

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

Luke Chicoine*

Bates College and IZA

Emily Lyons

Bates College

Alexia Sahue

Bates College

Appendix
(For Online Publication)

JEL classification: I15; I25; O55

Keywords: HIV/AIDS; mortality risk; schooling

*Corresponding author: Department of Economics, Bates College and IZA. Email: lchicoin@bates.edu.

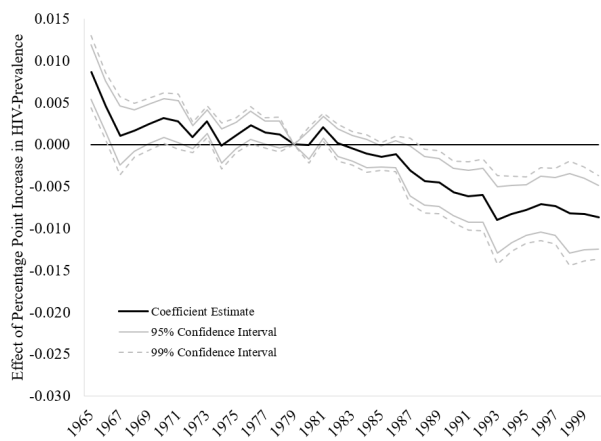
Appendix

A Cohort-Specific HIV Prevalence Rates

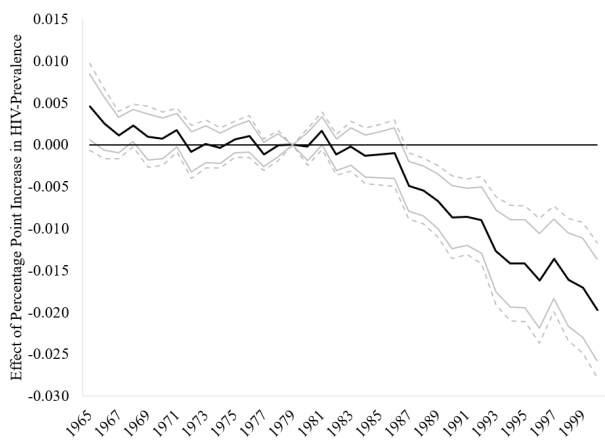
We construct cohort-specific estimates of regional HIV prevalence rates for the time period covering the initial outbreak of the disease through 2016. This cohort-specific measure is constructed using annual country-level estimates from the Global Burden of Disease (GBD) Study ([Global Burden of Disease Collaborative Network, 2018](#); [Murray et al., 2018](#); [Roth et al., 2018](#)) and combined with relative DHS prevalence rates in regions across each country, from each survey wave. For example, the prevalence rate in the northern region of Malawi was 68% of the national prevalence in the 2004 wave of the DHS (8.04 relative to 11.79); therefore, the northern region is assigned 68% of the GBD's national estimate in 2004 (10.5 versus 15.4). For non-survey years, prevalence is estimated assuming a linear trend between survey years. For all years prior to the first DHS survey (and after the most recent wave), the dispersion across regions is assumed to be constant, and equal to the relative prevalence in the initial (final) survey. When these data are used in analysis, observations from each birth cohort are matched with the prevalence rate in their region during their age six year. This is to simulate the mortality risk environment in which they, or their family, make their initial schooling decisions. Using an annual estimate allows us to both expand the sample to include DHS household survey data from the 1990s to account for the effect of HIV prevalence on human capital investment over an additional decade, and to introduce variation in the prevalence rate across cohorts.

The survey wave HIV prevalence measure used throughout the paper is likely the most accurate measure of HIV prevalence that we have available for the year of the survey. However, this variable is measuring the prevalence of the disease for adults, after most schooling decisions are completed. Although a number of assumptions are needed to create the cohort-specific regional prevalence estimates described in the above paragraph, these cohort-specific estimates concede some accuracy to approximate the ideal dataset of the HIV prevalence rate at the time an individual entered school. Furthermore, using the cohort-specific measures allow us to expand the sample to DHS waves without attached HIV data. When the cohort-specific measure is used 45 additional DHS survey waves without an attached HIV module are included in the sample. The additional surveys are: Burkina Faso (1993, 1998); Cameroon (1991, 1998); Chad (1996); Congo (2005, 2011); Cote d'Ivoire (1994); Ethiopia (2000); Gabon (2000); Ghana (1993, 1998); Guinea (1999); Kenya (1993, 1998); Malawi (1992, 2000); Mali (1995, 2001, 2012); Mozambique (1997, 2003, 2011); Namibia (1992, 2000, 2006); Niger (1992, 1998); Senegal (1992, 1997); South Africa (1998); Tanzania (1991, 1996, 1999); Togo (1998); Uganda (1995, 2000, 2006, 2016); Zambia (1992; 1996); Zimbabwe (1994, 1999). Data from Rwanda are removed from this sample, time consistent regional boundaries could not be determined for the entirety of the period.

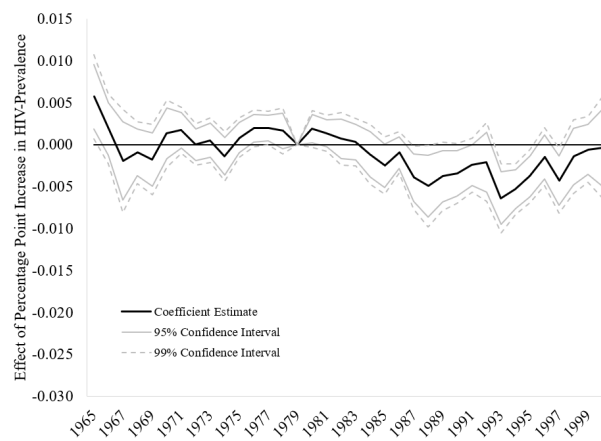
Appendix Figures and Tables



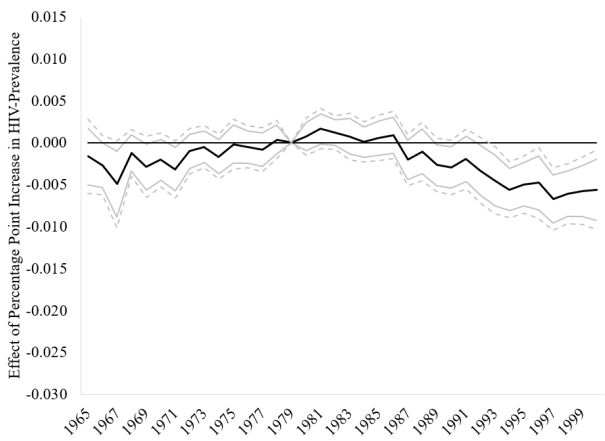
(a) Any School – Men (Ages 15-49)



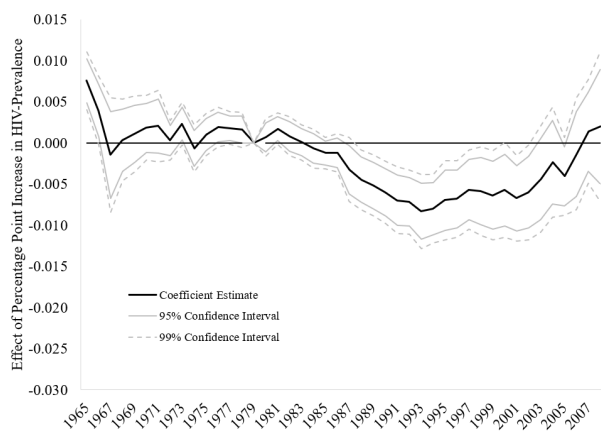
(b) Any School – Women (Ages 15-49)



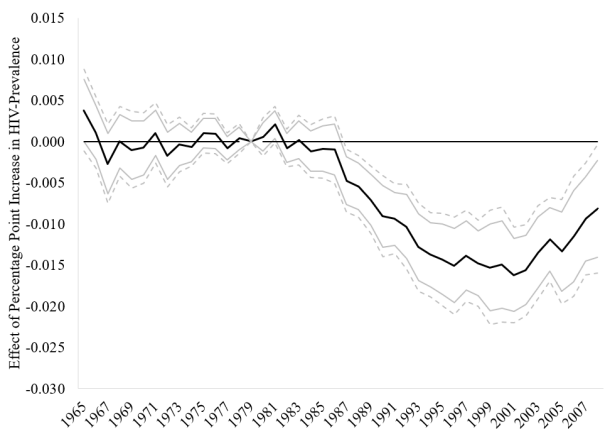
(c) Completed Primary School – Men (Ages 15-49)



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

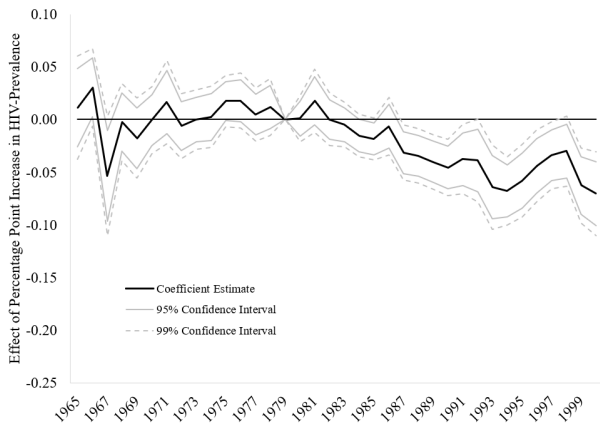


(e) Any School – Men (Ages 7-49)

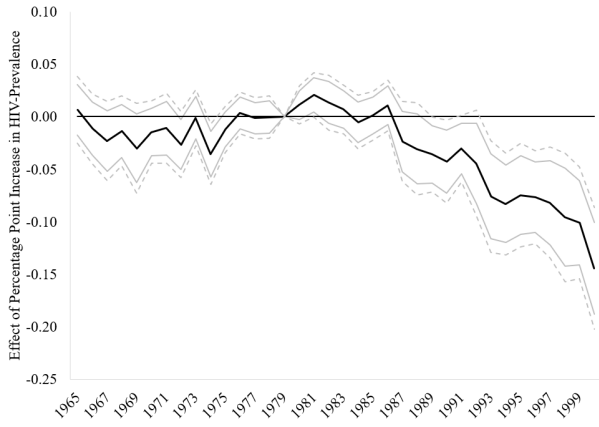


(f) Any School – Women (Ages 7-49)

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



(a) Years of Schooling – Urban



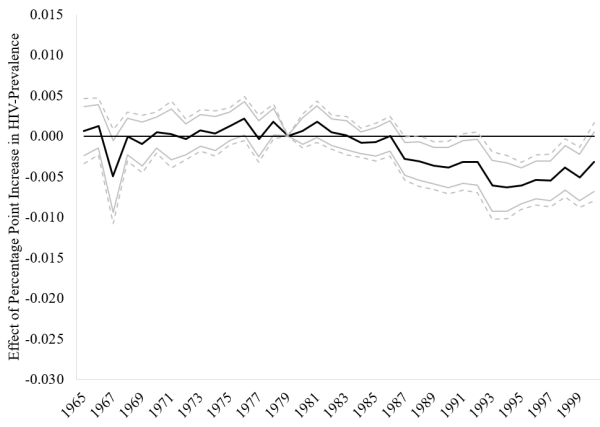
(b) Years of Schooling – Rural



(c) Any School – Urban



(d) Any School – Rural

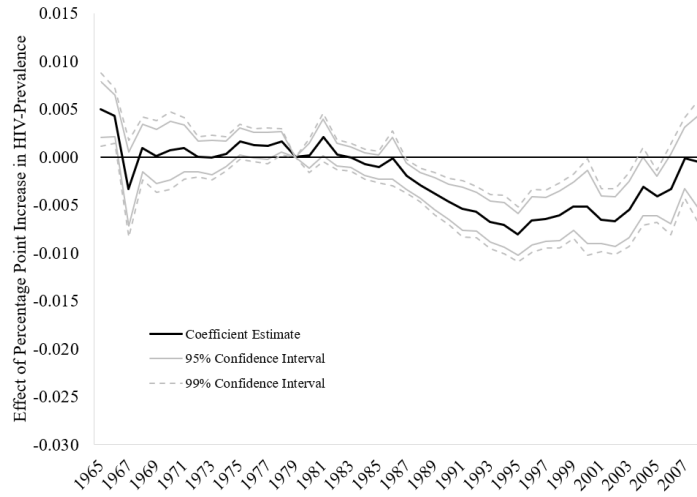


(e) Completed Primary School – Urban

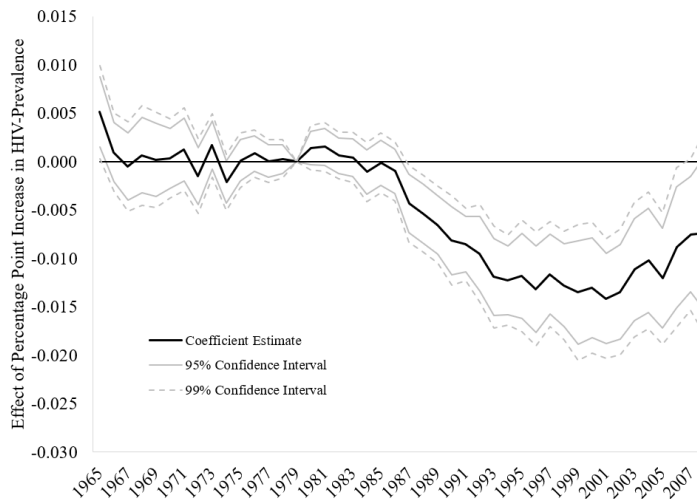


(f) Completed Primary School – Rural

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



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



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

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

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

	(1)	(2)		(3)		(4)		(5)	(6)		
	Angola	Burkina Faso		Burundi		Cameroon		Chad	Congo		
Survey Year	2015	2003	2010	2010	2016	2004	2011	2014	2009		
Mean	1.99	1.77	0.95	1.34	0.91	5.27	4.10	1.49	3.20		
S.d.	1.20	1.09	0.59	0.74	0.51	2.67	2.10	1.19	1.17		
Regions	18	13	13	4	4	12	12	21	12		
	(7)		(8)		(9)			(10)	(11)		
	DRC		Côte d'Ivoire		Ethiopia			Gabon	Gambia		
Survey Year	2007	2013	2005	2011	2005	2010	2016	2012	2013		
Mean	1.32	1.15	4.65	3.63	1.42	1.37	0.86	4.12	1.90		
S.d.	0.78	0.89	1.24	1.00	0.97	0.95	0.68	0.99	0.60		
Regions	11	11	10	10	11	11	11	10	8		
	(12)		(13)		(14)		(15)			(16)	
	Ghana		Guinea		Kenya		Lesotho			Liberia	
Survey Year	2003	2014	2005	2012	2003	2008	2004	2009	2014	2007	2013
Mean	2.21	1.99	1.50	1.71	6.74	6.36	23.35	22.77	24.17	1.55	2.04
S.d.	0.90	0.81	0.55	0.57	4.14	3.88	3.67	2.83	3.38	0.85	1.18
Regions	10	10	5	5	8	8	10	10	10	5	5
	(17)			(18)			(19)		(20)	(21)	
	Malawi			Mali			Mozambique		Namibia	Niger	
Survey Year	2004	2010	2015	2001	2006	2012	2009	2015	2013	2006	2012
Mean	11.64	10.59	9.13	1.72	1.29	1.12	11.13	13.11	14.29	0.69	0.33
S.d.	6.57	4.33	4.68	0.55	0.48	0.32	6.12	6.38	3.61	0.41	0.25
Regions	3	3	3	9	9	6	11	11	13	8	8
	(22)			(23)			(24)		(25)		
	Rwanda			Senegal			Sierra Leone		South Africa		
Survey Year	2005	2010	2014	2005	2010	2017	2008	2013	2016		
Mean	2.99	2.92	2.87	0.71	0.67	0.47	1.47	1.41	21.25		
S.d.	1.33	1.65	1.37	0.59	0.43	0.34	0.79	0.64	5.96		
Regions	5	5	5	4	4	4	4	4	9		
	(26)		(27)	(28)	(29)			(30)			
	Tanzania		Togo	Uganda	Zambia			Zimbabwe			
Survey Year	2003	2007	2011	2013	2011	2001	2007	2013	2005	2010	2015
Mean	6.85	5.67	5.04	2.40	7.29	15.33	13.85	12.95	18.12	15.29	13.81
S.d.	3.20	3.41	2.50	1.01	2.07	4.94	4.93	3.77	1.75	2.17	2.64
Regions	19	20	20	6	10	9	9	9	10	10	10

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table A.8: Replication of Fortson (2011) Table 7 –
Effect of HIV on Education Outcomes - Migration and Mortality (Age < 25)

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

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

Table A.9: Effect of HIV on Schooling and Child Labor: Ages 7 to 14

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

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

Table A.10: Replication of Fortson (2011) Table 9: Differences by Sex

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

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

Table A.11: Effect of HIV on Education Outcomes

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

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

Table A.12: Effect of HIV on Education Outcomes – Removing Regions with Highest and Lowest HIV Prevalence

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

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