# The Impact of HIV/AIDS on Human Capital Investment in Sub-Saharan Africa: New Evidence

Chicoine, Luke and Lyons, Emily and Sahue, Alexia

Bates College, Bates College, Bates College

May 2020

# The Impact of HIV/AIDS on Human Capital Investment in Sub-Saharan Africa: New Evidence

**Luke Chicoine***
Bates College and IZA

**Emily Lyons**
Bates College

**Alexia Sahue**
Bates College

May 2020

The risk of AIDS-related mortality increased dramatically throughout the 1990s. This paper updates previous work by Fortson (2011) to examine the impact of mortality risk on human capital investment during the deadliest period of the pandemic. We combine Demographic Health Survey data from 30 countries, across 60 survey waves, to generate a sample of over 1,300,000 observations. Cohort-specific analysis using the updated sample yields new evidence that the negative relationship between HIV prevalence and schooling steepened as mortality risk increased. The reduction in schooling is largest for women, and along the extensive margin of the schooling decision. The findings indicate that the decline in human capital investment associated with the HIV/AIDS pandemic prior to the availability of treatment was larger in magnitude than previously understood, but may be reversing rapidly as access to treatment is expanded.

*JEL classification:* I15; I25; O55

*Keywords: HIV/AIDS; mortality risk; schooling*

*Corresponding author: Department of Economics, Bates College and IZA. Email: lchicoin@bates.edu. We would like to thank Jane Fortson, Susan Godlonton, Erick Gong, Michael Murray, Nathan Tefft, and Nicholas Wilson for their extremely helpful feedback.

# 1 Introduction

Over the course of three decades, the HIV/AIDS pandemic has significantly changed life expectancy and mortality risk on a global scale. Today, there are over 37 million people living with HIV, and roughly 1.7 million new people become infected every year.[1] However, the burden of infection is not equally distributed across countries. Sub-Saharan Africa accounts for 71 percent of the global infection rate, with prevalence rates exceeding 20 percent in some southern regions of the continent (Global Burden of Disease Collaborative Network, 2018). Mortality rates from HIV are highest for prime-age adults, making this pandemic uniquely consequential for human capital accumulation, labor market productivity, and economic growth.

Using 2001 to 2006 survey data for fifteen Sub-Saharan African countries, Fortson (2011) identified a negative relationship between HIV prevalence and human capital investment. This effect was found to primarily operate through an individual's expectation of a shorter life span, leaving fewer years to accumulate returns to human capital, and disincentivizing investment in schooling. Due to the increased salience of the mortality risk generated by the HIV/AIDS pandemic for post-1990 cohorts – the youngest fully captured birth cohort in Fortson (2011) – we investigate whether the effect of HIV on schooling changed for individuals born in the following decade. We use data from 2001 to 2017 Demographic and Health Surveys (DHS) for a sample of 30 Sub-Saharan Africa countries, and construct region definitions with consistent boundaries across time.

We find that the HIV pandemic has reduced educational attainment, and that the relationship between HIV prevalence and schooling became increasingly negative as mortality risk increased. The evidence is consistent with the economic theory that investment in schooling is curtailed in response reductions in life expectancy. A number of alternative explanations are ruled out; for example, there is no evidence that the results are driven by decisions to forfeit schooling to care for HIV-positive family members, nor due to increases in orphanhood. In contrast to Fortson (2011), we find that HIV has a larger effect on women's schooling, which is most associated with reductions on the extensive margin decision to enter school. Finally, we introduce evidence that the negative effect on the school entry decision has begun to move back towards zero for cohorts entering school following the rapid expansion of antiretroviral therapy (ART).

The difference in disease environment for the newly added cohorts is described in Figure 1 using time trends for HIV prevalence and AIDS related mortality counts at age six – the age at which the initial schooling decision is made. The expanded sample used throughout the paper includes the 1991 to 2000 cohorts, a period during which prevalence in Sub-Saharan Africa peaked and AIDS-related death counts continued to rise in the region. The early schooling decisions for these cohorts were being made prior to the

[1] Data are for the year 2018, and can be found at http://aidsinfo.unaids.org.

large expansion in access to treatment. The central analysis in this paper investigates whether investments in human capital evolved as the mortality risk of the disease continued to intensify. Supplementary analysis extends the sample to include children as young as seven years old and cohorts entering school after the peak in AIDS-related mortality.

The formal theory of the relationship between life expectancy and human capital investment was laid out in Jayachandran and Lleras-Muney (2009), and the negative relationship is shown to empirically exist in Sri Lanka, a setting entirely independent from the one studied here. Godlonton and Thornton (2013) and Gong (2014) have documented that updated information regarding the severity of HIV/AIDS risk or HIV status can lead to changes in behavior. However, as adult access to ART continues to expand and life expectancy losses are reversed (Bor et al., 2013), increases in expected longevity is a channel that may motivate parents to reinvest in the education of their children (Baranov and Kohler, 2018; Lucas et al., 2019), their child's health (Lucas and Wilson, 2013), increase savings (Baranov and Kohler, 2018), and enter the labor market (Baranov et al., 2015). The expansion of ART has also been found to explain a significant portion of recent increases in economic growth throughout Africa (Tompsett, 2020).

The findings in this paper suggest that the previous work examining HIV/AIDS related mortality and human capital investment was too early to capture the evolution of the relationship as the pandemic peaked and has now begun to retreat. Furthermore, the greatest reductions in schooling are found in the earliest grades and for women, this introduces the question of whether the pandemic has contributed to broader inequality in schooling attainment.

# 2 Data and Empirical Methodology

## 2.1 Data

Fortson (2011) used household survey data from 2001 to 2006 DHS for 157 regions in 15 Sub-Saharan African countries, and the corresponding HIV-test result data to assign regional HIV prevalence rates. Shown in the following section, our narrow replication of Fortson (2011) uses her exact data and specifications, and our slightly modified replication makes two changes: *(i)* population weights are adjusted using World Development Indicators (World Bank, 2020) population estimates for the survey years, instead of 2007 population counts for all survey waves, and *(ii)* pre-1965 cohorts are excluded to maintain a consistent sample when analyzing cohort specific effects while ensuring a sufficient number of observations in each cohort. Following successful replication, we expand on Fortson (2011)'s findings in a number of ways. First, we include updated survey waves from 2007 to 2017 for the original countries to account for the progression of HIV infection across an additional decade.[2] Second, 15 additional countries are included using survey waves

[2] In addition to the original survey waves from Fortson (2011), the updated survey years for the original fifteen countries

from 2007 to 2016.[3] Each survey wave includes an HIV-testing module that is used to calculate regional prevalence rates. In each sample, regions are redefined to be consistent across survey rounds; therefore, when the sample of countries is held constant, but the range of survey years is increased, the number of observed boundary consistent regions is reduced to accommodate changes in region definitions between survey rounds.[4]

Table 1 provides summary statistics of individual characteristics for children and adults using the sample of all 30 countries. In the adult sample, the average regional HIV prevalence rate is 5.3 percent, with a standard deviation of 6.4; the children sample yields similar information with an average of 4.7 and standard deviation of 5.8. A summary of regional HIV prevalence rates by survey year for the thirty countries in our sample is provided in Appendix Table A.1. A change of 6.4 percentage points, the standard deviation in the adult sample, is similar to a change in the prevalence rate from Rwanda (2014) to Malawi (2015). Prevalence rates vary considerably across and within countries, and over survey years. Over the last two decades, a number of countries have been making considerable progress in reducing prevalence rates (e.g. Malawi, Zambia, Zimbabwe), while other have stagnated or continued to see rates increase (e.g. Côte d'Ivoire, Lesotho, Mozambique). Nationally, the highest HIV prevalence rate in the sample is 24.17 percent in Lesotho in 2014, and the lowest is 0.33 in Niger in 2012.

For adults, we measure education attainment using years of schooling, and two indicator variables for whether the individual completed any school (at least one year), and if they completed their primary school education. As shown in Table 1, adults in the sample completed an average of 5.7 years of schooling, 74 percent have completed at least one year, and 48 percent completed primary school. Educational attainment information is also included for children; however, due to their age, schooling is likely to be incomplete. Therefore, we also measure children's progression through school by calculating the number of years children are behind the appropriate grade for their age. In this sample, children are an average of 2.2 years behind grade for age.

Finally, we construct cohort-specific estimates of regional HIV prevalence rates for the time period covering the initial outbreak of the disease through 2016. This cohort-specific measure is constructed using annual country-level estimates from the Global Burden of Disease (GBD) Study (Global Burden of Disease Collaborative Network, 2018; Murray et al., 2018; Roth et al., 2018) and combined with relative DHS prevalence rates in regions across each country, from each survey wave. For example, the prevalence rate in the

include, Burkina Faso (2010), Cameroon (2011), Cote d'Ivoire (2011), Ethiopia (2011, 2016), Ghana (2008, 2014), Guinea (2012), Kenya (2008, 2012), Lesotho (2009, 2014), Malawi (2010, 2015), Mali (2006, 2012), Niger (2012), Rwanda (2010, 2014), Senegal (2010, 2014), Tanzania (2007, 2011), and Zambia (2007, 2013).

[3] The new countries and survey years include Angola (2015), Burundi (2010, 2016), Chad (2014), Congo (2005, 2009, 2011), Democratic Republic of Congo (2007, 2013), Gabon (2012), Gambia (2013), Liberia (2007, 2013), Mozambique (2009, 2015), Namibia (2006, 2013), Sierra Leone (2008, 2013), South Africa (2016), Togo (2013), Uganda (2006, 2011, 2016), and Zimbabwe (2005, 2010, 2015).

[4] Boundary consistent region definitions are taken from IPUMS-DHS when possible, and supplemented when necessary using shapefiles available at the DHS Program's Spatial Data Repository and information from statoids.com.

northern region of Malawi was 68 percent of the national prevalence in the 2004 wave of the DHS (8.04 relative to 11.79); therefore, the northern region is assigned 68 percent of the GBD's national estimate in 2004 (10.5 versus 15.4). For non-survey years, national prevalence is estimated assuming a linear trend in regional prevalence between survey years. For all years prior to the first DHS survey (and after the most recent wave), the dispersion across regions is assumed to be constant, and equal to the relative prevalence in the initial (final) survey. When these data are used in analysis, observations from each birth cohort are matched with the prevalence rate in their region during their age six year. This is to simulate the mortality risk environment in which they, or their family, make their initial schooling decisions. Using an annual estimate allows us to both expand the sample to include DHS household survey data from the 1990s to account for the effect of HIV prevalence on human capital investment over an additional decade, and to introduce variation in the prevalence rate across cohorts.

## 2.2 Empirical Methodology

Fortson (2011) used a difference-in-differences (DiD) method comparing educational outcomes of pre- and post-1980 cohorts across regions with varying HIV prevalence rates. Educational outcomes (i.e. years of schooling, any schooling, and primary school completion) were regressed on HIV prevalence in the survey year, interacted with an indicator for the post-1980 cohort. The schooling decisions for the first post-1980 cohorts were made based on prevalence information from the late 1980s and early 1990s, largely before the scale of the mortality risk was fully understood.[5] In both the replication, and when using the expanded sample, we largely follow the same estimation equation:

$$\begin{aligned} S_{iycr} = \quad & \beta_0 + \beta_1 HIV_{yr} \times \mathbb{1}\left[c \geq 1980\right] + \beta_2 Female_{iycr} + \beta_3 Rural_{iycr} \\ & + \theta g\left(age_{iycr}\right) + \gamma_c + \alpha_{ry} + \varepsilon_{iycr}. \end{aligned} \tag{1}$$

$S_{iycr}$ is an educational outcome for respondent $i$, observed in year $y$, in birth cohort $c$, and region $r$. Initial replication estimates include only a single survey for each country. In this case, the main explanatory variable interacts regional HIV prevalence from the only available survey year, $HIV_r$, and an indicator for being born in or after 1980, $\mathbb{1}\left[c \geq 1980\right]$. The coefficient on this variable, $\beta_1$, is the estimated effect of an additional percentage point of HIV prevalence on the human capital measure, $S_{iycr}$, for the post-1980 cohorts relative to individuals born before the cutoff. There are two additional indicators, one for whether the individual is female, and one for living in a rural area; $\gamma_c$ is a fixed effect for birth cohort. In samples without multiple survey rounds for each country, $\alpha_r$ is a fixed effect for region $r$; when the sample is expanded, we substitute

[5]The first rapid HIV test was not approved by the U.S. Food and Drug Administration (FDA) until in 1992; the current definition of progression to AIDS, a CD4 count below 200, was only established U.S. Center for Disease Control in 1993 (Kaiser Family Foundation); UNAIDS came into existence in 1996 (avert.org).

a vector of region-by-year fixed effects ($\alpha_{ry}$) and a cubic control for age. Finally, because the ideal data would have cohort specific measures of prevalence, instead of variation in the year of survey, an additional specification substitutes the cohort- and region-specific GBD estimates described in the previous section for the interaction term in equation (1). In all specifications, standard errors are clustered at the region level to allow for arbitrary within region correlation in the error term (Bertrand et al., 2004).

The model described in equation (1) assumes that there is a change in the relationship between schooling and future HIV prevalence for individuals born after 1980. To investigate this assumption, we allow the model to estimate the relationship between a time-invariant prevalence measure ($HIV_r^{max}$) and schooling for each cohort, where $HIV_r^{max}$ is the maximum observed prevalence rate in each region across DHS survey rounds. The set of cohort-specific estimates are generated using the following model:

$$\begin{aligned} S_{iycr} = \quad & \beta_0 + HIV_r^{max}\left[\sum_{b=1965}^{1978}\beta_b \mathbb{1}\left[c=b\right] + \sum_{b=1980}^{2000}\beta_b \mathbb{1}\left[c=b\right]\right] \\ & + \beta_2 Female_{iycr} + \beta_3 Rural_{iycr} + \theta g\left(age_{iycr}\right) + \gamma_c + \alpha_{ry} + \varepsilon_{iycr}. \end{aligned} \tag{2}$$

The outcome variable, $S_{iycr}$, is the same as the previous equation. $HIV_r^{max}$ is the HIV prevalence measure in each region $r$. $\mathbb{1}\left[c=b\right]$ is an indicator equal to one when individual $i$ is born in year $c$. The 1979 cohort, defined as the final pre-1980 cohort in Fortson (2011), is left to be the reference group. All other variables are the same as equation (1). The coefficient, $\beta_b$, is the cohort specific effect of an additional percentage point of HIV prevalence, relative to the omitted cohort. If the underlying assumptions of the DiD model described in equation (1) are valid, these estimates should not be different from zero prior to the 1980 cohort.

# 3 Results

## 3.1 Baseline Difference-in-Differences Estimates

The results of the replication of Fortson (2011)'s central model are shown in column (1) of each panel in Table 2. All results in Table 2 use a sample of adults from 15 to 49 years old who were born in or after 1965. Our estimates for the effect of HIV prevalence on attending any school, column (1) in Panel B, and completing primary school, Panel C, are almost identical to the original findings, while our estimate for the effect of HIV on years of schooling is slightly larger (-0.067 compared to -0.053 in Fortson, 2011).[6]

In Table 2, column (2) of each panel includes new survey waves for the fifteen original countries, adding data from 2006–2017. In this specification, we use a regional HIV prevalence measure specific to both region and survey year, include a cubic polynomial for age, and reduce the number of regions to 131 to

[6]We are also able to directly replicate Fortson (2011)'s exact model, and match within 0.001 of her estimated coefficient on the interaction term. The output comparing the original model to the alternations in our preferred specifications can be found in Appendix Table A.2. The only differences between column (1) of Table 2 and Fortson (2011)'s Table 4 are the use of survey year specific populations to scale sample weights, instead of 2007 populations for all survey years, and restricting the sample to the post-1965 set of cohorts.

maintain regional boundaries consistent throughout the extended time period. In the updated sample, a one percentage point increase in HIV prevalence decreases educational attainment by 0.06 years, the likelihood of attending any school by 0.9 percentage points, and the completion of primary school by 0.5 percentage points.

The remaining columns in Table 2 expand the sample of countries to include a total of 30 countries, and 289 regions, using survey waves from 2001-2017. Findings for the effect of HIV prevalence on years of schooling with two different definitions of regional fixed effects are shown in columns (3) and (4). Adding the region by survey year fixed effects in column (4) yields nearly identical results. Notably, when the sample is expanded to 30 countries, the relationship between HIV prevalence and the completion of primary school, shown in Panel C of Table 2, is no longer statistically significant. The effect on primary school completion may be less robust than previously understood, suggesting a more prominent relationship between HIV and dropout in the early years of primary school, and specifically, the extensive margin of the school entry decision.[7]

The results in the first four columns of Table 2 use prevalence rates associated with the year each adult is observed. To ensure these findings are not attributable to variation in HIV prevalence across the survey waves of each country, we use the highest regional HIV prevalence rate across the survey waves as our HIV measure. When doing so, the magnitude and statistical significance remain consistent for all education outcomes, shown in column (5). In fact, although smaller in magnitude, the result for primary school completion is again statistically significant.

To this point, the measures of HIV prevalence from DHS modules are likely accurate but may not reflect the distribution of perceived mortality risk at the time of each individual's schooling decision. To address this issue, we introduce a cohort-specific estimate of regional HIV prevalence that assigns prevalence at age six to each observation, which we consider to be roughly the age at which the initial human capital investment decision for children is made. Although a number of assumptions are needed to create these cohort-specific regional prevalence estimates, this method concedes some accuracy to approximate the ideal dataset of cohort specific mortality risk. Furthermore, the cohort-specific measures allow us to expand the sample to DHS waves without attached HIV data; 45 additional DHS waves without an attached HIV module – largely from the 1990s – are included in the sample. The results for this expanded sample and cohort-specific measure, reported in column (6), remain largely unchanged. A one-percentage point increase in age six HIV prevalence is associated with a decrease of 0.039 years of schooling and a 0.6 percentage point decrease in the likelihood of attending any school. The coefficient on primary completion is again statistically significant but smaller

[7] Further evidence that the declines in schooling are concentrated among those dropping out of the earliest years is shown in Appendix Table A.4. The reduced likelihood of grade completion is largest for grade one and remains negative and statistically significant through at least grade five, but the magnitude of the effect becomes more muted at each successive level.

in magnitude than the original estimates.

From these results, we can confirm the negative effect of HIV prevalence on human capital investment originally found in Fortson (2011) for a larger sample of countries, with over two decades of HIV prevalence rate data. In our fully extended sample with most flexible controls, column (4), we find that a one percentage point increase in HIV prevalence is associated with a reduction in schooling of 0.033 years, and a decreased likelihood of ever attending school of 0.7 percentage points. Evidence of reductions in schooling in the early years of primary school suggests that the HIV/AIDS pandemic may be further amplifying the inequality in human capital investment between those who see significant returns from large investments, and marginal students for whom small reductions in expected returns make any schooling investment suboptimal.

### 3.2 Cohort Specific Effect of HIV Prevalence

The expanded sample allows for a more detailed cohort-specific analysis. Equation (2) is used to estimate a set of cohort-specific effects from 1965 to 2000, with the 1979 cohort set as the omitted group. The magnitude of these estimates and their confidence intervals are shown in Figure 2. For all three outcomes, prior to 1980, there is no persistent relationship between the future maximum HIV prevalence rate and the education outcome being examined. This finding reinforces the DiD assumption necessary to estimate equation (1), that future HIV prevalence rates are not correlated with schooling decisions for cohorts born prior to 1980.[8]

The results in Panels A and B of Figure 2 demonstrate that HIV has a progressively more negative effect on years of schooling and attending any school for younger cohorts. For each outcome in the first two panels, the estimated effect is negative and statistically significant beginning with the 1987 cohort, and then continues to increase in magnitude. In both cases, the estimated effect for the final cohort in the sample is more than twice as large as the average effect shown in Table 2. It is important to note that the schooling entry decision for the entire set of cohorts is largely made prior to the expansion of access to treatment; this extensive margin decision impacts a relatively large portion of the population, and changes in years of schooling moves along a similar pattern.

As seen in Panel C of Figure 2, the effect on primary completion is negative over the same range of cohorts. However, the decline does not become progressively more negative for later cohorts in the sample. The final decision to complete primary school is being made six to eight years after the school entry decision. For these cohorts, these decisions are more likely to be impacted by the availability of ART and increases in life expectancy that accompany the treatment. While certainly not conclusive, the findings in Panel C of

[8] A replication of the placebo test performed in Fortson (2011)'s Table 6 is included in Appendix Table A.3. Although there are some small negative estimates for the relationship between any schooling and prevalence, Panel B of Figure 2 yields more detailed evidence of a lack of a pre-1980 relationship, one that consistently hovers around zero for cohorts in the 1970s, and through 1986.

Figure 2 behave in a pattern consistent with an initial decline in life expectancy that is then halted if not reversed as the benefits of treatment are beginning to be realized.

In addition to the trends seen in Figure 2, since there is not a relationship between HIV and human capital investment for pre-1987 cohorts, the DiD estimate defining 1980 as the point when treatment is introduced seems to include a number of "pre-treatment" cohorts in the post-treatment period. This likely leads to an underestimate of the true effect, assuming that the understanding of the risk changed closer to 1987, as suggested by Figure 2. Re-estimating the DiD model, equation (1), using 1987 as the cutoff cohort behaves as expected, yielding larger estimated effects for all three outcomes, each statistically significant at the 99 percent confidence level. For example, the estimated effect on years of schooling increases from 0.046 in the original specification to 0.089 in the 1987 specification.[9]

### 3.3 Exploring Alternative Channels of Effect

The strong negative relationship between HIV prevalence and human capital investment is consistent with the human capital model described in Fortson (2011). However, in addition to mortality risk, there are several other mechanisms through which HIV prevalence can affect education outcomes. A number of these alternative mechanisms are explored here.

**Orphanhood and caretaking requirements.** Orphans are less likely to attend school than nonorphans (Evans and Miguel, 2007; Haacker, 2016); therefore, if high-HIV areas have a disproportionate number of orphans, it could explain the strong negative relationship between HIV prevalence and human capital investment. Another explanation is caretaking requirements, which could reduce time spent in school for children that live in households with HIV-positive adults. To test these pathways we measure human capital investment using years behind in school for age using a sample of children aged 7 to 14. From the data, we know whether the parents of these children are alive, and within the HIV-testing sample, the HIV status of the adults within their household. This replicates Table 8 in Fortson (2011).

The DiD model described by equation (1) is again used here; following Fortson (2011), the 1992 cohort is used as the treatment threshold. For this sample, HIV prevalence – defined by region and survey year – is ideally recorded at the time of the decision regarding the child's schooling. The results are shown in Table 3.[10] The effect of HIV prevalence on the prescribed progression through school is shown in column (1) for all children, and in column (2) for all nonorphan children (defined as having both a living mother and father). For both samples, HIV prevalence slows the progression through school to a similar degree, indicating that orphanhood is not a pathway through which HIV prevalence impacts human capital investment. The caretaking requirement pathway is tested in columns (3) and (4). The sample in column (3) includes only

[9] These estimates can be found in Appendix Table A.5.

[10] Estimates directly replicating Table 8 in Fortson (2011) can be found in Appendix Table A.6.

children whose household was selected for the HIV testing module, and column (4) restricts the sample to children who live in households with no HIV-positive members. Given that HIV prevalence has an almost equal effect on children with no caretaking requirements, an increased need for caretaking cannot explain the negative effect of HIV prevalence on schooling.

**Migration and Differential Mortality.** Due to migration, individuals observed as adults may have faced a significantly different HIV environment when their education decisions were made. By using children aged 7-14 in our analysis, we are able to avoid this concern. In this analysis, we remove the measurement error introduced when observing post-migration adults by matching regional prevalence data to individuals at the age when their schooling decision is made. The findings from Appendix Table 3 demonstrate that the negative relationship between HIV and schooling exists prior to the possibility of any adult migration decision.

Alternatively, if HIV prevalence is highest among those who are highly educated (Fortson, 2008) then the association between HIV prevalence and lower levels of schooling could be biased a survivor bias. To investigate this the analysis from Fortson (2011)'s Table 7 is repeated and the sample is restricted to those between the ages of 15 and 25. The mortality rate of this age range is less than one-fifth of the 25 to 49 age range (IHME, 2015). The estimates for the paper's three main schooling outcomes are negative and statistically significant, and are shown in Appendix Table A.7. These findings, along with the negative effect on schooling for the children aged 7-14, demonstrate that the results exist for samples not greatly affected by the high rates of mortality. These two sets of findings suggest that the relationship between HIV and education outcomes is not driven by migration and differential mortality.

**Child Labor.** Heightened levels of prime-aged adult mortality could be leading to shortages in labor supply which would generate an upward pressure on wages. It is possible that children are leaving school to fill these higher paying vacancies. To examine this possibility Fortson (2011)'s model for children aged 7 to 14 is again estimated, using data from the 20 (of 60) survey waves with a child labor module.[11] For this sample, we find no evidence that higher rates of HIV prevalence are associated with an increased likelihood of working for pay for non-family members, or working in any type of family enterprise. We then show that higher rates of HIV prevalence continue to be associated with falling further behind in school, for both the full sample of children from these 20 surveys and after dropping all children reported to be engaging in either type of work. These results demonstrate that the negative impact on schooling persists even for children not attracted to the labor market due to potentially higher wage rates.

[11] These results are shown in Appendix Table A.8.

### 3.4 Differences between Men and Women

We might expect HIV prevalence to affect human capital investment for men and women differently due to differences in the timing of HIV mortality rates, generally affecting women at younger ages.[12] Figure 3 shows the effect of regional HIV prevalence on years of school completed for men (Panel A) and women (Panel B) born from 1965 to 2000. The decline seems to begin slightly earlier for men; however, the first statistically significant negative result is estimated in 1987 for women and 1988 for men. The pronounced decline in women's schooling is driven by a significantly larger decline in the school entry decision. This suggests the possibility that, on average, the decision to send women to school may be more marginal. In all cases, the effect on men's schooling investment does not continue to decline for the later cohorts; in fact, the effect for men's primary completion moves back towards zero for the latest cohorts in the sample.[13]

The findings seen in Figure 3 are the opposite of the those in Table 9 of Fortson (2011).[14] However, the key conclusion taken from these results holds, differences in the evolution of the effect on educational outcomes between the sexes suggests that HIV prevalence is not affecting human capital investment through a third possible channel of schooling provision. If educational attainment was falling due to a reduction in supply, possibly caused by fewer available teachers, the pattern of the effect across cohorts would likely be similar for both men and women. However, the decline in years of schooling for women begins slightly later and all outcomes show evidence of a steadily increasingly negative effect. Alternatively, the effect on men's schooling seems to plateau for the later cohorts.

### 3.5 Expansion of ART and Completing Any School

Antiretroviral therapy is able to substantially increase life expectancy (Bor et al., 2013), and since 2000, access to this treatment has rapidly expanded. As late as 2003 less than 1% of the HIV positive population in the 30 countries examined here had access to treatment (World Bank, 2020). By 2009 – the year the first students from the 2002 cohort completed grade one – more than 18% of the HIV positive population had access to ART, and in 2015 more than 50% of the relevant population were being treated.

If the decline in schooling is driven by increased mortality risk generated by HIV/AIDS, the expansion of treatment should reduce the effect once the improvement in life expectancy is incorporated into the decision-making process. However, the most recent survey used in the study is from 2017; those who are 15 years

[12]In 2005, HIV/AIDS mortality rates were twice as high for women between the ages of 20 and 24, 40% higher between the ages of 25 and 29, and 13% higher between the ages of 30 and 34 (IHME, 2015). Cohort specific estimates of AIDS-caused mortality rates for South Africa yield a similar finding (Chicoine, 2012).

[13]The charts with grade one and primary completion for both men and women can be found in Appendix Figure A.1.

[14]There are three key explanations for the divergence of the sex-specific results from Fortson (2011). These are explored in Appendix Table A.9. First, the original model was not a fully specified triple difference model, once the missing interaction terms are added the effect of HIV prevalence on women's schooling becomes negative – column (2). Second, the data from the fifteen new countries removes any evidence that the effect for men differs – column (4). Finally, the cohort-specific estimates uncover dynamics across time that are not visible when assuming treatment occurs at a single point in time.

old in this survey wave were born in 2002, prior to a majority of the expansion of ART. Therefore, to study the period of ART expansion younger ages have to be included in the sample. To do this, we expand the sample to include ages from 7 to 49 and focus on the any schooling outcome. Although there are a number of children between the ages of 7 and 15 who have not yet attended school but will do so in the future, this outcome provides the best comparison across these ages and an initial understanding of whether the trend we have seen through the 2000 cohort using the baseline sample persists for these younger cohorts.

To conduct this analysis, equation (2) is estimated for the expanded sample; the output is shown in Figure 4. The pattern is nearly identical to Panel B of Figure 2 through the 1997 cohort. After 1997, the decline is less steep over the next four cohorts, before beginning to move back towards zero for those born in 2002. In fact, by 2008 the effect is no longer statistically significant and is less than 30% the size of the effect for the 2001 cohort – the largest estimate in the sample. While there is certainly more work to be done to definitively link these findings to local measures of mortality risk, this evidence is consistent with the theory that increased mortality risk generated by the HIV/AIDS pandemic led to significant declines in human capital investment.[15]

# 4 Conclusion

Consistent with Fortson (2011), we find a strong negative relationship between HIV prevalence and human capital investment. The effect of HIV on educational attainment is consistent with economic theory – an expectation of a shorter lifespan can lead to reduced investments in schooling. We find no evidence that orphanhood, caretaking requirements, migration patterns, survivor bias, increases in child labor, or the provision of schooling are able to explain the relationship between HIV and reduced human capital investment found throughout the paper.

Importantly, our findings provide evidence that the consequences of the HIV/AIDS pandemic for schooling may be greater than previously documented. The larger effects on the extensive margin suggest that the increased mortality risk may be driving students at the margin of attending school out of the classroom earlier. Students are completing fewer year of school and are less likely to attend any school as the risk of future mortality increases. In the aggregate, this effect may be widening the gap in educational attainment between students with high expected returns and those on the margin of attendance. The expanded inequality has the potential to amplify over time, as it has been shown that worsened economic conditions, in this case possibly generated by lower levels of schooling, can lead to higher rates of HIV (Burke et al., 2014), and higher rates of HIV have also been shown to feed back to lower wages and employment rates (Chicoine, 2012;

[15] Additional estimates of the effect of HIV prevalence on the likelihood of completing at least grade one using the expanded sample can be found in the appendix. In Appendix Figure A.1, for men (Panel E) and women (Panel F); and in Appendix Figure A.3, for urban (Panel A) and rural (Panel B) households.

Levinsohn et al., 2013).

The findings are consistent with increased access to AIDS treatment improving life expectancy and a reducing in the impact of the disease on human capital investment. As more data on ART become available, it will be important to further investigate how reductions in mortality, increases in life expectancy (Bor et al., 2013), and changes in behavior (Thirumurthy and Graff Zivin, 2012; Baranov et al., 2015; Baranov and Kohler, 2018; Friedman, 2018) can lead to updated expectations of longevity and further reverse the decline in human capital investments found in this paper. Additionally, if socioeconomic status is positively correlated with access to treatment, the treatments themselves could further compound inequalities in human capital accumulation already magnified by the pandemic. Finally, future research aimed at better understanding the differences in response by gender could play a vital role in uncovering the more detailed mechanisms of how mortality risk associated with HIV/AIDS has altered the human capital investment decision.

## References


Baranov V, Bennett D, Kohler HP. 2015. The Indirect Impact of Antiretroviral Therapy: Mortality Risk, Mental Health, and HIV-negative Labor Supply. *Journal of Health Economics* **44**: 195–211.

Baranov V, Kohler HP. 2018. The Impact of AIDS Treatment on Savings and Human Capital Investment in Malawi. *American Economic Journal: Applied Economics* **10**: 266–306.

Bertrand M, Duflo E, Mullainathan S. 2004. How Much Should We Trust Differences-In-Differences Estimates? *Quarterly Journal of Economics* **119**: 249–275.

Bor J, Herbst AJ, Newell ML, Bärnighausen T. 2013. Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-up of HIV Treatment. *Science* **339**: 961–965.

Burke M, Gong E, Jones K. 2014. Income Shocks and HIV in Africa. *Economic Journal* **125**: 1157–1189.

Chicoine L. 2012. AIDS Mortality and Its Effect on the Labor Market: Evidence from South Africa. *Journal of Development Economics* **98**: 256–269.

Evans DK, Miguel E. 2007. Orphans and Schooling in Africa: A Longitudinal Analysis. *Demography* **44**: 35–57.

Fortson JG. 2008. The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS. *Demography* **45**: 303–322.

Fortson JG. 2011. Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa. *Review of Economics and Statistics* **93**: 1–15.

Friedman WH. 2018. Antiretroviral Drug Access and Behavior Change. *Journal of Development Economics* **135**: 392–411.

Global Burden of Disease Collaborative Network. 2018. Global Burden of Disease Study 2017 (GBD 2017) Covariates 1980-2017. University of Washington: Institute for Health Metrics and Evaluation (IHME).

Godlonton S, Thornton RL. 2013. Learning from Others' HIV Testing: Updating Beliefs and Responding to Risk. *American Economic Review* **103**: 439–44.

Gong E. 2014. HIV Testing and Risky Sexual Behaviour. *Economic Journal* **125**: 32–60.

Haacker M. 2016. *The Economics of the Global Response to HIV/AIDS*. Oxford University Press.

Institute for Health Metrics and Evaluation (IHME). 2015. HIV Worldwide 1990-2013. Seattle, WA: IHME, University of Washington. Available from http://vizhub.healthdata.org/hiv (Accessed 2/11/20).

Jayachandran S, Lleras-Muney A. 2009. Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines. *Quarterly Journal of Economics* **124**: 349–397.

Levinsohn J, McLaren ZM, Shisana O, Zuma K. 2013. HIV Status and Labor Market Participation in South Africa. *Review of Economics and Statistics* **95**: 98–108.

Lucas AM, Chidothe M, Wilson NL. 2019. Effects of Adult Health Interventions at Scale on Children's Schooling: Evidence from Antiretroviral Therapy in Zambia. *Economics of Education Review* **72**: 107–120.

Lucas AM, Wilson NL. 2013. Adult Antiretroviral Therapy and Child Health: Evidence from Scale-up in Zambia. *American Economic Review* **103**: 456–61.

Murray CJ, Callender CS, Kulikoff XR, Srinivasan V, Abate D, Abate KH, Abay SM, Abbasi N, Abbastabar H, Abdela J, et al. 2018. Population and Fertility by Age and Sex for 195 Countries and Territories, 1950–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**: 1995–2051.

Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A, et al. 2018. Global, Regional, and National Age-sex-specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**: 1736–1788.

Thirumurthy H, Graff Zivin J. 2012. Health and Labor Supply in the Context of HIV/AIDS: The Long-run Economic Impacts of Antiretroviral Therapy. *Economic Development and Cultural Change* **61**: 73–96.

Tompsett A. 2020. The lazarus drug: The impact of antiretroviral therapy on economic growth. *Journal of Development Economics* **143**: 102409.

World Bank. 2020. World Development Indicators. Washington, D.C. Available from http://wdi.worldbank.org (Accessed 2/11/20).

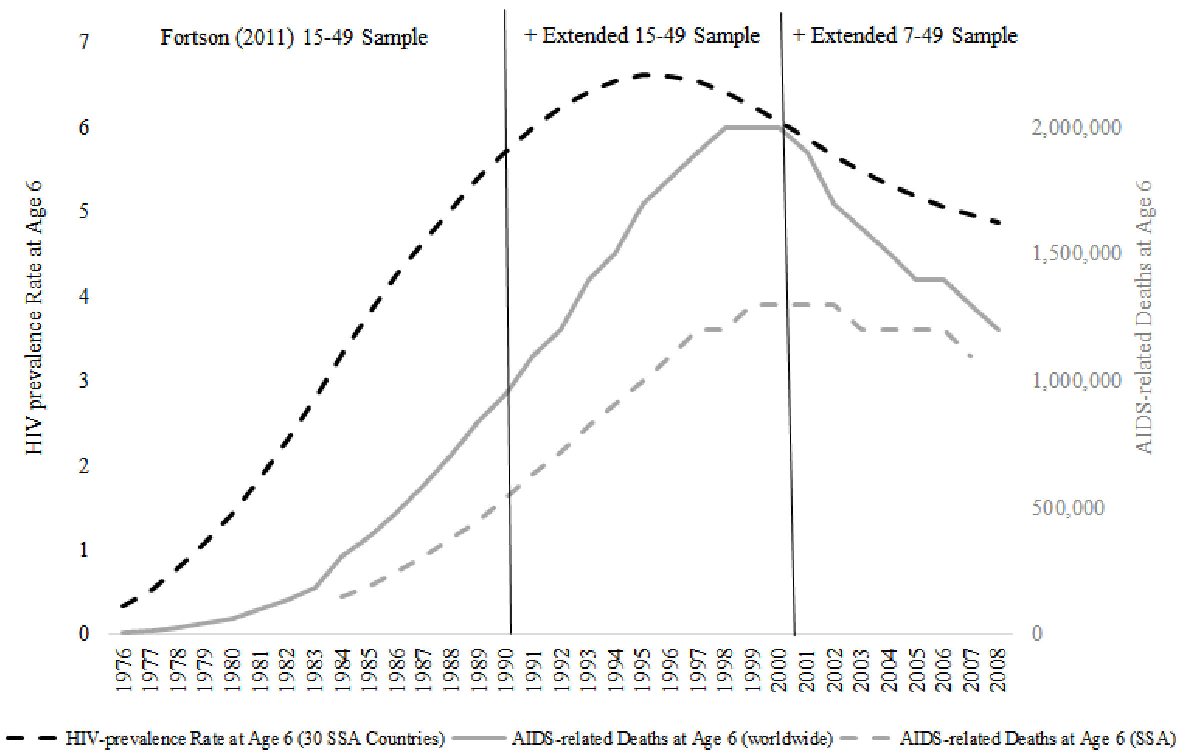
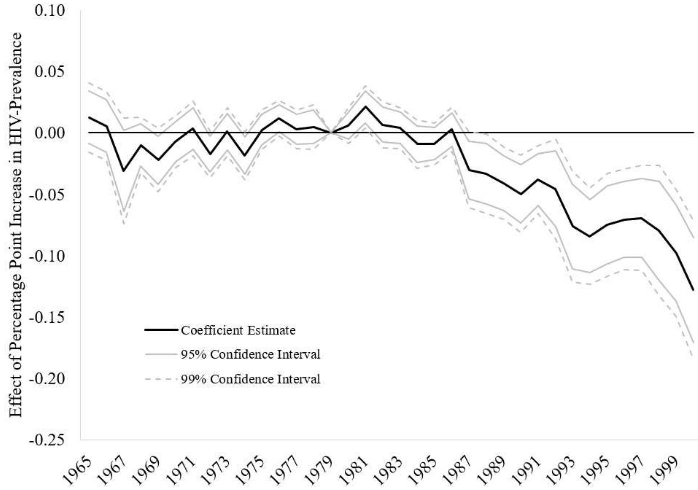
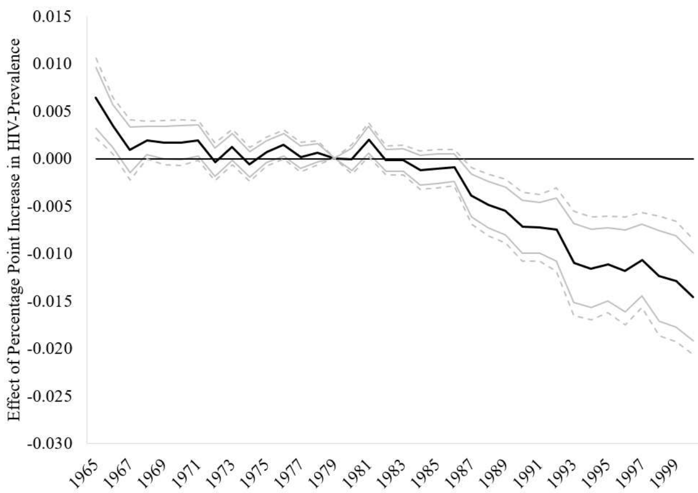



Figure 1: AIDS-related Deaths and HIV Prevalence at Age Six, by Birth Year

Note: Birth years are matched with statistics for the age six year of each cohort, approximately when the initial schooling decision is made. Time trends represent HIV prevalence rates across the 30 country sample (Global Burden of Disease Collaborative Network, 2018) and count estimates of AIDS-related deaths (worldwide from avert.org; Sub-Saharan Africa from IHME, 2015).



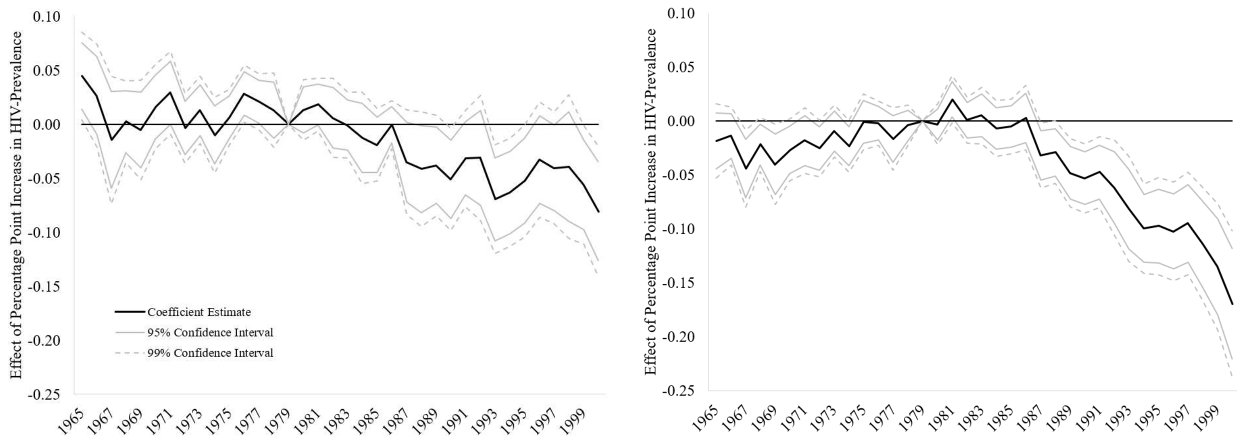

(a) Years of Schooling




(b) Any School (Years of School > 0)


(c) Completed Primary School

Figure 2: Effect of HIV Prevalence on Education Outcomes, by Birth Year




(a) Men (b) Women

Figure 3: Effect of HIV Prevalence on Years of School, by Birth Year: Differences by Sex

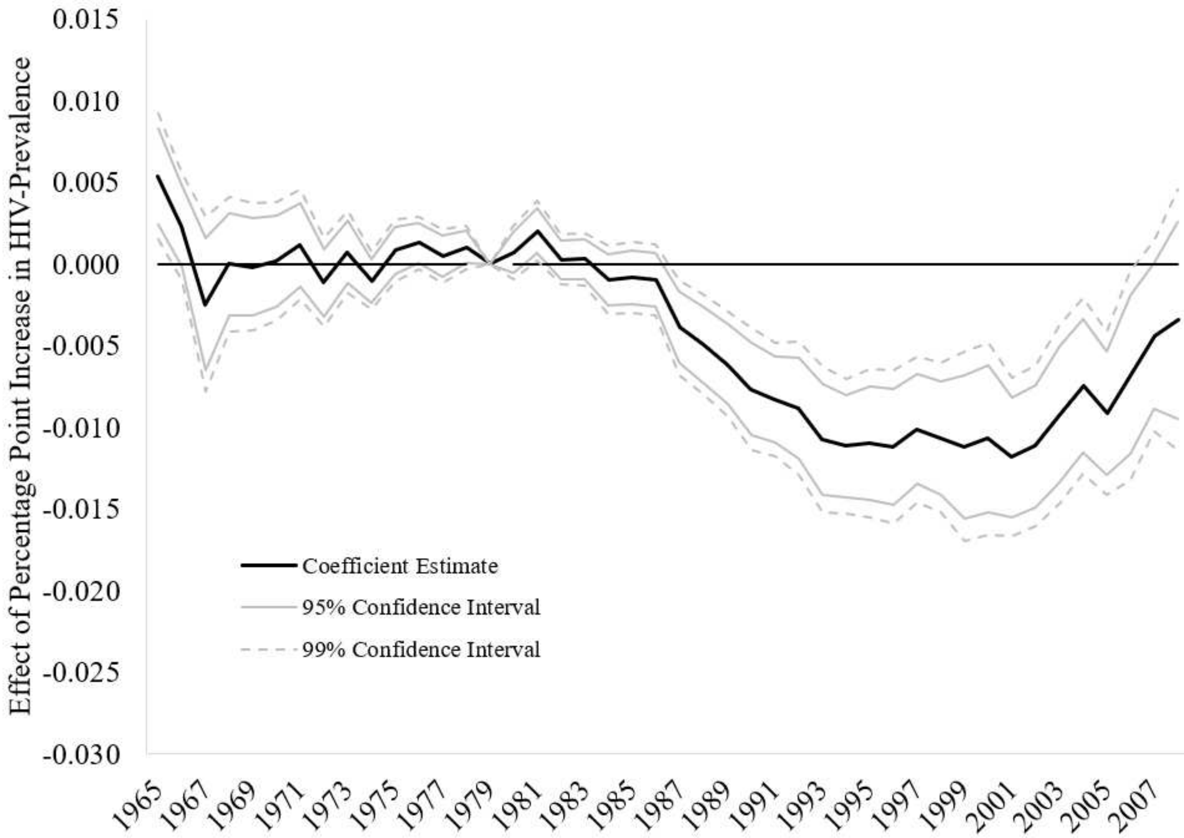



Figure 4: Effect of HIV Prevalence on Completing Any School (Years of Schooling > 0; Ages 7-49)

Table 1: Summary Statistics: Individuals

| | Age 7−14 | | | Age 15−49 | | |
|---|---|---|---|---|---|---|
| | N | Mean | s.d. | N | Mean | s.d. |
| HIV Prevalence (region; survey year) | 787,840 | 4.682 | 5.763 | 1,329,306 | 5.331 | 6.407 |
| Years of schooling | 784,438 | 2.179 | 2.167 | 1,320,655 | 5.651 | 4.528 |
| Any Schooling (Years of schooling > 0) | 784,438 | 0.680 | 0.467 | 1,320,655 | 0.737 | 0.440 |
| Completed primary school | 785,173 | 0.058 | 0.234 | 1,323,191 | 0.479 | 0.500 |
| Years behind grade for age | 775,083 | 2.240 | 1.921 | – | – | – |
| Year of birth | 787,840 | 2000 | 4.933 | 1,329,306 | 1983 | 9.119 |
| Female | 787,821 | 0.493 | 0.500 | 1,329,304 | 0.530 | 0.499 |
| Rural | 787,840 | 0.750 | 0.433 | 1,329,306 | 0.650 | 0.477 |

Note: The sample includes individuals from DHS survey waves from 2001-2017 for the 30 countries described in footnotes 2 and 3. Each panel includes observations within the stated age range, and born in or after 1965.

Table 2: Effect of HIV on Education Outcomes

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **A. Dependent Variable: Years of School** | | | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.067*** | -0.059*** | -0.034*** | -0.033*** | | |
| | (0.022) | (0.016) | (0.011) | (0.011) | | |
| Regional HIV Prevalence$_{ry}$ | | 0.027 | 0.037 | | | |
| | | (0.034) | (0.023) | | | |
| Max. Regional HIV Prevalence$_r$ × Post-1980 Cohort$_c$ | | | | | -0.033*** | |
| | | | | | (0.011) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | | | | -0.039*** |
| | | | | | | (0.015) |
| Regions | 157 | 131 | 289 | 289 | 289 | 254 |
| N | 253,324 | 845,684 | 1,320,654 | 1,320,654 | 1,320,654 | 1,751,328 |
| **B. Dependent Variable: Any Schooling (Years of School > 0)** | | | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.006*** | -0.009*** | -0.007*** | -0.007*** | | |
| | (0.002) | (0.002) | (0.001) | (0.001) | | |
| Regional HIV Prevalence$_{ry}$ | | 0.011*** | 0.008** | | | |
| | | (0.003) | (0.003) | | | |
| Max. Regional HIV Prevalence$_r$ × Post-1980 Cohort$_c$ | | | | | -0.007*** | |
| | | | | | (0.001) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | | | | -0.006*** |
| | | | | | | (0.002) |
| Regions | 157 | 131 | 289 | 289 | 289 | 254 |
| N | 253,324 | 845,684 | 1,320,654 | 1,320,654 | 1,320,654 | 1,751,328 |
| **C. Dependent Variable: Primary Completed** | | | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.009*** | -0.005*** | -0.001 | -0.001 | | |
| | (0.002) | (0.001) | (0.001) | (0.001) | | |
| Regional HIV Prevalence$_{ry}$ | | -0.001 | 0.002 | | | |
| | | (0.004) | (0.003) | | | |
| Max. Regional HIV Prevalence$_r$ × Post-1980 Cohort$_c$ | | | | | -0.002** | |
| | | | | | (0.001) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | | | | -0.003** |
| | | | | | | (0.001) |
| Regions | 157 | 131 | 289 | 289 | 289 | 254 |
| N | 253,513 | 846,325 | 1,323,189 | 1,323,189 | 1,323,189 | 1,754,126 |
| Sample | Original countries and survey waves (2001−2005) | Original countries; expanded survey waves (2001−2017) | Expanded countries and survey waves (2001−2017) | | | Expanded countries and survey waves (1991−2017) |
| Age (cubic) | | X | X | X | X | X |
| Female indicator | X | X | X | X | X | X |
| Rural indicator | X | X | X | X | X | X |
| Birth year F.E. | X | X | X | X | X | X |
| Region F.E. | X | X | X | | | |
| Region × survey wave F.E. | | | | X | X | X |

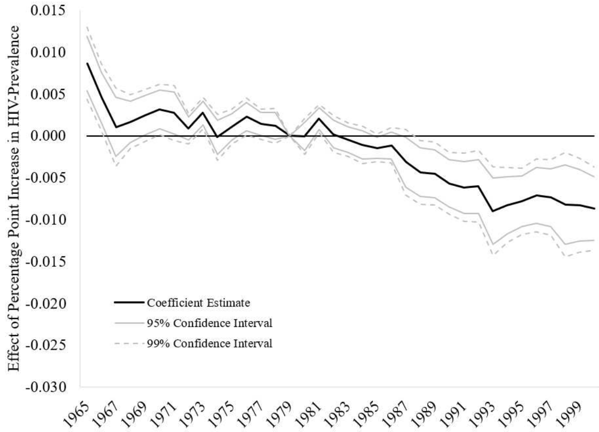
Note: The dependent variable is described at the top of each panel. In Panel A, it is years of schooling; an indicator for completing any schooling in Panel B; and in Panel C, an indicator for completing primary school. The set of controls used in each regression is detailed at the bottom of the table. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Max. regional HIV prevalence is a time-consistent measure of the highest recorded prevalence rate in the region across all survey years, and regional HIV prevalence at age six is a cohort specific estimate of the regional prevalence rate for each cohort's age six year. All samples include adults between the ages of 15 and 49, born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table 3: Effect of HIV on Years Behind Grade for age: Ages 7 to 14

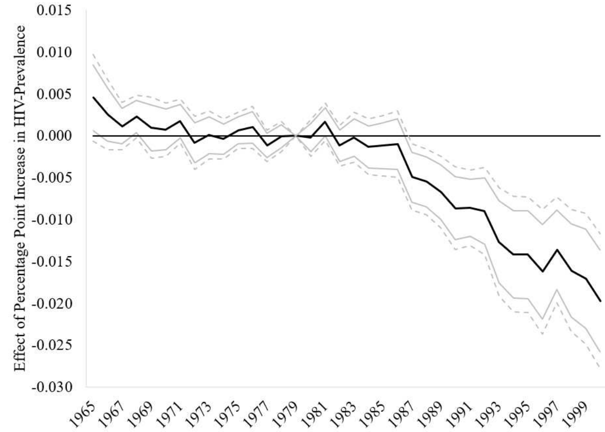
| | Full Sample | No Orphans | Testing Sample | No HIV+ Member |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Regional HIV Prevalence$_{ry}$ | 0.076*** | 0.079*** | 0.075*** | 0.075*** |
| × Post-1992 Cohort$_c$ | (0.010) | (0.011) | (0.014) | (0.016) |
| Sample | Expanded countries and survey waves (2001−2017) | Nonorphans; Expanded countries and survey waves (2001−2017) | Nonorphans in HIV testing sample; expanded countries and survey waves (2001−2017) | Nonorphans in HH with no HIV-positive individual; expanded countries and survey waves (2001−2017) |
| Regions | 289 | 289 | 289 | 289 |
| N | 775,064 | 653,798 | 367,377 | 332,555 |

Note: The dependent variable is years behind grade for age. Each regression includes a cubic for age, indicators for female and living in a rural area, a set of birth year fixed effects, and region by survey wave fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Nonorphan is defined as having both mother and father alive; columns (3) and (4) include only households within the HIV testing sample. All samples include observations between the ages of 7 and 14. Standard errors are clustered by region and shown in parentheses.

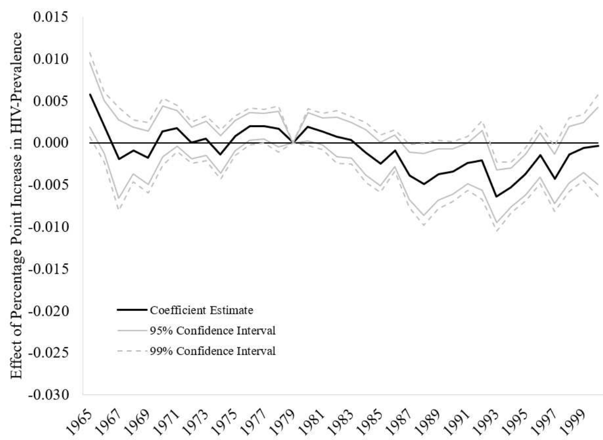
# Appendix (For Online Publication)



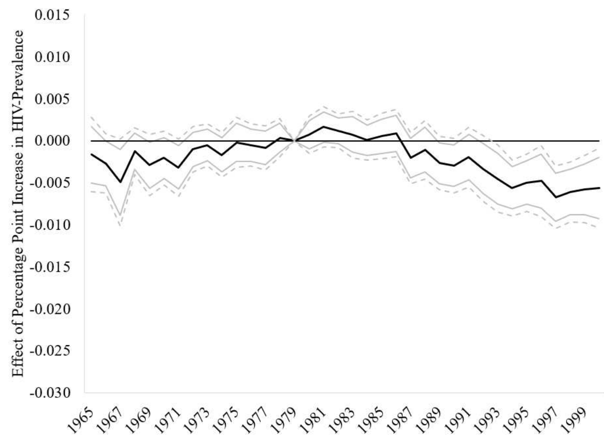

(a) Any School – Men (Ages 15-49)



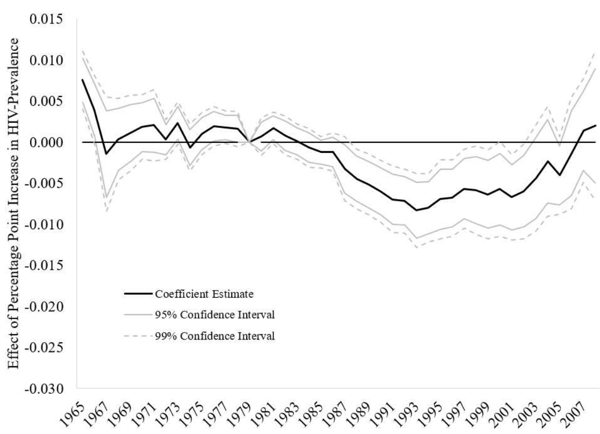

(b) Any School – Women (Ages 15-49)



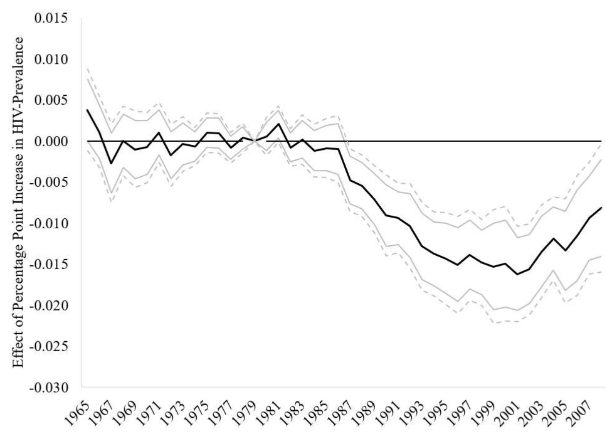

(c) Completed Primary School – Men (Ages 15-49)




(d) Completed Primary School – Women (Ages 15-49)

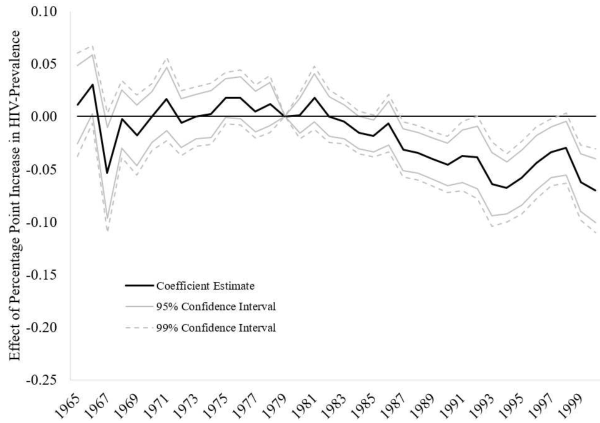



(e) Any School – Men (Ages 7-49)

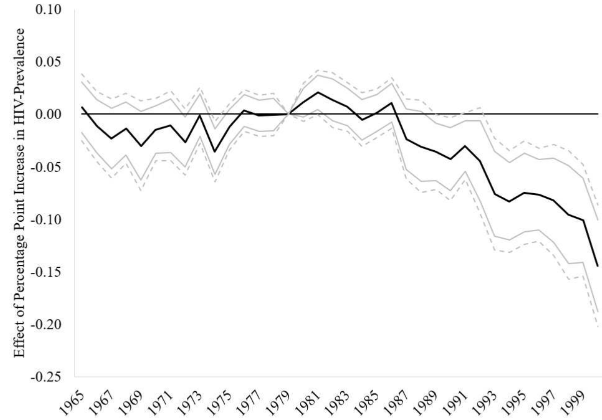



(f) Any School – Wiomen (Ages 7-49)

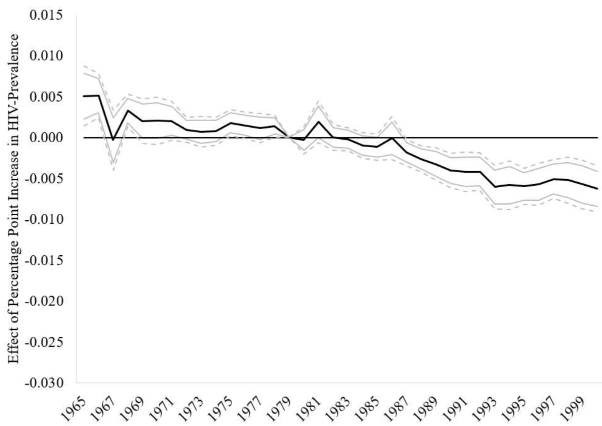
Figure A.1: Effect of HIV Prevalence on Education Outcomes by Birth Year: Differences by Sex




(a) Years of Schooling – Urban

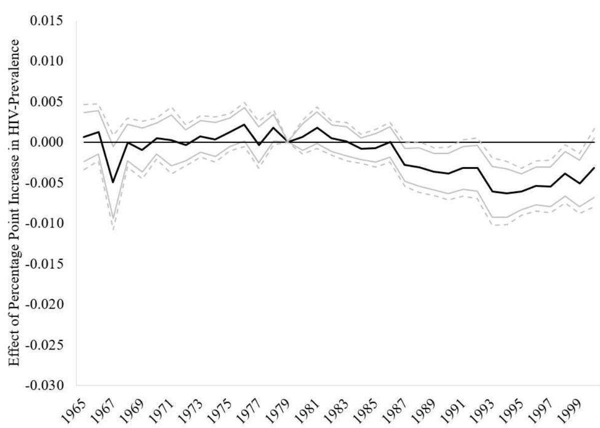



(b) Years of Schooling – Rural




(c) Any School – Urban

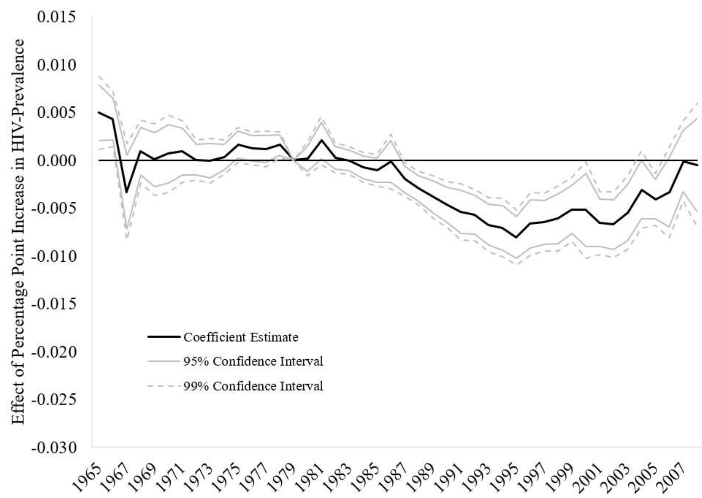

(d) Any School – Rural



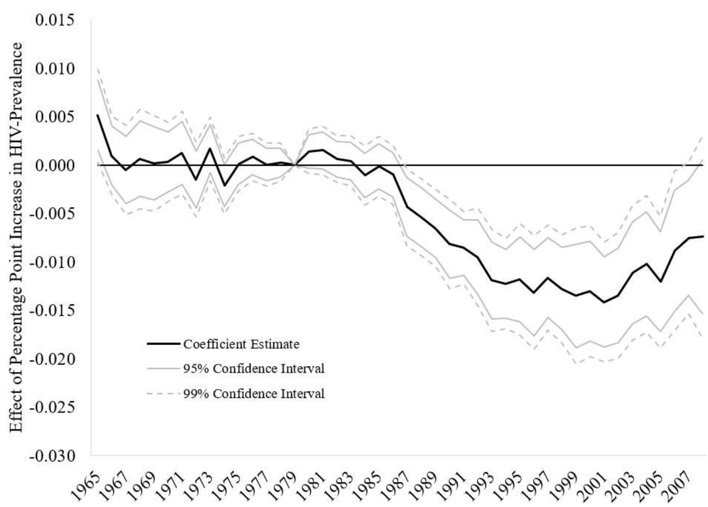

(e) Completed Primary School – Urban


(f) Completed Primary School – Rural

Figure A.2: Effect of HIV Prevalence on Education Outcomes by Birth Year, by Urban / Rural Status




(a) Any School (Ages 7-49); Urban Only




(b) Any School (Ages 7-49); Rural Only

Figure A.3: Effect of HIV Prevalence on Any School (Years of Schooling > 0; Ages 7-49) by Urban / Rural Status

Table A.1: Regional HIV Prevalence, by Country and Survey Year

| | (1) Angola | (2) Burkina Faso | | (3) Burundi | | (4) Cameroon | | (5) Chad | (6) Congo |
|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2015 | 2003 | 2010 | 2010 | 2016 | 2004 | 2011 | 2014 | 2009 |
| Mean | 1.99 | 1.77 | 0.95 | 1.34 | 0.91 | 5.27 | 4.10 | 1.49 | 3.20 |
| S.d. | 1.20 | 1.09 | 0.56 | 0.74 | 0.51 | 2.67 | 2.10 | 1.19 | 1.17 |
| Regions | 18 | 13 | 13 | 4 | 4 | 12 | 12 | 21 | 12 |

| | (7) DRC | | (8) Côte d'Ivoire | | (9) Ethiopia | | | (10) Gabon | (11) Gambia |
|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2007 | 2013 | 2005 | 2011 | 2005 | 2010 | 2016 | 2012 | 2013 |
| Mean | 1.32 | 1.15 | 4.65 | 3.63 | 1.42 | 1.37 | 0.86 | 4.12 | 1.9 |
| S.d. | 0.78 | 0.89 | 1.24 | 1.00 | 0.97 | 0.95 | 0.68 | 0.99 | 0.6 |
| Regions | 11 | 11 | 10 | 10 | 11 | 11 | 11 | 10 | 8 |

| | (12) Ghana | | (13) Guinea | | (14) Kenya | | (15) Lesotho | | | (16) Liberia | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2003 | 2014 | 2005 | 2012 | 2003 | 2008 | 2004 | 2009 | 2014 | 2007 | 2013 |
| Mean | 2.21 | 1.99 | 1.50 | 1.71 | 6.74 | 6.36 | 23.35 | 22.77 | 24.17 | 1.55 | 2.04 |
| S.d. | 0.90 | 0.81 | 0.55 | 0.57 | 4.14 | 3.88 | 3.67 | 2.83 | 3.38 | 0.85 | 1.18 |
| Regions | 10 | 10 | 5 | 5 | 8 | 8 | 10 | 10 | 10 | 5 | 5 |

| | (17) Malawi | | | (18) Mali | | | (19) Mozambique | | (20) Namibia | (21) Niger | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2004 | 2010 | 2015 | 2001 | 2006 | 2012 | 2009 | 2015 | 2013 | 2006 | 2012 |
| Mean | 11.64 | 10.59 | 9.13 | 1.72 | 1.29 | 1.12 | 11.13 | 13.11 | 14.29 | 0.69 | 0.33 |
| S.d. | 6.57 | 4.33 | 4.68 | 0.55 | 0.48 | 0.32 | 6.12 | 6.38 | 3.61 | 0.41 | 0.25 |
| Regions | 3 | 3 | 3 | 9 | 9 | 6 | 11 | 11 | 13 | 8 | 8 |

| | (22) Rwanda | | | (23) Senegal | | | (24) Sierra Leone | | (25) South Africa |
|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2005 | 2010 | 2014 | 2005 | 2010 | 2017 | 2008 | 2013 | 2016 |
| Mean | 2.99 | 2.92 | 2.87 | 0.71 | 0.67 | 0.47 | 1.47 | 1.41 | 21.25 |
| S.d. | 1.33 | 1.65 | 1.37 | 0.59 | 0.43 | 0.34 | 0.79 | 0.64 | 5.96 |
| Regions | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 9 |

| | (26) Tanzania | | | (27) Togo | (28) Uganda | (29) Zambia | | | (30) Zimbabwe | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Survey Year | 2003 | 2007 | 2011 | 2013 | 2011 | 2001 | 2007 | 2013 | 2005 | 2010 | 2015 |
| Mean | 6.85 | 5.67 | 5.04 | 2.40 | 7.29 | 15.33 | 13.85 | 12.95 | 18.12 | 15.29 | 13.81 |
| S.d. | 3.20 | 3.41 | 2.50 | 1.01 | 2.07 | 4.94 | 4.93 | 3.77 | 1.75 | 2.17 | 2.64 |
| Regions | 19 | 20 | 20 | 6 | 10 | 9 | 9 | 9 | 10 | 10 | 10 |

Note: In 2012, data are missing for the regions Tombouctou, Gao and Kidal in Mali and are therefore omitted from this last wave of data. The same applies to Tanzania in 2003 for missing data in Zanzibar.

Table A.2: Replication of Fortson (2011) Table 4: Consequences of Country Population Selection and Cohort Restriction

| | Years of School | | Any Schooling | | Completed Primary School | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | A. Direct Replication of Fortson (2011) – No Cohort Restriction | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.054** | -0.056*** | -0.006*** | -0.006*** | -0.008*** | -0.008*** |
| × Post-1980 Cohort$_c$ | (0.021) | (0.021) | (0.002) | (0.002) | (0.002) | (0.002) |
| Rural$_{iycr}$ | -2.705*** | -2.729*** | -0.195*** | -0.198*** | -0.276*** | -0.278*** |
| | (0.272) | (0.284) | (0.026) | (0.027) | (0.027) | (0.028) |
| Female$_{iycr}$ | -1.288*** | -1.297*** | -0.141*** | -0.143*** | -0.112*** | -0.113*** |
| | (0.084) | (0.086) | (0.018) | (0.019) | (0.008) | (0.008) |
| Sample | Original Population Weights (2007) | Survey Year Population Weights | Original Population Weights (2007) | Survey Year Population Weights | Original Population Weights (2007) | Survey Year Population Weights |
| Regions | 157 | 157 | 157 | 157 | 157 | 157 |
| Observations | 302,494 | 302,494 | 302,494 | 302,494 | 302,745 | 302,745 |
| | B. Restricted to Birth Year ≥ 1965 | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.066*** | -0.067*** | -0.006*** | -0.006*** | -0.009*** | -0.009*** |
| × Post-1980 Cohort$_c$ | (0.022) | (0.022) | (0.002) | (0.002) | (0.002) | (0.002) |
| Rural$_{iycr}$ | -2.717*** | -2.744*** | -0.196*** | -0.199*** | -0.282*** | -0.285*** |
| | (0.288) | (0.301) | (0.027) | (0.028) | (0.029) | (0.030) |
| Female$_{iycr}$ | -1.153*** | -1.166*** | -0.132*** | -0.135*** | -0.098*** | -0.100*** |
| | (0.095) | (0.098) | (0.019) | (0.019) | (0.009) | (0.009) |
| Sample | Original Population Weights (2007) | Survey Year Population Weights | Original Population Weights (2007) | Survey Year Population Weights | Original Population Weights (2007) | Survey Year Population Weights |
| Regions | 157 | 157 | 157 | 157 | 157 | 157 |
| Observations | 253,324 | 253,324 | 253,324 | 253,324 | 253,513 | 253,513 |

Note: The dependent variable is described at the top of each column. In columns (1) and (2), it is years of schooling; an indicator for completing any schooling in columns (3) and (4); and in columns (5) and (6), an indicator for completing primary school. Each regression includes indicators for female and living in a rural area, birth year fixed effects, and region fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Odd numbered columns use 2007 CIA World Factbook population data to adjust DHS weights so that the sum of each country's weights equals the population; even numbered columns use survey year WDI population data – as is done in the rest of the paper. All samples include adults between the ages of 15 and 49; restricted to cohorts born in or after 1965, in Panel B. Standard errors are clustered by region and shown in parentheses.

Table A.3: Replication of Fortson (2011) Table 6 – Differences Prior to Affected Time Period

| | (1) | (2) | | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|
| | | | **A. Dependent Variable: Years of Schooling** | | | | |
| Regional HIV Prevalence$_{ry}$ | 0.021* | -0.003 | Regional HIV Prevalence$_{ry}$ | 0.009 | | 0.013** | |
| $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.012) | (0.010) | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.006) | | (0.006) | |
| | | | Max. Regional HIV Prevalence$_r$ | | 0.006 | | 0.011* |
| | | | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | | (0.005) | | (0.005) |
| Regions | 157 | 157 | Regions | 289 | 289 | 289 | 289 |
| N | 170,542 | 121,372 | N | 452,771 | 452,771 | 681,018 | 681,018 |
| | | | **B. Dependent Variable: Any Schooling (Years of School > 0)** | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.002** | -0.003*** | Regional HIV Prevalence$_{ry}$ | -0.002*** | | -0.001* | |
| $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.001) | (0.001) | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.001) | | (0.001) | |
| | | | Max. Regional HIV Prevalence$_r$ | | -0.002*** | | -0.001** |
| | | | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | | (0.001) | | (0.001) |
| Regions | 157 | 157 | Regions | 289 | 289 | 289 | 289 |
| N | 170,542 | 121,372 | N | 452,771 | 452,771 | 681,018 | 681,018 |
| | | | **C. Dependent Variable: Completed Primary School** | | | | |
| Regional HIV Prevalence$_{ry}$ | 0.001 | -0.002*** | Regional HIV Prevalence$_{ry}$ | 0.001 | | 0.002*** | |
| $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.001) | (0.001) | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | (0.001) | | (0.000) | |
| | | | Max. Regional HIV Prevalence$_r$ | | 0.001 | | 0.001** |
| | | | $\times$ $\mathbb{1}$[Birth Year $\geq$ Placebo$_c$] | | (0.001) | | (0.000) |
| Regions | 157 | 157 | Regions | 289 | 289 | 289 | 289 |
| N | 170,741 | 121,509 | N | 454,031 | 454,031 | 682,794 | 682,794 |
| Sample | Original countries and survey years | | Sample | Expanded countries and survey years | | | |
| Placebo Birth Year Cutoff | 1970 | 1970 | Placebo Birth Year Cutoff | 1970 | 1970 | 1975 | 1975 |
| Cohort restrictions | | | Cohort restrictions | | | | |
| Early | | $\geq$ 1965 | Early | $\geq$ 1965 | $\geq$ 1965 | $\geq$ 1965 | $\geq$ 1965 |
| Late | < 1980 | < 1980 | Late | < 1980 | < 1980 | < 1980 | < 1980 |
| Region FE | X | X | Region FE | | | | |
| Region $\times$ Wave FE | | | Region $\times$ Wave FE | X | X | X | X |

Note: The dependent variable is described at the top of each panel. In Panel A, it is years of schooling; an indicator for completing any schooling in Panel B; and in Panel C, an indicator for completing primary school. Each regression includes indicators for female and living in a rural area, birth year fixed effects, and the set of region or region by survey wave fixed effects, as denoted at the bottom of the table. In the expanded sample, a cubic for age is also included. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Max. regional HIV prevalence is a time-consistent measure of the highest recorded prevalence rate in the region across all survey years. The sample only includes pre-1980 cohorts, and observations between the ages of 15 and 49, born in or after 1965; the exact placebo cutoff and cohort restrictions are denoted in each column. Standard errors are clustered by region and shown in parentheses.

Table A.4: Effect of HIV on Grade-by-Grade School Completion

| At least X Years Completed: | 1 | 2 | 3 | 4 | 5 | Primary |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | A. Regional HIV Prevalence × Post-1980 Cohort | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.007*** | -0.006*** | -0.005*** | -0.004*** | -0.003** | -0.001 |
| × Post-1980 Cohort$_c$ | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Regions | 289 | 289 | 289 | 289 | 289 | 289 |
| N | 1,320,654 | 1,320,654 | 1,320,654 | 1,320,654 | 1,320,654 | 1,323,189 |
| | B. Max. Regional HIV Prevalence × Post-1980 Cohort | | | | | |
| Max. Regional HIV Prevalence$_r$ | -0.007*** | -0.006*** | -0.005*** | -0.004*** | -0.003*** | -0.002** |
| × Post-1980 Cohort$_c$ | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Regions | 289 | 289 | 289 | 289 | 289 | 289 |
| N | 1,320,654 | 1,320,654 | 1,320,654 | 1,320,654 | 1,320,654 | 1,323,189 |
| | C. Regional HIV Prevalence at Age 6 | | | | | |
| Regional HIV Prevalence | -0.006*** | -0.005*** | -0.005*** | -0.004*** | -0.004*** | -0.003** |
| at Age 6$_{rc}$ | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| Regions | 254 | 254 | 254 | 254 | 254 | 254 |
| N | 1,751,328 | 1,751,328 | 1,751,328 | 1,751,328 | 1,751,328 | 1,754,126 |

Note: The dependent variable is described at the top of each column. In columns (1) through (5), it is an indicator equal to one if at least the stated number of years of schooling were completed. In column (6), the dependent variable is an indicator equal to one if primary school was completed. Each regression includes a cubic for age, indicators for female and living in a rural area, birth year fixed effects, and region by survey wave fixed effects. In Panel A, Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed; in Panel B, HIV prevalence is measured as a time-consistent measure of the highest record prevalence rate in the region across all survey years; in Panel C, regional HIV prevalence is a cohort specific estimate of the regional prevalence rate for each cohort's age six year. All samples include adults between the ages of 15 and 49; restricted to cohorts born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table A.5: Effect of HIV on Education Outcomes - 1987 Treatment Cutoff

| | Years of School | Any Schooling | Completed Primary School |
|---|---|---|---|
| | (1) | (2) | (3) |
| Regional HIV Prevalence$_{ry}$ | -0.063*** | -0.009*** | -0.003*** |
| × Post-1987 Cohort$_c$ | (0.013) | (0.002) | (0.001) |
| Clusters | 289 | 289 | 289 |
| Observations | 1,320,654 | 1,320,654 | 1,323,189 |

Note: The dependent variable is described at the top of each column. In column (1), it is years of schooling; an indicator for completing any schooling in column (2); and in column (3), an indicator for completing primary school. Each regression includes a cubic for age, indicators for female and living in a rural area, birth year fixed effects, and region by survey wave fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. All samples include adults between the ages of 15 and 49; restricted to cohorts born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table A.6: Replication of Fortson (2011) Table 8 –
Effect of HIV on Years Behind Grade for Age: Ages 7 to 14

| | Full Sample | No Orphans | Testing Sample | No HIV+ Member |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Regional HIV Prevalence$_{ry}$ | 0.069*** | 0.073*** | 0.068*** | 0.070*** |
| × Post-1992 Cohort$_c$ | (0.012) | (0.013) | (0.017) | (0.019) |
| Sample | Original countries and survey waves (2001−2005) | Nonorphans; original countries and survey waves (2001−2005) | Nonorphans in HIV testing sample; original countries and survey waves (2001−2005) | Nonorphans in HH with no HIV-positive individual; original countries and survey waves (2001−2005) |
| Regions | 157 | 157 | 139 | 139 |
| N | 163,601 | 137,799 | 53,643 | 49,772 |

Note: The dependent variable is years behind grade for age. Each regression includes indicators for female and living in a rural area, birth year fixed effects, and region fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Nonorphan is defined as having both mother and father alive; columns (3) and (4) include only households within the HIV testing sample. All samples include observations between the ages of 7 and 14. Standard errors are clustered by region and shown in parentheses.

Table A.7: Replication of Fortson (2011) Table 7 –
Effect of HIV on Education Outcomes - Migration and Mortality (Age < 25)

| | Years of School | Any Schooling | Completed Primary School |
|---|---|---|---|
| | (1) | (2) | (3) |
| | A. Original Countries and Survey Waves | | |
| Regional HIV Prevalence$_{ry}$ | -0.038** | -0.003** | -0.005*** |
| × Post-1980 Cohort$_c$ | (0.015) | (0.001) | (0.002) |
| Clusters | 157 | 157 | 157 |
| Observations | 143,960 | 143,960 | 144,017 |
| | B. Extended Sample (2001−2017) | | |
| Regional HIV Prevalence$_{ry}$ | -0.042*** | -0.003** | -0.005*** |
| × Post-1980 Cohort$_c$ | (0.014) | (0.001) | (0.002) |
| Clusters | 289 | 289 | 289 |
| Observations | 650,001 | 650,001 | 650,773 |

Note: The dependent variable is described at the top of each column. In column (1), it is years of schooling; an indicator for completing any schooling in column (2); and in column (3), an indicator for completing primary school. In Panel A, the sample and specification are described in Table 7 of Fortson (2011). In Panel B, each regression includes a cubic for age, indicators for female and living in a rural area, birth year fixed effects, and region by survey wave fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. All samples include adults between the ages of 15 and 25; restricted to cohorts born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table A.8: Effect of HIV on Schooling and Child Labor: Ages 7 to 14

| | | In the Past Seven Days | | Not Working in Past Seven Days |
|---|---|---|---|---|
| | Years Behind Grade for Age | Work for Pay – Non-Family | Work for Family | Years Behind Grade for Age |
| | (1) | (2) | (3) | (4) |
| Regional HIV Prevalence$_{ry}$ × Post-1992 Cohort$_{c}$ | 0.052*** | -0.002** | -0.014*** | 0.058*** |
| | (0.012) | (0.001) | (0.002) | (0.014) |
| Mean of Dependent | 2.544 | 0.127 | 0.274 | 2.350 |
| Regions | 114 | 114 | 114 | 114 |
| N | 283,504 | 263,851 | 263,028 | 159,535 |

Note: The dependent variable in column (1) and (4) is years behind grade for age, in columns (2) and (3) the dependent variable is an indicator equal to one if doing the work described at the top of the column. Each regression includes a cubic for age, indicators for female and living in a rural area, birth year fixed effects, and region by survey wave fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Survey waves include: Burkina Faso (2010); Burundi (2010); Cameroon (2011); Democratic Republic of the Congo (2007); Cote d'Ivoire (2011); Ethiopia (2011); Gabon (2012); Guinea (2012); Liberia (2007); Malawi (2004); Mali (2001; 2006; 2012); Niger (2006; 2012); Rwanda (2010); Senegal (2005; 2010); Sierra Leona (2008; 2013). All samples include observations between the ages of 7 and 14. Standard errors are clustered by region and shown in parentheses.

Table A.9: Replication of Fortson (2011) Table 9: Differences by Sex

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | A. Dependent Variable: Years of School | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ × Male$_{iycr}$ | -0.083*** | -0.049*** | -0.047*** | -0.012 | -0.007 |
| | (0.014) | (0.014) | (0.015) | (0.010) | (0.011) |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.026 | -0.043* | -0.035* | -0.028*** | -0.031*** |
| | (0.018) | (0.023) | (0.019) | (0.010) | (0.010) |
| Clusters | 157 | 157 | 137 | 289 | 289 |
| Observations | 253,324 | 253,324 | 867,584 | 1,320,654 | 1,320,654 |
| | B. Dependent Variable: Any School (Years of School > 0) | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ × Male$_{iycr}$ | -0.009*** | -0.000 | 0.001 | 0.002 | 0.002 |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.001 | -0.006* | -0.009*** | -0.007*** | -0.007*** |
| | (0.002) | (0.003) | (0.003) | (0.002) | (0.002) |
| Clusters | 157 | 157 | 137 | 289 | 289 |
| Observations | 253,324 | 253,324 | 867,584 | 1,320,654 | 1,320,654 |
| | C. Dependent Variable: Completed Primary | | | | |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ × Male$_{iycr}$ | -0.009*** | -0.008*** | -0.007*** | -0.002 | -0.002 |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) |
| Regional HIV Prevalence$_{ry}$ × Post-1980 Cohort$_c$ | -0.005*** | -0.005*** | -0.003* | -0.000 | -0.001 |
| | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| Clusters | 157 | 157 | 137 | 289 | 289 |
| Observations | 253,513 | 253,513 | 868,232 | 1,323,189 | 1,323,189 |
| Sample | Original Countries and survey waves (2001 – 2005) | Original Countries and survey waves (2001 – 2005) | Original Countries; expanded survey waves (2001 – 2017) | Expanded Countries and survey waves (2001 – 2017) | Expanded Countries and survey waves (2001 – 2017) |
| Age (cubic) | — | — | X | X | X |
| Rural Indicator | X | X | X | X | X |
| HIV Prevalence × Male | — | X | X | X | — |
| Post-1980 Cohort × Male | — | X | X | X | — |
| Male Indicator | X | X | X | X | — |
| Birth Year F.E. | X | X | X | X | — |
| Birth Year F.E. × Male | — | — | — | — | X |
| Region F.E. | X | X | — | — | — |
| Region × Survey Wave F.E. | — | — | X | X | — |
| Region × Survey Wave F.E. × Male | — | — | — | — | X |

Note: The dependent variable is described at the top of each panel. In Panel A, it is years of schooling; an indicator for completing any schooling in Panel B; and in Panel C, an indicator for completing primary school. The set of controls used in each regression is detailed at the bottom of the table. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. The sample includes all observations between the ages of 15 and 49, born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table A.10: Effect of HIV on Education Outcomes

| | Men | | | Women | | |
|---|---|---|---|---|---|---|
| | A. Dependent Variable: Years of School | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Regional HIV Prevalence$_{ry}$ | -0.039*** | | | -0.029*** | | |
| × Post-1980 Cohort$_{c}$ | (0.013) | | | (0.010) | | |
| Max. Regional HIV Prevalence$_{r}$ | | -0.040*** | | | -0.030*** | |
| × Post-1980 Cohort$_{c}$ | | (0.013) | | | (0.010) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | -0.032** | | | -0.037** |
| | | | (0.013) | | | (0.017) |
| Regions | 289 | 289 | 254 | 289 | 289 | 254 |
| N | 619,910 | 619,910 | 914,767 | 700,744 | 700,744 | 1,024,316 |
| | B. Dependent Variable: Any Schooling (Years of School > 0) | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.006*** | | | -0.007*** | | |
| × Post-1980 Cohort$_{c}$ | (0.001) | | | (0.001) | | |
| Max. Regional HIV Prevalence$_{r}$ | | -0.006*** | | | -0.007*** | |
| × Post-1980 Cohort$_{c}$ | | (0.001) | | | (0.001) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | -0.005*** | | | -0.006*** |
| | | | (0.001) | | | (0.002) |
| Regions | 289 | 289 | 254 | 289 | 289 | 254 |
| N | 619,910 | 619,910 | 914,767 | 700,744 | 700,744 | 1,024,316 |
| | C. Dependent Variable: Primary Completed | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.002** | | | -0.000 | | |
| × Post-1980 Cohort$_{c}$ | (0.001) | | | (0.001) | | |
| Max. Regional HIV Prevalence$_{r}$ | | -0.003** | | | -0.001 | |
| × Post-1980 Cohort$_{c}$ | | (0.001) | | | (0.001) | |
| Regional HIV Prevalence at Age Six$_{rc}$ | | | -0.003** | | | -0.002 |
| | | | (0.001) | | | (0.001) |
| Regions | 289 | 289 | 254 | 289 | 289 | 254 |
| N | 620,938 | 620,938 | 916,223 | 702,251 | 702,251 | 1,026,027 |

Note: The dependent variable is described at the top of each panel. In Panel A, it is years of schooling; an indicator for completing any schooling in Panel B; and in Panel C, an indicator for completing primary school. Each regression includes a cubic for age, an indicator for living in a rural area, and birth year and region by survey wave fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. Max. regional HIV prevalence is a time-consistent measure of the highest recorded prevalence rate in the region across all survey years, and regional HIV prevalence at age six is a cohort specific estimate of the regional prevalence rate for each cohort's age six year. All samples include adults between the ages of 15 and 49, born in or after 1965. Standard errors are clustered by region and shown in parentheses.

Table A.11: Effect of HIV on Education Outcomes – Removing Regions with Highest and Lowest HIV Prevalence

| | Full | Only Include Regions In Following Percentile Range | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample | > 1 | > 5 | > 10 | > 25 | < 99 | < 95 | < 90 | < 75 | > 1; < 99 | > 5; < 95 | > 10; < 90 | > 25; < 75 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| | | | | | | A. Dependent Variable: Years of School | | | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.033*** | -0.033*** | -0.035*** | -0.033*** | -0.027* | -0.037*** | -0.041*** | -0.035* | -0.074** | -0.038*** | -0.046*** | -0.037 | -0.077 |
| × Post-1980 Cohort$_c$ | (0.011) | (0.011) | (0.011) | (0.012) | (0.015) | (0.011) | (0.015) | (0.021) | (0.037) | (0.012) | (0.015) | (0.024) | (0.065) |
| Regions | 289 | 285 | 279 | 268 | 235 | 286 | 271 | 253 | 211 | 282 | 261 | 232 | 157 |
| N | 1,320,654 | 1,308,567 | 1,271,269 | 1,195,979 | 1,010,660 | 1,304,139 | 1,252,069 | 1,170,436 | 968,849 | 1,292,052 | 1,202,684 | 1,045,761 | 658,855 |
| | | | | | | B. Dependent Variable: Any Schooling (Years of School > 0) | | | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.007*** | -0.007*** | -0.007*** | -0.006*** | -0.006*** | -0.007*** | -0.008*** | -0.008*** | -0.018*** | -0.007*** | -0.009*** | -0.009*** | -0.020*** |
| × Post-1980 Cohort$_c$ | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) | (0.002) | (0.003) | (0.004) | (0.001) | (0.002) | (0.003) | (0.007) |
| Regions | 289 | 285 | 279 | 268 | 235 | 286 | 271 | 253 | 211 | 282 | 261 | 232 | 157 |
| N | 1,320,654 | 1,308,567 | 1,271,269 | 1,195,979 | 1,010,660 | 1,304,139 | 1,252,069 | 1,170,436 | 968,849 | 1,292,052 | 1,202,684 | 1,045,761 | 658,855 |
| | | | | | | C. Dependent Variable: Primary Completed | | | | | | | |
| Regional HIV Prevalence$_{ry}$ | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.002* | -0.003** | -0.004** | -0.012*** | -0.002* | -0.003** | -0.004** | -0.015*** |
| × Post-1980 Cohort$_c$ | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.003) | (0.001) | (0.001) | (0.002) | (0.005) |
| Regions | 289 | 285 | 279 | 268 | 235 | 286 | 271 | 253 | 211 | 282 | 261 | 232 | 157 |
| N | 1,323,189 | 1,311,096 | 1,273,737 | 1,198,322 | 1,012,846 | 1,306,673 | 1,254,441 | 1,172,537 | 970,271 | 1,294,580 | 1,204,989 | 1,047,670 | 659,928 |

Note: The dependent variable is described at the top of each panel. In Panel A, it is years of schooling; an indicator for completing any schooling in Panel B; and in Panel C, an indicator for completing primary school. Each regression includes indicators for female and living in a rural area, birth year fixed effects, and wave specific region fixed effects. Regional HIV prevalence is the regional prevalence rate in the survey year in which each individual was observed. All samples include adults between the ages of 15 and 49, born in or after 1965. Standard errors are clustered by region and shown in parentheses.