MORTALITY RISK AND HUMAN CAPITAL INVESTMENT: THE IMPACT OF HIV/AIDS IN SUB-SAHARAN AFRICA

Jane G. Fortson*
Becker Center on Chicago Price Theory
University of Chicago

This Version: September 2007
First Draft: September 2006

Abstract

Over the past several decades, the HIV/AIDS epidemic has dramatically altered patterns of morbidity and mortality in sub-Saharan Africa, with consequences for human capital investment and economic growth. Previous studies of orphans in Africa have established that losing a parent is associated with worse educational outcomes; however, HIV/AIDS could influence human capital investment more broadly. Using data from Demographic and Health Surveys for seven countries in sub-Saharan Africa, I estimate the effect of local HIV prevalence on individual human capital investment. I find that the HIV/AIDS epidemic has indeed affected human capital investment quite broadly: living in an area with higher HIV prevalence is associated with lower levels of completed schooling and slower progress through school, even among non-orphans. Robustness checks suggest that these results are not driven by omitted variable bias or by differential migration or differential survival by educational level. These results are consistent with a model of human capital investment in which parents and children respond to changes in the expected return to schooling driven by mortality risk.

* *Email:* jfortson@uchicago.edu
Address: Becker Center on Chicago Price Theory, 5807 S. Woodlawn Avenue, Chicago, IL 60637

I am very grateful to Gary Becker, Sandy Black, Anne Case, Erica Field, Ken Fortson, Jon Guryan, Alan Krueger, Steve Levitt, John List, Giovanni Mastrobuoni, Emily Oster, Chris Paxson, Jesse Rothstein, and Ceci Rouse for helpful discussions; Silvia Barcellos, Marianne Bertrand, Leandro Carvalho, Melissa Clark, Angus Deaton, Wiola Dziuda, Molly Fifer, Marco González-Navarro, Luojia Hu, Adriana Lleras-Muney, Marlaine Lockheed, Jesse Shapiro, Jonathan Robinson, Ian Walker, Susan Yeh, two anonymous referees, and seminar participants at Princeton University, the University of Chicago, Mathematica Policy Research, and the Georgetown Public Policy Institute for their suggestions; Bridgette James and Martin Wulfe for help with the data; and Lizzie DeYoung for research assistance.

I. INTRODUCTION

Twenty-five years after AIDS was first reported by the CDC in 1981, 38.6 million people worldwide are estimated to be infected with HIV. Sub-Saharan Africa, the "global epicenter of the AIDS pandemic," is home to over 60 percent of HIV-infected people, with prevalence among adults ages 15-49 estimated to be 6.1 percent (UNAIDS, 2006). AIDS killed some two million Africans in 2005, and has reduced life expectancies in the most affected countries to their pre-1970 levels (US Census Bureau International Data Base, 2005). And yet the impact of the AIDS epidemic extends far beyond its first-order effect on longevity. In some areas, AIDS has wiped out a substantial fraction of prime-age adults, threatening economic stability and growth today and in the coming decades.

The full impact of HIV/AIDS has not yet been realized, and available empirical evidence on the effect of HIV on economic growth is mixed. Bloom and Mahal (1997) find that, as of the mid-1990s, the AIDS epidemic had an insignificant effect on GDP growth. Bonnel (2000) disputes this, showing that HIV prevalence is negatively related to growth. Recent work by Werker, Ahuja, and Wendell (2006) using data through 2002 finds no impact of the epidemic on growth, savings, or fertility.

Several recent studies forecast the long-term effects of the AIDS epidemic by jointly analyzing fertility and human capital. In particular, Young (2005), in an analysis of South Africa, argues that AIDS-induced fertility declines, by reducing the number of people among whom limited resources must be shared, will offset AIDS-related reductions in human capital and actually improve living standards of future generations. This is at odds with Kalemli-Ozcan (2006), who suggests that fertility may rise in response to the HIV/AIDS epidemic, implying an aggregate decline in welfare. Both Young (2005) and Kalemli-Ozcan

(2006) assume that human capital investment has fallen in response to HIV/AIDS, but provide only limited empirical support.[1]

Most existing estimates of the effect of HIV on human capital investment have focused on orphans. Orphans in Africa are significantly less likely to be enrolled in school than non-orphans, and progress more slowly when enrolled (e.g., Case, Paxson, and Ableidinger, 2004; Case and Ardington, 2006; Evans and Miguel, 2007). However, the HIV/AIDS epidemic may well affect human capital investment more broadly. For instance, changes in mortality could have a non-negligible effect on human capital investment (e.g., Meltzer, 1992). Because the presence of HIV/AIDS may increase mortality risk in adulthood, it could *ceteris paribus* reduce the expected return to schooling. To the extent that human capital investment responds to changes in the expected return to schooling, we might think that human capital investment would fall among both orphans and non-orphans.[2] Through this and other channels, the HIV/AIDS epidemic may have a widespread effect on human capital investment and, therefore, growth.

The contribution of this paper is twofold. First, I provide estimates of the overall impact of the HIV/AIDS epidemic on educational attainment; to my knowledge, these are the first estimates of this effect. Second, I evaluate empirically several competing hypotheses for why HIV might affect human capital investment. These findings can inform public

[1] Young (2005) relies on previous estimates of schooling declines due to orphanhood, overlooking other possible channels. Kalemli-Ozcan (2006) uses country-level data to show that HIV incidence is negatively related to school enrollment. As Kalemli-Ozcan (2006) cautions, school enrollment rates have serious limitations as measures of human capital investment. Not only are they fraught with reporting errors, but enrollment rates are a poor proxy for educational attainment, as progress through school is extremely heterogeneous in Africa. Furthermore, analyzing the link between HIV and education at the country level is potentially problematic, both because it is difficult to separate the effect of the country's HIV rate from country-specific shocks and because HIV rates can vary tremendously within country. Lastly, as Kalemli-Ozcan (2006) concedes, country-level analysis cannot distinguish between possible channels through which HIV could influence schooling.

[2] There are indeed numerous other channels through which HIV could influence schooling. A related hypothesis, pursued by Kalemli-Ozcan (2006), is that HIV jointly influences fertility and human capital decisions, and that, as a form of insurance, parents may have more children but invest less in their education (a "quality-quantity trade-off").

policies aimed at minimizing the devastating effects of the HIV/AIDS epidemic, and may generalize to other settings as well.

In my principal empirical tests, I use nationally-representative cross-sections from Demographic and Health Surveys (DHS) for Burkina Faso, Cameroon, Ghana, Kenya, Mali, Tanzania, and Zambia to estimate the relationship between local HIV prevalence and the educational outcomes of successive birth cohorts. These surveys include newly-available estimates of HIV based on household testing – arguably the most accurate and geographically-specific estimates available. Merging these cross-sectional data on local HIV prevalence with assumptions about the time path of HIV, my analysis exploits the substantial geographic and time variation in HIV prevalence to estimate the effect of HIV on human capital investment. I find that local HIV prevalence at school-going age is negatively associated with years of schooling, school attendance, primary school completion, and progress through school. Children living in areas with HIV prevalence of 10 percent complete about 0.5 fewer years of schooling than children living in areas without HIV.

Before investigating possible explanations for these results, I evaluate several threats to validity. First, I perform a series of tests to assess whether my estimates are confounded by omitted correlates of HIV and schooling. In one such test, I show that high- and low-HIV areas did not have differing educational trends prior to the HIV epidemic, supporting a causal interpretation of the subsequent divergence. Separately, I test whether attrition through migration and mortality might influence my results by constructing a sample of young adults from multiple waves of the DHS, rather than drawing a sample of respondents of many ages from a single cross-section. Young adults are less likely to have died and may be less likely to have migrated from their childhood region, and hence should be more

representative of those who grew up in a particular area. I find that results incorporating these adjustments confirm that local HIV prevalence reduces human capital investment.

Lastly, I examine possible explanations for these results. Building a model of human capital investment that incorporates lifetime uncertainty, I demonstrate that mortality risk reduces the value of the incremental increase in earnings from an additional year of schooling. Under credible assumptions, the model suggests that completed schooling is inversely related to mortality risk. Employing estimates from past literature, I argue that the model implies that local HIV prevalence of 10 percent should reduce completed years of schooling by about 0.3 years, a slightly smaller decline than I estimate empirically.

In addition to the mortality risk channel explored by the model, I evaluate five possible channels through which the epidemic could affect schooling: orphanhood, caretaking requirements, the provision of schooling, family resources, and market wages. For instance, it could be that the local HIV rate is negatively associated with schooling because a disproportionate number of orphans live in high-HIV areas. I test whether my results are robust to the exclusion of orphans, and find that local HIV prevalence is associated with substantial declines in human capital investment even among non-orphans. I also test whether these results are driven by the increased care-taking requirements that might come from AIDS-related illness. Specifically, using individual HIV test results available in these DHS data, I exclude from the analysis children living in households known to include an HIV-positive member. I find that local HIV prevalence is negatively and significantly associated with educational outcomes in this restricted sample as well. To test whether the provision of schooling or family resources drive the effect, I test for differences between men and women in the effect of HIV on education. Because mortality due to HIV early in the epidemic was disproportionately concentrated among men (UNAIDS and

WHO, 1999; Oster, 2006b), we would expect to find a larger effect among men if expectations about mortality risk contribute substantially to the estimated effect. Provision and family resources, on the other hand, would in principle affect schooling among boys and girls equally.[3] I find significantly larger effects of HIV on educational outcomes for men, suggesting that this effect does not primarily operate through the provision of schooling or family resources.

Lastly, I test whether children leave school in response to market conditions. We might expect wages to increase in response to AIDS-related mortality, inducing children to leave school and enter the workforce. But if this were the case, we would not expect to find different effects of HIV across subgroups facing the same market conditions. While men and women may not be substitutes in the labor market, Muslims and non-Muslims arguably face the same market conditions. However, I find substantially larger HIV-related declines in educational attainment among non-Muslims, inconsistent with the effect being driven by market wages. These results provide evidence in support of a broad effect of HIV on human capital investment. Though I cannot definitively rule out other channels, the results are generally consistent with what we would expect to see if parents respond to declines in the expected return to schooling, here driven by mortality risk.

II. DATA

My empirical analysis uses nationally-representative cross-sections from Demographic and Health Surveys (DHS) for seven countries in sub-Saharan Africa: Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003),

[3] Under certain conditions, we might expect family resources to affect boys and girls differently. But in such situations, we would expect girls' schooling to fall by more than boys' schooling, contrary to what we find.

and Zambia (2001-2002).[4] In addition to providing demographic, economic, and fertility microdata, these waves of the DHS are linked to HIV test results which can be used to calculate local HIV prevalence.

HIV Prevalence.

In these seven DHS surveys, some adult respondents to the survey were asked to provide blood samples for HIV testing.[5] I calculate local HIV prevalence as the adult HIV rate in each sector of residence (urban or rural) within each region (generally defined by a government administrative division). Table I.A shows the considerable variability in local HIV prevalence across and within countries.

Estimates of HIV prevalence from other sources are generally based on the testing of sub-populations (e.g., pregnant women or commercial sex workers). Because DHS HIV testing was conducted at the household level, respondents should be more representative of area residents than other HIV testing samples. Not only are DHS HIV testing data representative, but they provide estimates of prevalence at a fine level of geographic detail. National HIV prevalence data, in contrast, obscure the substantial geographic variation within country.

However, local HIV rates could be measured with error. For about 17 percent of respondents, HIV test results are missing – primarily because of test refusal. Because HIV status may well be related to HIV test refusal and because data are missing at different rates across sectors and regions, missing HIV test data could bias my estimates of local HIV prevalence. However, the magnitude of any bias would most likely be quite small: Mishra et

[4] Source: DHS datasets, www.measuredhs.com, MEASURE DHS, Macro International Inc.
[5] The data appendix provides detail on sample selection and testing procedures.

al. (2006) show that such non-response has an insignificant effect on national HIV prevalence estimates using these data.

All the same, local HIV prevalence data have their limitations. Sampling variability could introduce classical measurement error, attenuating my estimates. More importantly, estimates of local HIV prevalence over time are unavailable. My estimates of local HIV prevalence from DHS testing are applicable only to the survey year. To address this shortcoming, I estimate local HIV prevalence over time by combining survey year estimates with several different time paths. In particular, I model past local HIV rates using local rates in the survey year and information about the spread of HIV over time more generally. I consider four time paths: a step function, a linear function, a nonlinear function, and a flexible time path, all of which are described in greater detail in Section III.

Individual Characteristics.

Data on age, educational attainment, and other individual characteristics are drawn from responses to the individual and household questionnaires of the DHS. For the most part, my analysis uses a sample of adults ages 15-49. I consider three educational outcomes for adults: years of schooling, school attendance (completed one or more years), and primary school completion. In Section V, I use a sample of children ages 7-14 to explore explanations for my main results. Because schooling for children may be incomplete, I use progress through school (years / (age – 6)) as an additional measure of schooling attainment for this younger group.

Child characteristics, including progress through school, are summarized in Table I.B while summary statistics for adult characteristics, including all three measures of education, are presented in Table I.C. (Survey sampling procedures, response rates, and variable definitions are described in detail in the data appendix.) In the countries under study, two-

thirds of adults and three-quarters of children reside in rural areas. On average, adults have completed 4.8 years of schooling, and two-thirds report having completed at least one year of school. Forty-three percent of adults have at least a primary school education.

Assignment of adults and children to birth cohorts is based on reported age. Because there is evidence that age may be misreported in DHS surveys (e.g., National Council for Population Development and Institute for Resource Development/Macro Systems, Inc., 1989), I address possible biases due to age misreporting in Section IV.

My data include information about the respondent's current region and sector of residence, but only limited information about his childhood residence. Because my ability to identify the effect of HIV on human capital investment depends heavily on my ability to match individuals to their childhood local HIV rates, migration (and mortality, for that matter) could be a source of bias. For simplicity, I assume that individuals have not moved from their childhood region/sectors of residence. I test the validity of this assumption in Section IV.

III. EMPIRICAL MODEL

My approach exploits geographic and time variation in HIV prevalence. Using fixed effects for region/sector of residence, I estimate the effect of local HIV prevalence on three educational outcomes: completed years of schooling, school attendance, and primary school completion.[6] In particular, I estimate the following model:

$$S_{icr} = \beta_0 + \beta_1 HIV_{cr} + \beta_2 F_{icr} + \gamma_c + \alpha_r + \varepsilon_{icr} \qquad (1)$$

where S_{icr} = educational outcome for respondent i in cohort c and region/sector r,

[6] I use these three educational outcomes to limit concern that incomplete schooling could bias my results. However, only 13 percent of adults ages 25-49 have more than 10 years of schooling, so any bias should be small. As an additional test, I top-code years of schooling at 10; the results are quite similar. Using a number of outcomes should also assuage concerns that my estimates are sensitive to transformations of the dependent variable (Meyer, 1995).

HIV_{cr} = HIV prevalence in region/sector r when cohort c was in childhood ,

F_{icr} = indicator for whether respondent i in cohort c in region/sector r is female ,

γ_c = fixed effect for year of birth c , and

α_r = fixed effect for region/sector r .

γ_c allows for a flexible trend in educational outcomes over time, and α_r accounts for underlying differences in educational outcomes across region/sectors of residence. My sample includes adults ages 15-49 in the survey year, and is weighted using DHS-provided weights that account for the sampling and response probabilities. I compute Huber-White standard errors that adjust for possible correlations of outcomes within each region/sector.[7]

I assume that parents make investment decisions for their children based on local HIV prevalence when children are age 10. (In results not shown, I test the sensitivity of my results to different assumptions about the relevant age. Local HIV prevalence at age 5 and local HIV prevalence at age 15, using any of the time paths, have negative and significant effects on educational outcomes.)

Though the DHS HIV testing component provides HIV status only in the survey year, I construct estimates of local HIV prevalence over time by making a variety of assumptions about the time path of HIV. I generate time variation in the local HIV rate using:

$$HIV_{cr} = HIV_r \times f(c+10),$$

where c = year of birth and $f(c+10)$ determines the time path of HIV, which varies across specifications.

[7] If there is serial correlation in the outcome variable – which is quite possible for schooling – conventional difference-in-differences techniques tend to over-reject a null hypothesis of no effect. While I do not adopt a true difference-in-differences approach, my standard errors could still be plagued by this problem. Clustering that accounts for intraregional auto-correlation has been found to be an effective solution in some applications when the number of regions is sufficiently large (Bertrand, Duflo, and Mullainathan, 2004).

First, I adopt a difference-in-differences framework. In particular, I assume that HIV spread instantaneously in 1990 and remained constant thereafter, so that $f(t)$ is a step function with a discontinuity at 1990 (i.e., $f(t) = I(t \geq 1990)$, or $f(c+10) = I(c \geq 1980)$). This amounts to assuming that HIV had no effect on the educational outcomes of cohorts born before 1980, and a constant effect on cohorts born in or after 1980.[8]

My sample of adults ages 15-49 covers birth cohorts 1951-1989. Respondents in the unaffected (1951-1979) cohorts would have, for the most part, completed their schooling by the late 1980s, before HIV was widespread. Respondents in the affected (1980-1989) cohorts reached school-going age in the late 1980s and 1990s, when HIV was more prevalent.

Table II shows results from this specification. When the time path of HIV is assumed to follow a step function, HIV has a negative and significant impact on all three educational outcomes. Relative to a base case of no HIV, local HIV prevalence of 10 percent is associated with about 0.6 fewer years of schooling, a 4 percentage point lower probability of attending school, and a 7 percentage point lower probability of completing primary school. These results suggest that educational outcomes improved less over time in areas harder hit by HIV.

Though intuitively appealing, the discontinuous time path used in Table II is somewhat unrealistic, and could be sensitive to my choice of cohorts. Because HIV has, in general, become more prevalent over time, a linear time path may be a more reasonable

[8] In this difference-in-differences framework, I compare differences in educational outcomes of unaffected and affected cohorts across regions with differing levels of treatment intensity (HIV prevalence). Region/sector fixed effects control for underlying differences in educational outcomes across regions. Unlike a simple difference-in-differences approach, I allow for a flexible time trend in schooling outcomes and control for sex. Difference-in-differences results without controls (not shown) are quite similar to these results.

approximation of the actual time path. In Table III, I assume that local HIV rates grew linearly over time since 1980, when they were zero, so that

$$f(t) = \begin{cases} 0 & \text{if } t \leq 1980 \\ \frac{t-1980}{t_s-1980} & \text{if } t > 1980 \end{cases} \quad \text{where } t_s = \text{survey year}.$$

The results show that the effect of local HIV prevalence is negative and significant for all three measures of education. The point estimates in Table III are about twice as large as those in Table II, and imply that a local HIV rate of 10 percent is associated with about 1.2 fewer years of schooling.[9]

In Tables II and III, I assume that HIV followed a relatively simple time path. While HIV prevalence trends at the local level are unknown, national HIV prevalence trends from UNAIDS suggest that HIV spread neither instantaneously nor linearly. To better mimic the true spread of the disease, I choose $f(t) = \frac{HIV_{tn}}{HIV_{t_s n}}$ where $t_s = \text{survey year}$ and n indexes the nation. In particular, I assume that the time path of local HIV prevalence follows the time path of national HIV prevalence (which is nonlinear and varies by country). This method assumes that local HIV prevalence and national HIV prevalence are in fixed proportion over time.

Table IV shows estimates of equation (1) using this nonlinear time path. These results imply that, relative to growing up in an area without HIV, growing up in an area with HIV prevalence of 10 percent is associated with 0.5 fewer years of schooling, a 4 percentage point lower probability of school attendance, and a 7 percentage point lower probability of primary school completion. Because levels of completed schooling are low in these countries, these effects are in fact quite large: local HIV prevalence of 10 percent in

[9] It is not altogether surprising that the linear time path implies a large effect of HIV prevalence at age 10 on schooling. This path assumes lower levels of HIV prevalence between 1990 and the survey year than assumed by the step function, attributing the change over this period to lower levels of HIV.

childhood implies an 11 percent reduction in completed years of schooling, a 6 percent reduction in the probability of attending school, and a 16 percent reduction in the probability of completing primary school.

The step, linear, and nonlinear time paths cover a range of assumptions about how HIV prevalence has changed over time, and yet all show that local HIV prevalence is associated with declines in human capital investment. However, to the extent that these differ from the true time path, the magnitude of these estimates could be biased. To test the robustness of these results, I consider a more flexible time path, $f(t) = \sum_{j \neq 1970} I(t = j)$, which is a set of year of birth indicators. I estimate equation (1) using this time path on a sample of adults born between 1954 and 1986. This specification produces not a single estimate β_1 of the effect of HIV but a series of estimates, β_{1c} – one for each birth cohort. In Figure 1, I plot these estimates by year of birth for each of the three educational outcomes, relative to the effect on the 1970 cohort. In principle, HIV should have no effect on early cohorts and an increasingly large negative effect on later cohorts. These figures show that the effect of HIV on schooling varies with year of birth in a manner consistent with my hypothesis and with the results in Tables II, III, and IV.

The flexible time path in Figure 1 has the advantage that it requires few assumptions about how HIV has changed over time. However, because it produces a series of estimates, the results are more difficult to interpret than those in Tables II, III, and IV. In Figure 2, I plot the results from Figure 1 and the implied effects from Tables II, III, and IV. In particular, I show $\hat{\beta}_1 \times f(c+10)$ by birth cohort for each of the four time paths (step, linear, nonlinear, and flexible). Comparing the estimated effects from the four time paths, we can see that the nonlinear time path (which forces local HIV to follow national HIV trends) is a

reasonably good approximation to the flexible time path.[10] In fact, the effects from the flexible time path are, if anything, larger in magnitude than those implied by the nonlinear time path results, suggesting that estimates from the nonlinear path may be a lower bound for the true effect.[11]

IV. THREATS TO VALIDITY

Endogenous and Omitted Regressors.

In Section III, I present evidence of a strong negative association between local HIV prevalence and human capital investment. However, it is possible that this association is driven by omitted correlates of both HIV and schooling, or by reverse causality.

There could be region/sector of residence level shocks that affect both education and HIV transmission, driving my results. If this were the case, different region/sectors would be subject to different shocks over time. This implies that in all likelihood educational trends would evolve differently across region/sectors. As a robustness check, I test whether there are differences in educational outcome trends between areas with different levels of HIV *before* HIV became very prevalent (i.e., among adults born before 1980). In principle, if it is HIV – rather than an omitted correlate – that causes educational outcome trends to differ across areas, there should be no relationship between the eventual HIV rate and the trends among early birth cohorts. Indeed, these results, shown in Table V, are consistent with my identifying assumption – that in the absence of HIV, trends in educational outcomes would not be significantly different across region/sectors.

[10] Differences between these need not reflect incorrect assumptions about the time path. It could be, for instance, that HIV has a nonlinear effect on schooling, producing an inconsistency between the flexible and nonlinear specifications, even if the nonlinear specification is an accurate representation of the time path.
[11] Separately, I test whether my results hold when using an alternate source of HIV prevalence data over time. Oster (2007) calculates estimates of national HIV prevalence using sibling histories from the DHS. Results from Burkina Faso, Cameroon, Kenya, Mali, and Zambia using these national prevalence estimates (in place of the region/sector estimates) in the same regression framework are similar in magnitude to the estimates from Tables II, III, and IV.

While this provides assurance that region/sectors did not otherwise change differently over time, I still might be concerned that the factors that make one region/sector susceptible to HIV might also have an impact on education. In particular, areas exposed to economic shocks could see both declines in educational attainment and increases in HIV. Using the 1991 wave from Cameroon, the 1992 wave from Zambia, and the 1993 waves from Ghana and Kenya (in regions that are comparable over time), I test whether household assets and amenities are related to the local HIV rate in 2003/2004. I find that the HIV rate is not significantly related to household asset ownership, which suggests that economic conditions were not a major factor driving increases in HIV prevalence.[12]

Reverse causality could also confound my estimates. While the observed relationship between the local HIV rate and educational outcomes is not likely to be driven by a direct effect of educational outcomes (of children) on the local HIV rate (of adults), the observed effect could in theory reflect the omission of parental schooling. We might think that parental schooling is positively correlated with children's schooling, and that parents' schooling is related to HIV status, which in turn affects local HIV prevalence. Fortson (2007) shows that HIV infection is positively related to educational attainment among adults. Therefore, to the extent that region/sector fixed effects do not fully account for parental

[12] In particular, I use a composite measure of household assets and amenities that is the fraction of eight assets or amenities (radio, television, refrigerator, bicycle, motorcycle, car, electricity, and a flush toilet or pit latrine) that the respondent's household has. I perform this analysis in a subset of region/sectors in Cameroon, Ghana, Kenya, and Zambia because these DHS surveys were conducted in the early 1990s, early in the HIV/AIDS epidemic. Regions in Burkina Faso's 1992/1993 survey and Tanzania's 1992 survey were not comparable to those in the current wave; the Mali 1987 data are not comparable to the more recent surveys. While assets may be an imperfect measure of current economic conditions, they are perhaps the best measure available in the DHS. In my main specification, I do not control for economic conditions because I have very few observations per region/sector. However, region/sector fixed effects should pick up the level effect of economic conditions.

schooling, my estimates will be positively biased by the omission of parental schooling, leading me to understate the magnitude of the effect.[13]

Migration and Mortality.

My estimates of the effect of local HIV prevalence on human capital investment are based on the observed educational outcomes of adults living in a given sector (urban or rural) within a region, and tying trends in the outcomes of successive birth cohorts to HIV prevalence in that region/sector. However, it is possible that the population of adults currently living in a particular region/sector is not representative of the population that lived in that region/sector in childhood, either because of migration or mortality. For instance, because individual HIV status is related to education (Fortson, 2007), differential survival could bias my estimates of the effect of local HIV prevalence on human capital investment.[14]

To test whether non-random migration or survival might influence my estimates, I perform a series of sensitivity checks. First, rather than using adults ages 15-49 from the most recent wave of the DHS, I construct a time-series of birth cohorts by drawing adults ages 15-25 from each of multiple waves of the DHS. This sample has the advantage that

[13] To confirm that my estimates are not driven by omitted variable bias, I would ideally identify a source of exogenous variation in the local HIV rate, and exploit this variation to estimate the effect of HIV absent omitted variable bias. The spread of HIV has been influenced by a number of factors, including the structure of sexual networks, the prevalence of other sexually transmitted diseases, and proximity to the geographic origin of HIV (Iliffe, 2006; Oster, 2005; Oster, 2006a). Because male circumcision has recently been shown to reduce the risk of HIV infection (Auvert et al., 2005; Wise, 2006), one potential candidate is the local male circumcision rate. There is precedent for using the male circumcision rate as an instrument for HIV prevalence (e.g., Werker, Ahuja, and Wendell, 2006; Oster, 2006b). Circumcision rates are plausibly unrelated to changes in educational attainment over time, except through their effect on HIV. Using self-reports of circumcision from the DHS from Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania, I calculate local male circumcision rates in each region/sector. Using a nonlinear time path for both HIV and circumcision, I find that the local circumcision rate is negatively and significantly related to the local HIV rate (not shown). However, the instrument is too weak to provide sufficient precision, and the two-stage least squares estimates are insignificant. Nevertheless, these results are not inconsistent with the fixed effects estimates.

[14] AIDS mortality will primarily affect early cohorts (Hosegood, Vanneste, and Timæus, 2004). Because the educated are disproportionately affected by HIV (Fortson, 2007), I am likely underestimating the educational outcomes of early cohorts in high-HIV areas, and therefore understating the negative effect of HIV on educational outcomes. However, if more educated people from younger cohorts (relative to those from older cohorts) are more likely to migrate away from high-HIV areas, then I might be overstating the effect.

young adults should be less affected by migration and mortality, and perhaps represent a more accurate time-series of birth cohorts.[15] However, region/sectors are not always comparable across waves of the DHS, and can only be compared in 100 of 158 region/sectors (for information about regions and years used, see the data appendix).

I assume a nonlinear time path for HIV, as in Table IV. Table VI, column (2) shows estimates from this robustness check. HIV has a negative and significant effect on all three educational outcomes (years of schooling, attendance, and primary school completion). The coefficients are similar in magnitude to those in columns (1) and (3), which use only the most recent cross-sections of the DHS. Column (1) duplicates the fixed effects results from Table IV, and column (3) shows fixed effects results from the latest wave using adults living in regions that are comparable across waves.

As an additional test of whether migration may be problematic for my results, I use information about the respondent's childhood sector of residence (when available). Because I do not know the respondent's childhood region of residence (or country), I assume that migration occurred only within region. I then use a nonlinear time path for HIV to estimate the effect of local HIV prevalence in the respondent's childhood sector (urban or rural) and current region on his educational outcomes.[16]

---

[15] However, migration among young adults is nevertheless quite high. Among men and women administered an individual questionnaire in the most recent wave (in all countries but Tanzania), only 46 percent of 15- to 25-year-olds say that they have always lived in the same place. Among the 54 percent who have moved, a majority (69 percent) moved after age 14. To the extent that migration is related to HIV prevalence, this could bias my estimates of the effect of HIV. I find that 15- to 25-year-olds who have moved are more likely to be educated and are more likely to live in high-HIV areas. Therefore, we can infer that the migration of young adults may attenuate my estimates, if anything. As an additional check, I construct a sample of children ages 7 to 10 from multiple waves; among 15- to 25-year-olds who moved, over 88 percent last moved after age 10. I find that HIV (using a linear time path) is negatively and significantly related to progress through school in this sample as well.

[16] Note that in this sensitivity check I do not adjust my estimates of local HIV prevalence. That is, while I may assign respondents the local HIV rate in a different region/sector, I do not recalculate the local HIV rate. Because respondents were, for the most part, not HIV positive in childhood, assigning their adult HIV status to their childhood sector does not seem sensible.

Table VI, column (4) shows that, using childhood sector, the local HIV rate is negatively and significantly associated with all three educational outcomes, and that the estimated effects are in fact slightly larger in magnitude than the comparison results.

However, these results have the disadvantage that they may still be biased by sample attrition due to mortality. On the other hand, unlike the robustness check in column (2) of Table VI, this approach may better adjust for sample selection bias due to childhood migration in response to the HIV/AIDS epidemic. For example, if more educated parents sent their children (who ultimately completed more schooling) to live in lower HIV areas, I might be overstating the effect of local HIV prevalence on educational outcomes. Assigning respondents to their childhood sector of residence should minimize this bias. Together, these robustness checks suggest that my results in Section III are not driven by migration or mortality.

Age Misreporting.

Previous studies have noted considerable age misreporting in DHS surveys (National Council for Population Development and Institute for Resource Development/Macro Systems, Inc., 1989). In particular, respondents have a tendency to report ages that end in zeros or fives, resulting in age heaping. Because I estimate the effect of HIV by looking at differences across birth cohorts, age misreports may influence my results. To test whether age heaping affects my estimates, I estimate the fraction of sample adults whose ages end in zeros or fives. About 26 percent of reported ages ended in zeros or fives, compared to the 20 percent that would be predicted if all ages were equally likely.

While this suggests that age heaping affected only a relatively small proportion of respondents, I investigate whether this biases my results by testing whether the fraction reporting ages ending in zeros or fives in a region/sector is related to the local HIV rate in

the survey year. I find that areas with higher levels of HIV have a lower fraction of respondents reporting ages that end in zeros or fives, even after controlling for average educational attainment. Looking at a histogram of reported ages, age misreports appear to be primarily underestimates of age for older cohorts and overestimates of age for younger cohorts. Because educational attainment has risen over time, I may have underestimated the educational gains over time in low-HIV areas relative to the gains in high-HIV areas, which would positively bias my estimates of the effect of HIV, making it less likely for me to detect a negative effect. Nevertheless, the magnitude of this bias is likely to be small. First, the relationship between age heaping and the local HIV rate appears to be driven by a handful of outliers – the relationship disappears in a median regression with controls. Second, because age is positively correlated with the probability of reporting an age ending in a zero or five, I can test whether my results are robust to the addition of controls for age. The adjusted fixed effects estimates are similar in magnitude to those in Table IV.

Additional Concerns.

Economic shocks or civil conflict could affect both local HIV rates and educational outcomes. In an extreme situation, this could lead to spurious correlation in one or two countries that is substantial enough to drive the pooled result. Though there are at most 42 region/sectors within one country, I estimate the effect of HIV on educational outcomes separately for each country. Country-by-country estimates show that HIV is negatively related to completed years of schooling in all countries, and the effect is significant in Cameroon and Zambia. The fact that these results are consistent across a very diverse set of countries suggests that my findings are likely to generalize to other African nations as well.

A separate issue is the timing of the surveys. In my estimates in Table II, I use an estimate of local HIV prevalence in the survey year—2001 in Mali, 2001/2002 in Zambia,

2003 in Burkina Faso, Ghana, Kenya, and Tanzania, and 2004 in Cameroon—interacted with a step function. If HIV prevalence changed sizeably between 2001 and 2004, this could bias my results. To investigate, I calculate predicted HIV prevalence in the region/sector in 2004 (see data appendix), and re-estimate equation (1) for the step time path. The revised fixed effects estimates are negative and significant, and are in fact somewhat larger in magnitude than the estimates in Table II.

In addition, effects of education on HIV infection at the individual level could bias my estimates of past HIV rates. Because my estimates of local HIV prevalence in childhood are based on estimates of local HIV prevalence when the respondents were adults, estimates of prevalence in childhood could be biased if education affects the probability of HIV infection. That is, if educational attainment affects HIV prevalence, this could confound my estimate of the effect of HIV prevalence in childhood on educational attainment. However, because HIV infection is positively correlated with educational attainment (Fortson, 2007), my estimates of the effect of HIV on human capital investment would be positively biased, making it less likely for me to find a negative effect.

I also test the sensitivity of my estimates to various specification changes. Estimates of the effect of HIV on attendance and primary school completion using a probit are negative and significant and larger in magnitude than the estimates in Table IV.

V. CHANNELS

Sections III and IV show evidence of a robust negative relationship between local HIV prevalence and human capital investment. The observed link may, however, arise through numerous channels.

Mortality Risk.

Local HIV prevalence, through its effect on expected longevity, may influence human capital investment. In an appendix, I develop an illustrative theoretical model of this channel. Using a straightforward extension of the standard human capital model, I show that changes in mortality risk, such as those generated by changes in local HIV prevalence, imply a decline in educational investment. Completing an additional year of schooling delays the age at which earnings commence but increases the value of those earnings; thus, individuals face a tradeoff between present and future consumption. By lowering the probability of survival, HIV lowers the value of future consumption. Hence, we would expect HIV to reduce completed schooling, assuming that the price of skill does not rise substantially.

I adopt a simple life-cycle model of human capital investment, incorporating aspects of Ben-Porath (1967), Card (1995, 1999), Preston (1980), and Meltzer (1992), the details of which are provided in the model appendix. My results show that $\frac{\partial S^*}{\partial HIV_{cr}}$, the effect of local HIV prevalence in childhood on completed schooling, should be negative. I calibrate the model using parameter estimates from the literature, and estimate that $\frac{\partial S^*}{\partial HIV_{cr}} = -3$, which implies that an increase in HIV prevalence of 10 percentage points should reduce completed years of schooling by about 0.3 years.

My empirical results indicate that HIV has a slightly larger effect on completed years of schooling than predicted by the model calibrations – that $\beta_1 = \frac{\partial S^*}{\partial HIV_{cr}}$ is on the order of -5, rather than -3. There are several possible explanations for this disparity. First, if parents and children are not myopic, it could be that the measure of HIV prevalence that I use in the econometric model (childhood HIV prevalence) understates the theoretically-relevant measure (expectations of adulthood HIV prevalence). If people expect HIV prevalence

(and, therefore, mortality risk) to rise, this could account for the larger estimated effect. Second, to the extent that this model is overly simplistic, the actual effect could differ from the model calibrations. For instance, increases in m and decreases in average levels of schooling could lead to increases in the return to schooling, which could in theory work to offset the decline in investment predicted by the model. A model accounting for general equilibrium effects could predict smaller and possibly positive effects on investment.

Conversely, the effect of local HIV prevalence on schooling could be larger than the model calibrations if parents or children think that schooling increases mortality risk, whether by increasing the risk of HIV infection or through other channels. Uncertainty about lifetime duration could further reduce schooling if parents or students are risk averse.

Also, it could be that the model ignores important features of investment decisions, causing the calibrations to underestimate the effect of HIV. Local HIV prevalence may be related to the probability of orphanhood, which has been shown to adversely affect educational outcomes. In addition, local HIV prevalence might reduce completed schooling because it affects the supply of teachers, economic wellbeing of families, responsibilities of children (to care for sick relatives), or the health of children (UNAIDS, 2002). I test empirically for the influence of five other possible determinants of human capital investment: orphanhood, caretaking requirements, the provision of schooling, family resources, and market wages.[17]

[17] We might also expect worsening child health to reduce schooling. While HIV may have had a substantial impact on child health (particularly through mother-to-child transmission of the virus), the fixed effects estimates reported in Section III likely reflect other channels. Children ill with HIV may be less likely to attend school or may progress more slowly through school. But because the effects in Section III are estimated on a sample of adults born before 1990, very few of these respondents would have been infected at birth, and any who were infected would be unlikely to survive to adulthood. And because the primary post-infancy mode of transmission in Africa is heterosexual sex, it is unlikely that infection after birth would directly affect completed years of schooling, especially since schooling levels are so low.

Orphanhood.

As noted in the introduction, several previous studies have shown that orphans are less likely to be enrolled in school than non-orphans. In theory, the negative relationship that I estimate between HIV and educational outcomes could reflect the fact that orphans have a greater likelihood of living in high-HIV areas. Using the fixed effects estimation strategy from Section III, I test whether the effect of local HIV prevalence on educational outcomes is driven by orphans. In particular, I compare the effect of local HIV prevalence between a sample of all children and a sample of non-orphans alone. Because orphan status is known only for sample children (but not adults), I estimate the effect of local HIV prevalence on a measure of progress through school (grades completed relative to potential) for children ages 7-14. I assume that local HIV prevalence follows a linear time path.[18] Table VII, column (1) presents the results for the full sample of children, which shows that local HIV prevalence is associated with slower progress through school. Children growing up in areas with HIV prevalence of 10 percent complete about 0.4 fewer grades each year than children growing up in areas without HIV. Excluding double orphans (children who lost both parents), the effect of HIV on progress through school is nearly unchanged. When I exclude both single and double orphans, again the estimate is negative and significant, and of the same magnitude as the comparison results. The results in Table VII indicate that the effect of HIV of human capital investment is not limited to orphans – there are in fact much broader effects of the HIV/AIDS epidemic.[19]

[18] The step function used in Table II would not be appropriate here because all sample children ages 7-14 were born after 1980. The nonlinear time path used in Table IV is also ill-suited to this analysis because some of the respondents were not yet age 10 at the time of the survey.

[19] I also estimate the effect of HIV using enrollment (an indicator for whether the respondent is currently in school) as the dependent variable (not shown), assuming HIV follows a linear time path. The fixed effects estimates are negative and significant in all three samples, and imply a large effect of local HIV prevalence on enrollment (around -4). Very low levels of HIV prevalence (of 1 or 2 percent) imply enrollment effects on par

Indeed, the effect of local HIV prevalence on schooling for orphans is actually smaller than for non-orphans. In results not shown, I test the effect of HIV on progress through school for a sample of (single and double) orphans. Local HIV prevalence is significantly related to progress, but is smaller in magnitude than the estimates for non-orphans. This is not surprising. Because losing a parent significantly reduces the probability of enrollment, educational outcomes will be quite poor for orphans in all areas. That is, conditional on being an orphan, I would not expect a child's educational outcomes to vary much with the local HIV rate.

Caretaking Requirements.

Children may also miss school to care for sick family members, contributing to the large negative effect of local HIV prevalence on schooling. To test whether this is driving my results, I exclude from the analysis children belonging to households known to have HIV-infected members (based on the DHS HIV test). Because I do not know HIV test results at the household level for Mali and Zambia, the sample is restricted to households undergoing HIV testing in Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania. I then test, on this sample of non-orphans with no known HIV-infected household members, whether the local HIV rate is negatively and significantly associated with attendance. The results, shown in Table VIII, indicate that my estimates are not driven by family caretaking requirements, as the effect of HIV on progress through school is negative and significant in this restricted sample.

---

with the effects of orphanhood estimated elsewhere (Case, Paxson, and Ableidinger, 2004; Evans and Miguel, 2007).

Provision of Schooling.

Another possibility is that reductions in the provision of schooling, perhaps driven by teacher death, are responsible for the estimated effect. To estimate the contribution of changes in provision to the decline in human capital investment, I test whether the effect of local HIV prevalence on schooling differs for men and women. Though there are differences in educational attainment between men and women, these are small enough that we would not expect the effect of HIV to differ much if provision were driving the effect. The results, shown in Table IX, show that local HIV prevalence has a significantly larger (negative) effect on the educational outcomes of men: the effect of HIV on men's attendance is over five times as large as the effect for women, even though men are only 18 percent more likely to have attended school. If provision were the primary channel through which HIV reduced schooling, we would not expect to see a large differential by sex.[20] In fact, the results show that the effect of local HIV prevalence on educational attainment operates almost entirely through its effect on men's educational outcomes.

Family Resources.

If local HIV prevalence reduces family resources even among households without illness or death, this could in theory drive the estimated decline. However, if this were the case, we would expect to see comparable declines in schooling for men and women, assuming that the returns to schooling are likewise comparable (as indicated by Schultz, 2004).[21] However, as shown in Table IX, there are substantially larger reductions in

[20] Furthermore, recent survey data for Uganda show that only 1.8 percent of parents cited poor school quality as a reason for why their children dropped out of primary school (Uganda Bureau of Statistics and ORC Macro, 2001). If provision changes were a primary determinant of the decline in schooling, we would expect a greater proportion of parents to cite school quality as a reason for attrition.

[21] If parents' preferences prioritize boys before girls when it comes to schooling, we might expect to find differences by sex. But in this situation, we would expect girls' schooling to fall by more than boys' schooling – the opposite of what we find.

completed schooling for men. These results suggest that the effects of HIV on family resources do not explain a large part of the estimated decline.

Market Wages.

It has been proposed that increases in mortality due to AIDS may drive up wages among survivors (Young, 2005). If this is the case, it could be that the relatively larger declines in schooling in high-HIV areas represent children's responses to increased wages, rather than increased mortality risk in adulthood. And while I find different effects of HIV by sex, this is not necessarily inconsistent with this hypothesis. If boys and girls are not substitutes in the labor market, then the greater educational declines among men could reflect a relatively larger increase in men's wages. However, if changes in wages are driving the schooling decline, we might not expect to find differences in the effect of HIV across other subgroups. For example, Muslims and non-Muslims are arguably substitutes in the labor market, but may have very different expectations of the risk of contracting HIV in adulthood.[22] In results not shown, I find much larger HIV-related declines in educational attainment among non-Muslims than among Muslims. Such large differences in the effect of HIV by religion would be unlikely to arise if market wages were the primary channel through which HIV reduced schooling.

Discussion.

I find that there are large declines in schooling related to local HIV prevalence, even after excluding orphans and children growing up in households with infected members. I

[22] These data show that Muslim men are more likely to be circumcised, which should reduce their risk of contracting HIV. However, within a region/sector, Muslims and non-Muslims do not show statistically different probabilities of HIV infection, controlling for sex. Nevertheless, Muslims are significantly less likely to say that they know someone who has or died of AIDS, which may be related to their expectations about the probability of infection.

also find that men experience significantly larger declines than women, consistent with the higher mortality risk among men in the early part of the HIV/AIDS epidemic (UNAIDS and WHO, 1999; Oster, 2006b).

While these results support the notion that HIV-related declines in human capital investment may be driven by expectations about mortality risk in adulthood, they are nevertheless inconclusive. The large estimated sex differential in the effect of HIV, while directionally consistent with early gender differences in rates of HIV infection, is much larger in magnitude than we would expect. And while this differential has declined over time, as we would have expected, it has not disappeared. Among children, boys show larger declines in educational attainment than girls in the same area – even though HIV is now more prevalent among women than men. So while this supports the mortality risk channel, a simple model of human capital investment does not fully explain these results.

VI. CONCLUSION

Exploiting regional and time variation in local HIV prevalence, I estimate the effect of the local HIV rate on human capital investment. I find that local HIV prevalence is associated with substantially worse educational outcomes: children living in areas with HIV rates of 10 percent complete about 0.5 fewer years of schooling than children living in areas without HIV. Children in areas with higher levels of HIV fare worse along a number of dimensions – they are less likely to attend school, less likely to complete primary school, and progress more slowly through school. These results are robust to numerous sensitivity checks, including tests for omitted variable bias, sample selection bias, and attrition.

I extend the analysis to empirically distinguish between channels from HIV to education. Building on a simple model of human capital investment, I show that HIV, through its effect on longevity, may reduce human capital investment among non-orphans

and orphans alike. Using parameter estimates from previous literature, I calibrate the model to show that longevity declines due to HIV would be expected to reduce educational attainment among children living in areas with HIV prevalence of 10 percent by about 0.3 years, more than half of the estimated effect. I evaluate five other channels – orphanhood, caretaking requirements, provision, family resources, and market wages – and conclude that none of these single-handedly drives the effect. My empirical findings are inconclusive but consistent with the theoretical model; the other channels perhaps contribute marginally to the larger estimated decline.

These results suggest that, when making schooling decisions, parents and children may respond to changes in the expected return to investment due to mortality risk. In fact, this suggests that the model developed here may have the potential to provide ballpark estimates of the effects of other diseases or interventions in other settings. And in the context of HIV/AIDS, my results provide some hope that expanded access to antiretroviral therapy would increase longevity among the HIV-infected and thus mitigate this effect.[23]

But absent marked improvements in longevity, the HIV/AIDS epidemic may impose a substantial economic cost. Based on these estimated effects and overall HIV prevalence figures, I would expect average eventual educational attainment among children today in sub-Saharan Africa to be about 0.3 years fewer than it would have been in the absence of HIV, with much larger declines in areas hardest hit by HIV. Moreover, HIV has not only endangered human capital investment and GDP (Krueger and Lindahl, 2001), but it has deprived a large fraction of adults of their productive prime, a disproportionate fraction of them relatively well-educated (Fortson, 2007). The substantial reductions in human

[23] Indeed, new longitudinal evidence from Kenya shows that the availability of antiretrovirals increases school-going among children whose parents receive treatment (Graff Zivin, Thirumurthy, and Goldstein, 2006). However, this study cannot distinguish between an effect on expectations of longevity and an effect through parental wellbeing.

capital investment estimated here, coupled with the high levels of mortality among adults today, have reduced both the stock and flow of human capital in the region, contributing to sub-Saharan Africa's economic woes.

REFERENCES

Auvert, Bertran, Dirk Taljaard, Emmanuel Lagarde, Joëlle Sobngwi-Tambekou, Rémi Sitta, and Adrian Puren (2005), "Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial," *PLoS Med*, 2(11), e298.

Ben-Porath, Yoram (1967), "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), Part 1, pp. 352-365.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Difference-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1), pp. 249-275.

Bloom, David E., and Ajay S. Mahal (1997), "Does the AIDS Epidemic Threaten Economic Growth?" *Journal of Econometrics*, 77(1), pp. 105-124.

Bonnel, René (2000), "HIV/AIDS and Economic Growth: A Global Perspective," *South African Journal of Economics*, 68(5), pp. 360-379.

Card, David (1995), "Earnings, Schooling, and Ability Revisited," in Polachek, Solomon, ed., *Research in Labor Economics*, Vol. 14. Greenwich, CT: JAI Press, pp. 23-48.

Card, David (1999), "The Causal Effect of Education on Earnings," in Ashenfelter, Orley and David Card, eds., *Handbook of Labor Economics, Volume 3*. Oxford, UK: Elsevier, pp. 1801-1863.

Case, Anne, and Cally Ardington (2006), "The Impact of Parental Death on School Outcomes: Longitudinal Evidence from South Africa," *Demography*, 43(3), pp. 401-420.

Case, Anne, Christina Paxson, and Joseph Ableidinger (2004), "Orphans in Africa: Parental Death, Poverty and School Enrollment," *Demography*, 41(3), pp. 483-508.

Evans, David K., and Edward Miguel (2007), "Orphans and Schooling in Africa: A Longitudinal Analysis," *Demography*, 44(1), pp. 35-57.

Fortson, Jane (2007), "The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS," working paper, University of Chicago.

Graff Zivin, Joshua, Harsha Thirumurthy, and Markus Goldstein (2006), "AIDS Treatment and Intrahousehold Resource Allocations: Children's Nutrition and Schooling in Kenya," NBER Working Paper 12689.

Hosegood, Victoria, Anna-Maria Vanneste, and Ian M. Timæus (2004), "Levels and Causes of Adult Mortality in Rural South Africa: The Impact of AIDS," *AIDS*, 18(4), pp. 663-671.

Iliffe, John (2006), *The African AIDS Epidemic: A History*. Oxford, UK: James Currey, Ltd.

Kalemli-Ozcan, Sebnem (2006), "AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa," NBER Working Paper 12181.

Krueger, Alan B., and Mikael Lindahl (2001), "Education for Growth: Why and For Whom?" *Journal of Economic Literature*, 39(4), pp. 1101-1136.

Meltzer, David (1992), "Mortality Decline, the Demographic Transition, and Economic Growth," Ph.D. Dissertation, University of Chicago.

Meyer, Bruce D. (1995), "Natural and Quasi-Experiments in Economics," *Journal of Business & Economic Statistics*, 13(2), pp. 151-161.

Mishra, Vinod, Martin Vaessen, J. Ties Boerma, Fred Arnold, Ann Way, Bernard Barrere, Anne Cross, Rathavuth Hong, and Jasbir Sangha (2006), "HIV Testing in National Population-Based Surveys: Experiences from the Demographic and Health Surveys," *Bulletin of the World Health Organization*, 84(7), pp. 537-545.

National Council for Population Development (Kenya) and Institute for Resource Development/Macro Systems, Inc. (1989), *Kenya Demographic and Health Survey, 1989*. Columbia, MD: Institute for Resource Development/Macro Systems, Inc.

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

Oster, Emily (2006a), "HIV and Sexual Behavior Change: Why Not Africa?" working paper, University of Chicago.

Oster, Emily (2006b), "Measuring the Magnitude of HIV in Africa," working paper, University of Chicago.

Oster, Emily (2007), "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa," working paper, University of Chicago.

Preston, Samuel H. (1980), "Causes and Consequences of Mortality Declines in Less Developed Countries During the Twentieth Century," in Easterlin, Richard A., ed., *Population and Economic Change in Developing Countries*. Chicago and London: University of Chicago Press, pp. 289-360.

Psacharopoulos, George and Harry Anthony Patrinos (2004), "Returns to Investment in Education: A Further Update," *Education Economics*, 12(2), pp. 111-134.

Rutstein, Shea Oscar and Guillermo Rojas (2003), *Guide to DHS Statistics*. Calverton, MD: Demographic and Health Surveys, ORC Macro.

Schultz, T. Paul (2004), "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education," *Journal of African Economies*, 13(AERC Supplement 2), pp. ii95-ii148.

Timæus, Ian M. and Momodou Jasseh (2004), "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys," *Demography*, 41(4), pp. 757-772.

Uganda Bureau of Statistics and ORC Macro (2001), *Uganda DHS EdData Survey 2001: Education Data for Decision-Making*. Calverton, Maryland, USA: Uganda Bureau of Statistics and ORC Macro.

UNAIDS (2002), *2002 Report on the Global HIV/AIDS Epidemic*. Geneva, Switzerland: UNAIDS.

UNAIDS (2004), *2004 Report on the Global AIDS Epidemic*. Geneva, Switzerland: UNAIDS.

UNAIDS (2006), *2006 Report on the Global AIDS Epidemic*. Geneva, Switzerland: UNAIDS.

UNAIDS and WHO (1999), *AIDS Epidemic Update: December 1999*. (Joint United Nations Programme on HIV/AIDS and World Health Organization), Geneva, Switzerland: UNAIDS and WHO.

US Census Bureau International Data Base (2005), U.S. Census Bureau, Population Division, International Programs Center, Data Update April 26, 2005.

Werker, Eric, Amrita Ahuja, and Brian Wendell (2006), "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis," Harvard Business School Working Paper 07-025.

Wise, Jacqui (2006), "Demand for Male Circumcision Rises in a Bid to Prevent HIV," *Bulletin of the World Health Organization*, 84(7), pp. 509-511.

Young, Alwyn (2005), "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," *Quarterly Journal of Economics*, 120(2), pp. 423-466.

DATA APPENDIX

Data for the analysis come from Demographic and Health Surveys (DHS), which are available from ORC Macro (http://www.measuredhs.com). The sample includes seven countries: Burkina Faso, Cameroon, Ghana, Kenya, Mali, Tanzania, and Zambia. The majority of the analysis uses only the most recent wave of the DHS for each country (2001 for Mali; 2001/2002 for Zambia; 2003 for Burkina Faso, Ghana, Kenya, and Tanzania; and 2004 for Cameroon).[24] However, in some robustness checks, I use previous waves of the data.

The DHS has several survey components, including a household questionnaire, women's questionnaire, and men's questionnaire. HIV testing was also conducted in the most recent wave in the seven countries in my sample. In Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania, these HIV test results can be linked to other respondent characteristics. In Mali and Zambia, HIV test results are unlinked to the survey data; however, test results can be used to calculate HIV prevalence for various subgroups, including local prevalence. Local HIV prevalence is estimated at the region and sector (urban or rural) level, and is calculated as HIV prevalence among adults ages 15-49 (for comparability across countries).

All households selected for the survey were asked to respond to a household questionnaire, which provides information about the age, sex, educational attainment, and school enrollment status of household members and visitors (in addition to other information). Adult women (ages 15-49) in all households were eligible for the women's survey. In Burkina Faso, Mali, and Zambia, men were interviewed and HIV tests were conducted in a one-in-three subsample of households. In Cameroon and Kenya, men were

[24] The 2003 DHS for Tanzania is also referred to as the HIV/AIDS Indicator Survey (AIS), and covers only mainland Tanzania. The data used in this analysis from that survey is from a preliminary release of the data.

interviewed and HIV tests were conducted in a one-in-two subsample of households. In Ghana and Tanzania, men were interviewed and HIV tests were conducted in all households. Men ages 15-59 were eligible for the men's survey, except in Kenya (ages 15-54) and Tanzania (ages 15-49). The analysis uses data on individual characteristics from the women's and men's surveys, if they are available; if unavailable, data are drawn from the household questionnaire.

Each respondent eligible for HIV testing was asked to provide a blood sample for testing. In Burkina Faso, Cameroon, Ghana, Kenya, Mali, and Tanzania, HIV testing was conducted on dried blood spot specimens collected by finger prick. In Zambia, the dried blood spot specimen came from a venous blood specimen.

Survey and HIV test non-response rates are shown in Appendix Table A.I. Response rates for the household, women's, and men's questionnaires are quite high. Response rates for the HIV test components are somewhat lower.

This analysis uses household weights in nearly all specifications, since in most cases information is drawn from the household questionnaire.[25] These weights adjust for the household sampling probability and household response rate (Rutstein and Rojas, 2003). Local HIV prevalence is calculated using DHS-provided HIV weights, which adjust for individual sampling probabilities and test non-response (separately by sex).

In some specifications, I use data on national HIV prevalence over time. Because UNAIDS HIV prevalence data from a temporally-comparable series are available only through 2001, I calculate predicted HIV in the survey year (or 2004) for these calculations.

[25] Childhood sector of residence is drawn from the men's and women's questionnaires, and therefore results using childhood sector are weighted using individual weights, which adjust the household weights by the individual sampling probabilities and response rates. Results in columns (2) and (3) of Table VIII use the product of HIV weights from the cluster, rescaled by the country mean for non-orphan 7- to 14- year-olds in tested households.

Predicted HIV is calculated by fitting a cubic (one per country) to the UNAIDS national HIV prevalence data and projecting HIV prevalence.

In robustness checks in Section IV, I use the local male circumcision rate as an instrument for the local HIV rate. The men's questionnaire for Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania asks adult men if they are circumcised. Local circumcision rates are calculated as the percentage of male respondents ages 20-30 in a region/sector who indicate that they are circumcised.

Separately, I also use data from previous waves of the DHS when regions in past cross-sections are an exact match to regions in the latest waves. In particular, the following regions are comparable across waves of the DHS (survey years in parentheses):

Ghana (1993, 1998, 2003):

Western, Central, Greater Accra, Volta, Eastern, Ashanti, Brong Ahafo, Northern, Upper West, Upper East

Kenya (1993, 1998, 2003):

Nairobi, Central, Coast, Nyanza, Western

Mali (1995/1996, 2001):

Kayes, Koulikoro, Sikasso, Segou, Mopti, Tombouctou, Gao, Bamako

Tanzania (1996, 1999, 2003):

Dodoma, Kilimanjaro, Tanga, Morogoro, Pwani, Dar es Salaam, Lindi, Mtwara, Ruvuma, Iringa, Mbeya, Singida, Tabora, Rukwa, Kigoma, Shinyanga, Kagera, Mwanza, Mara

Zambia (1992, 1996, 2001/2002):

Central, Copperbelt, Eastern, Luapula, Lusaka, Northern, North-Western, Southern, Western

I am also able to use the Cameroon (1991) data in a robustness check. These data have fewer regions than the 2004 data, but each 1991 region corresponds exactly to a set of

regions in the 2004 data. I do not use these data in the migration/mortality analysis, however, because the 2004 regions do not correspond one-to-one to the earlier waves.

In other robustness checks, I use the childhood sector of residence reported by the respondent. Childhood sector of residence is defined as where the respondent lived "most of the time until [he was] 12 years old."

TABLE A.I
SURVEY RESPONSE RATES

	Household	Women's	Men's	HIV Test		
Country	Questionnaire	Questionnaire	Questionnaire	Men	Women	Total
Burkina Faso	99.4	96.3	90.5	85.8	92.3	89.3
Cameroon	97.6	94.3	93.0	89.8	92.1	91.0
Ghana	98.7	95.7	93.8	80.0	89.3	84.9
Kenya	96.3	94.0	85.5	70.3	76.3	73.4
Mali	97.9	94.9	83.8	75.6	85.2	80.7
Tanzania	98.5	95.9	91.3	77.0	83.5	80.5
Zambia	98.2	96.4	88.7	73.3	79.4	76.5

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). Percent surveyed or tested among eligible respondents.

MODEL APPENDIX

Assume that an individual chooses schooling level S to maximize the expected present value of lifetime utility subject to a budget constraint:

$$\max_{S} \int_0^{\infty} U(c_t)e^{-rt}\, p(t)dt \text{ subject to } \int_0^{\infty} c_t e^{-rt}\, p(t)dt = \int_0^{\infty} y_t(S)e^{-rt}\, p(t)dt \ , \tag{A.1}$$

where $U(c_t)$ is an increasing, concave utility function over consumption in period t, r is the discount rate, $p(t)$ is the probability of survival to age t, and $y_t(S)$ is earnings in period t, which are a function of schooling. Importantly, assume that $y_t(S)$ and the price of consumption do not depend on mortality risk or the average level of schooling.

Assume that the only cost of schooling is foregone earnings and that earnings are constant thereafter, and are determined by the increasing, concave human capital production function $g(S)$ as follows:

$$y_t(S) = \begin{cases} 0, & t \in [0, S] \\ e^{g(S)}, & t \in (S, \infty) \end{cases}.$$

Let $\mu(z)$ be the probability of death at age z. Assume that individuals face a risk of mortality m at each age t, so $p(t)$ is:

$$p(t) = e^{-\int_0^t \mu(z)dz} = e^{-mt}.$$

By assuming that capital markets function perfectly, we can simplify the representative individual's maximization problem as follows:

$$\max_{S} \int_S^{\infty} e^{g(S)} e^{-(r+m)t}\, dt = \frac{e^{g(S)-(r+m)S}}{r+m} \ . \tag{A.2}$$

Therefore, individuals discount future earnings at rate $r + m$. Solving (A.2) yields the optimal schooling choice, S^*, at which the rate of return to schooling is equal to the sum of the discount rate and mortality rate at each age:

$$g_s(S^*) = r + m.$$

I assume that m depends on HIV and other arguments, and solve for S^*:

$$S^* = g_s^{-1}(r + m(HIV,...)).$$

Using the chain rule, we can calculate the effect of HIV on the optimal level of schooling:

$$\frac{\partial S^*}{\partial HIV} = \frac{dg_s^{-1}(x^*)}{dx}\frac{d(r + m(HIV,...))}{dm}\frac{\partial m}{\partial HIV},$$

from which it follows that

$$\frac{\partial S^*}{\partial HIV} = \frac{1}{g_{ss}(S^*)}\frac{\partial m}{\partial HIV}. \qquad \text{(A.3)}$$

By concavity, $g_{ss}(S^*) < 0$, and it is straightforward to assume that mortality risk rises with the local HIV rate, $\frac{\partial m}{\partial HIV} > 0$, so $\frac{\partial S^*}{\partial HIV} < 0$.

Deriving estimates of $g_{ss}(S^*)$ and $\frac{\partial m}{\partial HIV}$ from previous literature, I can approximate the magnitude of $\frac{\partial S^*}{\partial HIV}$ predicted by the model. Relying on estimates of the return to schooling in sub-Saharan Africa from Psacharopoulos and Patrinos (2004), I estimate $g_{ss}(S^*)$. In particular, I assume that the rate of return to schooling diminishes at a constant rate, $g_{ss}(S^*) = k$, and calculate the difference in the return to secondary schooling and the return to primary schooling, divided by the number of years of secondary schooling. Because the number of years of secondary schooling differs across countries, I calculate two estimates of $g_{ss}(S^*)$: -0.026 and -0.022. To calculate $\frac{\partial m}{\partial HIV}$, I rely on results from Timæus and Jasseh (2004), who provide estimates of adult mortality risk over time in a sample of

African countries. By fully attributing changes in mortality to changes in HIV prevalence, I calculate $\frac{\partial m}{\partial HIV}$ for several countries. Based on these estimates, I construct two plausible estimates of $\frac{\partial m}{\partial HIV}$: 0.06 and 0.09. Taken together with the estimates of $g_{ss}(S^*)$, these parameter values suggest that $\frac{\partial S^*}{\partial HIV}$ should be about -3.

TABLE I.A
LOCAL HIV PREVALENCE IN SURVEY YEAR, DETAILED SUMMARY STATISTICS

HIV RATE	Burkina Faso	Cameroon	Ghana	Kenya	Mali	Tanzania	Zambia	Total
Mean	1.70	5.90	2.12	7.42	1.48	8.07	17.38	6.33
Standard Deviation	1.45	2.68	1.14	5.69	1.38	6.23	7.57	6.62
25th Percentile	0.50	3.71	1.29	4.09	0.40	4.08	10.19	1.71
Median	1.35	6.40	2.09	6.38	1.35	6.41	17.18	4.19
75th Percentile	2.47	7.91	2.62	10.39	2.11	10.71	22.37	8.29
Observations	25	22	20	15	16	42	18	158

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). Table shows detailed summary statistics for the local HIV rate in the survey year, which is estimated for each region and sector of residence (within each country). The unit of observation is a region/sector. These HIV rates are calculated from the DHS HIV data using a sample which includes men and women ages 15-49, weighted using appropriate HIV sample weights.

TABLE I.B
SAMPLE CHILD CHARACTERISTICS, SUMMARY STATISTICS

	N	Mean	Standard Deviation
Local HIV Prevalence in Survey Year	73,265	0.050	0.055
Progress	69,755	0.358	0.370
Year of Birth	73,265	1992	2.622
Female	73,252	0.494	0.500
Rural	73,265	0.727	0.446
Burkina Faso	73,265	0.195	0.396
Cameroon	73,265	0.159	0.366
Ghana	73,265	0.082	0.274
Kenya	73,265	0.118	0.323
Mali	73,265	0.218	0.413
Tanzania	73,265	0.105	0.306
Zambia	73,265	0.124	0.330

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All results are weighted using provided household sample weights. Sample includes boys and girls ages 7-14 at the time of the survey. Local HIV Prevalence in Survey Year is the estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Progress is the number of completed years of schooling divided by the potential (age – 6); values outside the [0,1] range are coded as missing. Year of Birth is the respondent's year of birth. Female is an indicator for whether the respondent is female.

TABLE I.C
SAMPLE ADULT CHARACTERISTICS, SUMMARY STATISTICS

	N	Mean	Standard Deviation
Local HIV Prevalence in Survey Year	133,949	0.054	0.057
Years of Schooling	133,563	4.780	4.515
Years of Schooling > 0	133,563	0.625	0.484
Completed Primary School	133,644	0.433	0.495
Year of Birth	133,949	1974	9.702
Female	133,947	0.527	0.499
Rural	133,949	0.641	0.480
Burkina Faso	133,949	0.185	0.388
Cameroon	133,949	0.174	0.379
Ghana	133,949	0.083	0.276
Kenya	133,949	0.130	0.336
Mali	133,949	0.195	0.396
Tanzania	133,949	0.107	0.309
Zambia	133,949	0.125	0.331

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All results are weighted using provided household sample weights. Sample includes men and women ages 15-49 at the time of the survey. Local HIV Prevalence in Survey Year is the estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Years of Schooling is completed years of schooling. Years of Schooling > 0 is an indicator for whether the respondent completed at least one year of schooling. Completed Primary School is an indicator for whether the respondent completed primary school. Year of Birth is the respondent's year of birth. Female is an indicator for whether the respondent is female.

TABLE II
FIXED EFFECTS REGRESSION: STEP FUNCTION

COMPLETED SCHOOLING	(1) YEARS	(2) YEARS > 0	(3) PRIMARY
Local HIV Prevalence	-5.920*	-0.403*	-0.687*
	(0.803)	(0.079)	(0.134)
Female	-1.161*	-0.104*	-0.107*
	(0.073)	(0.008)	(0.007)
Constant	4.643*	0.772*	0.389*
	(0.220)	(0.018)	(0.022)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 15-49	Ages 15-49	Ages 15-49
Observations	133,562	133,562	133,642

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes men and women ages 15-49. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Local HIV Prevalence is estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon) times an indicator for whether the respondent is in an affected (post-1979) birth cohort. Female is an indicator for whether the respondent is female. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE III
FIXED EFFECTS REGRESSION: LINEAR FUNCTION

COMPLETED SCHOOLING	(1) YEARS	(2) YEARS > 0	(3) PRIMARY
Local HIV Prevalence	-11.505*	-0.827*	-1.420*
	(1.440)	(0.156)	(0.263)
Female	-1.160*	-0.104*	-0.107*
	(0.073)	(0.008)	(0.007)
Constant	4.817*	0.786*	0.413*
	(0.218)	(0.019)	(0.024)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 15-49	Ages 15-49	Ages 15-49
Observations	133,562	133,562	133,642

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes men and women ages 15-49. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is calculated assuming that HIV was 0 through 1980 and grew linearly from 1980, using estimates (based on DHS HIV data) of HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Female is an indicator for whether the respondent is female. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE IV
FIXED EFFECTS REGRESSION: NONLINEAR FUNCTION

COMPLETED SCHOOLING	(1) YEARS	(2) YEARS > 0	(3) PRIMARY
Local HIV Prevalence	-5.230*	-0.356*	-0.713*
	(0.777)	(0.072)	(0.137)
Female	-1.162*	-0.104*	-0.107*
	(0.073)	(0.008)	(0.007)
Constant	4.602*	0.769*	0.390*
	(0.233)	(0.018)	(0.023)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 15-49	Ages 15-49	Ages 15-49
Observations	133,562	133,562	133,642

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes men and women ages 15-49. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is derived using estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon) in conjunction with national HIV prevalence trends from UNAIDS. This estimation assumes that local HIV prevalence follows the same time path as national HIV prevalence. Female is an indicator for whether the respondent is female. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE V
ROBUSTNESS CHECK: DIFFERENCES PRIOR TO AFFECTED TIME PERIOD

COMPLETED SCHOOLING	(1) YEARS	(2) YEARS > 0	(3) PRIMARY
Local HIV Prevalence (Survey Year) × Year of Birth	0.111	-0.002	0.006
	(0.081)	(0.007)	(0.007)
Female	-1.445*	-0.110*	-0.137*
	(0.088)	(0.008)	(0.009)
Constant	5.111*	0.604*	0.450*
	(0.132)	(0.018)	(0.014)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Birth Cohorts 1951-1979	Birth Cohorts 1951-1979	Birth Cohorts 1951-1979
Observations	85,105	85,105	85,171

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes men and women born 1951-1979 among those ages 15-49. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Local HIV Prevalence (Survey Year), which is interacted with a linear year of birth term, is the estimated HIV prevalence in the region and sector (urban or rural), and is calculated using estimates (based on DHS HIV data) of HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Female is an indicator for whether the respondent is female. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE VI
SENSITIVITY ANALYSIS: MIGRATION AND MORTALITY

	(1) BASE (TABLE IV)	(2) ALL WAVES	(3) LATEST WAVE	(4) CHILDHOOD SECTOR
YEARS OF SCHOOLING				
Local HIV Prevalence	-5.230*	-3.426*	-3.588*	-10.186*
	(0.777)	(1.197)	(0.984)	(3.876)
Observations	133,562	85,483	77,922	59,451
YEARS OF SCHOOLING > 0				
Local HIV Prevalence	-0.356*	-0.244*	-0.274*	-0.652*
	(0.072)	(0.068)	(0.076)	(0.259)
Observations	133,562	85,483	77,922	59,451
COMPLETED PRIMARY				
Local HIV Prevalence	-0.713*	-0.573*	-0.531*	-1.760*
	(0.137)	(0.203)	(0.139)	(0.356)
Observations	133,642	85,513	77,951	59,452
Sample	Ages 15-49, Latest Wave, 158 Areas	Ages 15-25, All Waves, 100 Areas	Ages 15-49, Latest Wave, 100 Areas	Ages 15-49, Latest Wave, 158 Areas
Additional Controls	Female, Constant, Region × Sector FEs, Birth Year FEs	Female, Constant, Region × Sector FEs, Birth Year FEs, Wave FEs	Female, Constant, Region × Sector FEs, Birth Year FEs	Female, Constant, Region × Sector FEs, Birth Year FEs

Notes. Base and Childhood Sector results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All Waves results are from the DHS for some regions in Ghana (1993, 1998, 2003), Kenya (1993, 1998, 2003), Mali (1995/1996, 2001), Tanzania (1996, 1999, 2003), and Zambia (1992, 1996, 2001/2002). Latest Wave results are from the DHS for some regions in Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household ((1), (2), and (3)) or individual (4) sample weights with clustering on the region/sector. Sample includes men and women ages 15-25 in specification (2), and ages 15-49 in specifications (1), (3), and (4). In specifications (2) and (3), the sample is restricted to respondents living in regions that are comparable across waves (see data appendix). In specification (4), local HIV prevalence is estimated using the respondent's current region of residence and childhood sector of residence; the latter is only known for some respondents. The dependent variable is completed years of schooling (top panel), an indicator for whether the respondent completed at least one year of schooling (middle panel), or an indicator for whether the respondent completed primary school (bottom panel). Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is derived using estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon) in conjunction with national HIV prevalence trends from UNAIDS. This estimation assumes that local HIV prevalence follows the same time path as national HIV prevalence. All regressions include controls for gender (coefficients not reported). Specification (2) also includes wave fixed effects. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE VII
CHANNELS: ORPHANHOOD

COMPLETED SCHOOLING	(1) PROGRESS	(2) PROGRESS	(3) PROGRESS
Local HIV Prevalence	-4.289*	-4.255*	-4.498*
	(0.443)	(0.448)	(0.464)
Female	-0.018*	-0.018*	-0.019*
	(0.006)	(0.006)	(0.006)
Constant	0.573*	0.571*	0.565*
	(0.025)	(0.025)	(0.024)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 7-14	Ages 7-14, Exclude Double Orphans	Ages 7-14, Exclude Single Orphans
Observations	69,742	67,979	59,764

Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes boys and girls ages 7-14. The dependent variable is the number of completed years of schooling divided by the potential (age – 6); values outside the [0,1] range are coded as missing. Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is calculated assuming that HIV was 0 through 1980 and grew linearly from 1980, using estimates (based on DHS HIV data) of HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Female is an indicator for whether the respondent is female. "Double Orphans" are respondents whose fathers and mothers are not known to be alive. "Single Orphans" are respondents whose fathers or mothers (or both) are not known to be alive. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE VIII
CHANNELS: CARETAKING REQUIREMENTS

COMPLETED SCHOOLING	(1) PROGRESS	(2) PROGRESS	(3) PROGRESS
Local HIV Prevalence	-4.498*	-4.435*	-4.972*
	(0.464)	(1.668)	(1.807)
Female	-0.019*	-0.001	-0.003
	(0.006)	(0.010)	(0.010)
Constant	0.565*	0.620*	0.635*
	(0.024)	(0.094)	(0.074)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 7-14, Exclude Single Orphans	Ages 7-14, Testing Sample	Ages 7-14, Testing Sample, Exclude if HIV+ Household Member
Observations	59,764	21,366	19,852

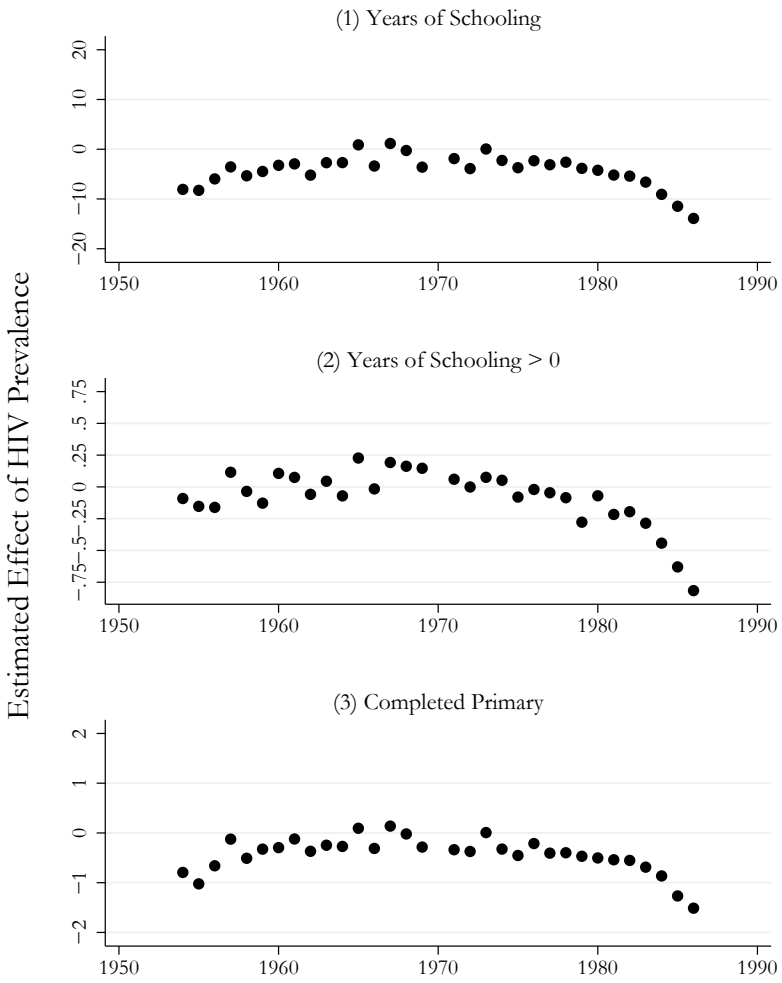
Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects with clustering on the region/sector. Weights in column (1) are household sample weights. Weights in columns (2) and (3) are the product of HIV weights from the cluster, rescaled by the country mean. Sample includes boys and girls ages 7-14. The dependent variable is the number of completed years of schooling divided by the potential (age – 6); values outside the [0,1] range are coded as missing. Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is calculated assuming that HIV was 0 through 1980 and grew linearly from 1980, using estimates (based on DHS HIV data) of HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Female is an indicator for whether the respondent is female. "Single Orphans" are respondents whose fathers or mothers (or both) are not known to be alive. "Testing Sample" includes households in which at least one member was tested for HIV, and excludes Mali and Zambia, in which HIV test results cannot be matched to the household. Column (3) excludes from the testing sample children living in households with at least one HIV-positive member. Huber-White standard errors are in parentheses. * = p-value < .05

TABLE IX
CHANNELS: DIFFERENCES BY SEX

COMPLETED SCHOOLING	(1) YEARS	(2) YEARS > 0	(3) PRIMARY
Local HIV Prevalence × Male	-7.738*	-0.526*	-0.889*
	(0.922)	(0.070)	(0.121)
Local HIV Prevalence	-1.633	-0.112	-0.300*
	(0.888)	(0.061)	(0.114)
Female	-1.327*	-0.115*	-0.126*
	(0.086)	(0.009)	(0.009)
Constant	4.681*	0.775*	0.399*
	(0.237)	(0.019)	(0.023)
Additional Controls	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs	Region × Sector FEs, Birth Year FEs
Sample	Ages 15-49	Ages 15-49	Ages 15-49
Observations	133,562	133,562	133,642

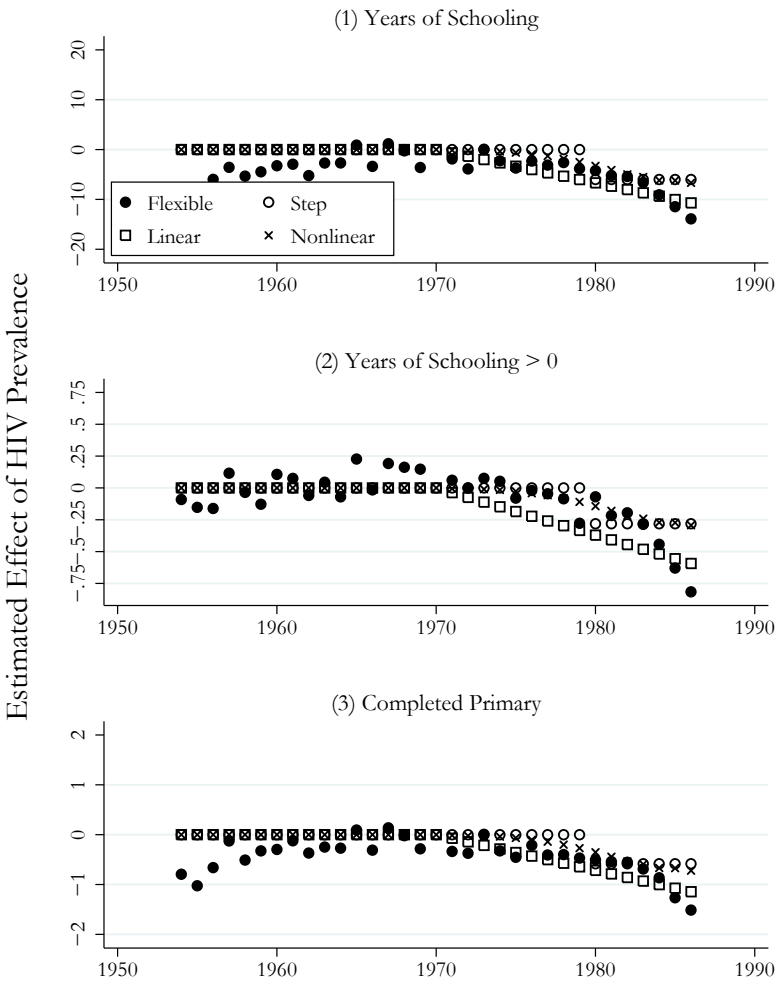
Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions with region by sector fixed effects and year of birth fixed effects, weighted using provided household sample weights with clustering on the region/sector. Sample includes men and women ages 15-49. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Local HIV Prevalence is the estimated HIV prevalence in the region and sector (urban or rural) at age 10, and is derived using estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon) in conjunction with national HIV prevalence trends from UNAIDS. This estimation assumes that local HIV prevalence follows the same time path as national HIV prevalence. Male is an indicator for whether the respondent is male. Female is an indicator for whether the respondent is female. Huber-White standard errors are in parentheses. * = p-value < .05

FIGURE 1
FIXED EFFECTS REGRESSION: FLEXIBLE FUNCTION



Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions, weighted using provided household sample weights. Sample includes men and women born 1954-1986. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Plotted results are coefficients from the interactions of local HIV prevalence and year of birth indicators, and can be interpreted as effects relative to the 1970 birth cohort. Local HIV Prevalence is the estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Regressions also include an indicator for whether the respondent is female, year of birth fixed effects, and region/sector fixed effects.

FIGURE 2
FIXED EFFECTS REGRESSION: COMPARISON OF TIME PATHS



Notes. Results are from the DHS for Burkina Faso (2003), Cameroon (2004), Ghana (2003), Kenya (2003), Mali (2001), Tanzania (2003), and Zambia (2001/2002). All regressions are weighted least squares regressions, weighted using provided household sample weights. Sample includes men and women born 1954-1986. The dependent variable is completed years of schooling (1), an indicator for whether the respondent completed at least one year of schooling (2), or an indicator for whether the respondent completed primary school (3). Plotted results for flexible specification are coefficients from the interactions of local HIV prevalence and year of birth indicators, relative to the 1970 birth cohort, where local HIV prevalence is the estimated (based on DHS HIV data) HIV prevalence among men and women ages 15-49 in the region and sector in 2001 (Mali), 2001-2002 (Zambia), 2003 (Burkina Faso, Ghana, Kenya, and Tanzania), or 2004 (Cameroon). Step, linear, and nonlinear results are the comparable to the results from Tables II, III, and IV, respectively, where the coefficient is interacted with the time path. Regressions also include an indicator for whether the respondent is female, year of birth fixed effects, and region/sector fixed effects.