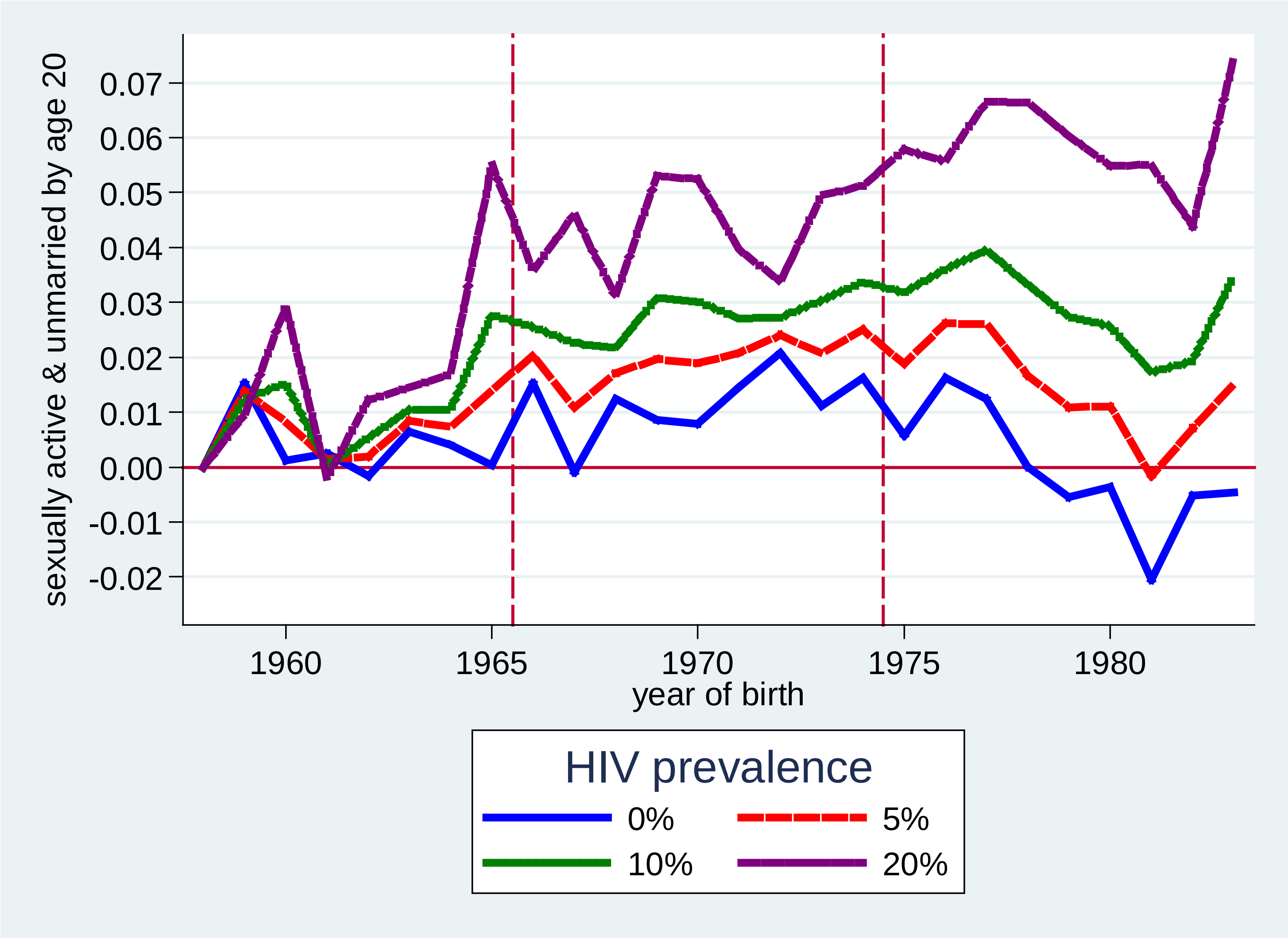

Figure 2: Teen sexually activity by birth cohort, across higher and lower HIV regions

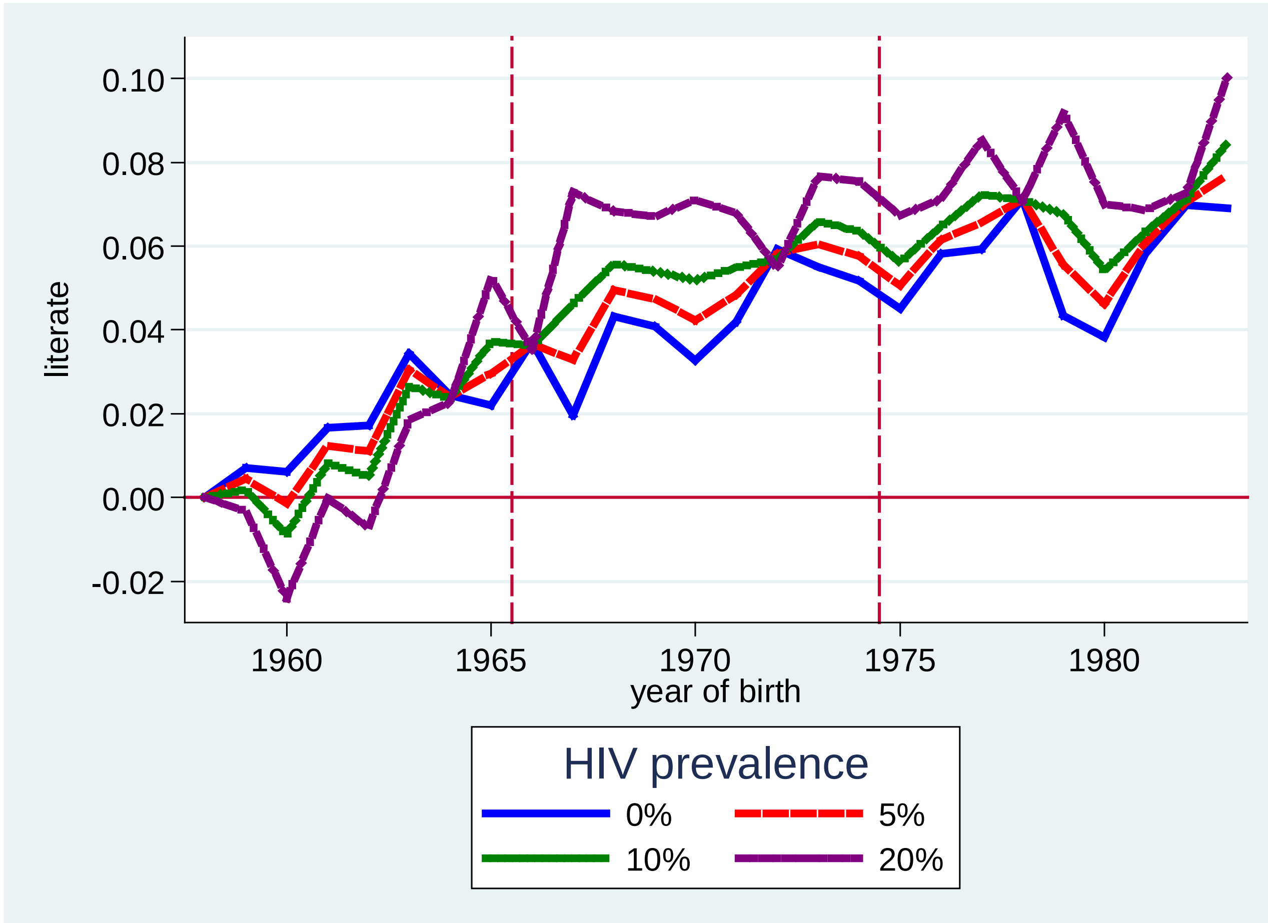

Figure 3: Literacy by birth cohort, across higher and lower HIV regions

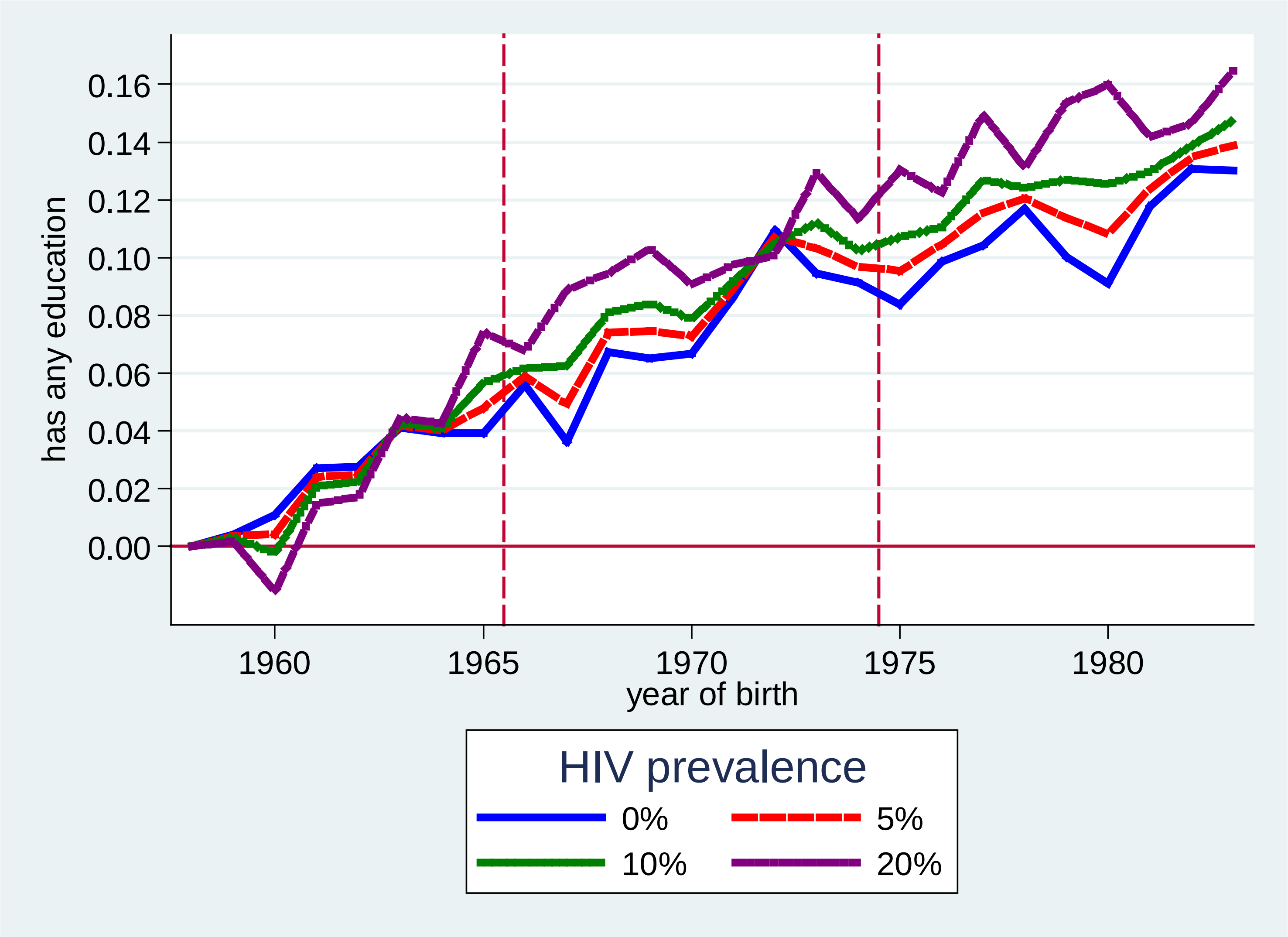

Figure 4: Any education by birth cohort, across higher and lower HIV regions