Applied Economics, 2006, **38**, 491–503

Does female education prevent the spread of HIV-AIDS in Sub-Saharan Africa?

Robert J. Brent

Fordham University, Bronx, NY, 10458
E-mail: Brent@fordham.edu

Cross-section estimates of 31 countries for the year 2000 are used to examine the effect of the level of female education, and the gap between male and female levels, on the prevalence of HIV/AIDS in Sub-Saharan Africa. Many standard types of education are analysed and a non-standard category involving under- and over-aged primary school enrollees are identified. Although it is found that female education is positively related to country infection rates, there is scope to reduce infection rates by substituting one type of education for another. It is estimated that female infection rates peak within 15–17 years of the first case reported.

I. Introduction

At this point in history, the HIV-AIDS pandemic is very much an African phenomenon and females bear the greatest burden. UNAIDS has estimated that by the end of 2003, over 40 million people were living with HIV/AIDS and 26.5 million live in Sub-Saharan Africa. African women were 1.2 times more likely to be infected than men. Moreover, there is strong evidence that the younger the age cohort the greater is the percentage of females involved.[1] This implies that policy interventions need to be focused on younger, school-aged females. The limited public health budgets of African countries mean that there are insufficient resources available for costly anti-retroviral drugs as used in middle and high-income countries. Prevention seems to be the most cost-effective option. Since school-aged females are most at risk, some sort of school-based prevention programme, i.e., formal or specialized education, would appear to have the greatest priority.

This study tests whether female education does actually prevent the spread of HIV-AIDS in Sub-Saharan Africa using a cross-section of 31 countries. Female education is examined in many different forms, and includes a new measure to cover non-standard students. Three alternative estimation frameworks are applied and a range of estimation techniques are used. Because it is found that female education is positively related to HIV/AIDS rates, the scope for lowering infections is analysed by substituting one type of female education for another. One advantage of the methods that is employed is that one can also provide an estimate of the turning points for national female infection rates.

The next section provides the necessary background to the work, summarizing what is known about this subject, how one intends to extend this knowledge by looking at the type of student involved, and explaining the data and sample of countries. Then three alternative estimation frameworks are outlined. This is followed by a presentation of the

[1] As the World Bank (2002) reports: ‘In 11 population-based studies, the average infection rates in teenage African girls were over five times higher than those in teenage boys.’

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00036840500392045

estimation equations, which covers the specification of all the variables and a statement of how the equations will be estimated. After giving the results, which includes estimates of the infection turning points, the paper closes with the summary and discussion section.

II. Preliminary Considerations

Previous research

A recent report by the World Bank (1999) summarizes the literature on the global impact of HIV-AIDS and how government policies can be structured to mitigate this impact. A number of background studies were commissioned for the report. An important study by Over (1998) attempted to explain why it is that some countries have higher rates of urban HIV/AIDS infection than others. In a sample of 72 countries (32 from Sub-Saharan Africa) they found that there were eight significant variables. Positively related to infection rates were: the age of the epidemic; the percentage of the population that are foreign born; the extent of relative income inequality (as measured by the Gini coefficient); the size of military personnel as a percentage of the population; and the magnitude of the male-female gender gap in urban areas. Negatively related to infection rates were GNP per capita and the percentage of the population that were Muslim.

The remaining determinant, that had a positive effect on country infection rates, was the male–female literacy gap. This factor seemed to have the most immediate policy relevance, for if a way could be found to increase female literacy (holding male literacy rates constant) then the gap would be lowered and so would the infection rate.

The value of doing exploratory research like the Over study is that it can subsequently lead to causal research that tests specific hypotheses. In a subsequent book, World Bank (2002), the authors focused explicitly on the education gap variable to argue that a good general, basic education was the most cost-effective way for governments to intervene in the HIV/AIDS epidemic. They listed a number of good reasons why expanding female education would be helpful in this regard:

> Better educated women are more likely, in comparison with their peers, to delay marriage and childbearing, have fewer children and healthier babies, enjoy better earning potential, have stronger decisionmaking and negotiation skills as well as higher self-esteem, and avoid commercial sex.

Clearly, in the African context, where the HIV infection is primarily caused by unprotected, heterosexual intercourse that has intergenerational mixing and is not always consensual, female education is likely to lead to greater economic and social empowerment. The specific mechanism by which the literacy gap was thought by the World Bank to have its impact on infection rates was via the gap leading to fewer conventional job opportunities or lower wages for women, thus leading to a greater supply of female sex workers.

This World Bank book was one of the first to emphasize that there was likely to be two-way causation between education and AIDS infection. There was not just the impact of education on AIDS, but also the reverse connection. Many teachers had AIDS in Sub-Saharan Africa.[2] Hence there is increased teacher mortality and absenteeism (stemming from ill-health due to AIDS-related sickness episodes and funeral attendance). The result is a reduction in the available supply of teachers in many African countries. The empirical work will try to test explicitly for this reverse causation between infection rates and education levels.[3]

Quantity of students versus type of education

Separate from the issue of how the total number of students educated affects HIV/AIDS infection rates is the question as to whether the type of student educated is important. We can distinguish students by stage of education, secondary or primary, and within the primary sector according to whether it is gross (consisting of all age groups) or net (covering only those who attend who are of the official school age). The gross primary ratio can be greater than 100% because the numerator adds to those enrolled of the official primary school age pupils above and below the official age, and also students who repeat years, while the denominator consists only those of the official primary school age. The net primary ratio has 100% as a maximum. The denominator is the same as for the gross counterpart, but the numerator

[2] As one graphic example they report that, in the Central African Republic, 85% of teachers who died between 1996 and 1998 were HIV-positive, dying on average 10 years before they were due to retire.
[3] Data on the number of teachers for the sample of counties was very incomplete, so this variable could not be used explicitly in the estimation.

consists only those of official school age who are enrolled.

The fact that the gross primary figures are always higher than the net primary figures is worth considering further. The gross primary percentage includes older (and some younger) students as well as those who are less able. One could argue that the behaviour of these students, which can be called non-standard students, is likely to be different from the standard students who are of the official age.[4] In particular, it could be that having these non-standard female students in the classroom, as opposed to being outside the school and vulnerable to the sexual advances of adults, would have a different impact on infection rates than educating only the standard student.

To isolate these non-standard students one needs to subtract the net primary ratios from the gross primary ratios. This will be one of three education difference variables that will be tested. The others involve the differences between the two primary rates and the secondary enrolment rate. If the various education level variables have differential impacts on infection rates, then substituting one type of education for another, in the form of one of these education difference variables, will have a net impact on the spread of HIV/AIDS.

Choice of data, selection of the sample, and list of all the educational variables

Most country studies rely on national averages covering males and females. Given the intended emphasis on how female education impacts the transmission of HIV-AIDS, infection rate data just for female adults, aged 15–49 will be used. Although education data that cover ages earlier than 15 will be used, the epidemic has existed for nearly two decades in Africa and so a cross-section study as at the end of 2001 can capture the long-run relation between teen and pre-teen education and adult infection rates.

The infection measure relates to country HIV-AIDS prevalence rates, the total number of cases over all years. For most purposes, say to monitor the effectiveness of particular interventions, it is the number of new cases each year, i.e., incidence, that is more useful and this information was not available. However, in the context of a cross-section analysis of HIV-AIDS, using prevalence measures is not such a drawback as annual infection reporting standards are likely to vary by country and change over time. If the infection rates were indeed precise annual estimates, and either a time series or panel data analysis were being utilized, then it would very much matter which year the education data related. Matching infection rates and education levels by year would then be essential. By working with data on stocks and not flows one can accommodate the incomplete education data that are available for Africa, as it is these data that the study had the most difficulty in obtaining conformability across countries.

Even though there were 40 Sub-Saharan African countries for which UNAIDS had data on adult female infection rates, only 31 countries' experience could be used in the cross-sectional analysis due to education data availability. To ensure that the study controlled for the different duration that each country accumulates the stock measures, the age of the epidemic and the age of the epidemic squared are specified as independent variables alongside the education variables. This quadratic specification has the additional advantage of allowing one to predict the turning point for the HIV/AIDS infection rate for the sample of countries.

In all, 12 different education variables are tested. General female literacy rates are covered as well as primary (gross and net) and secondary school enrolment rates. For each category of education, differences (called gaps) between male and female levels are used as separate variables. The idea of a non-standard student is introduced by including the difference between gross and net primary rates. Then variables are formulated that pick up switching enrolment from one type of education (secondary, primary, and non-standard) to another.

III. Three Alternative Estimation Frameworks

The estimation will be carried out in terms of a single equation and a pair of equations considered both separately and jointly. For convenience the subscript is suppressed for the particular country being considered.

Single equation estimation

The Over study based estimation on a single equation.[5] He regressed the infection rate I on a

[4] Seftor and Turner (2002) recently found that older, non-traditional students in the USA react very differently to government educational subsidies than did the traditional students. Similarly, the category of non-standard students may also be thought to behave differently, albeit in a social context, to educational opportunities.

[5] Strictly, Over had two groups (a high-risk and a low-risk) and did some estimation where the two groups interact. His main results took the form of pooled estimates and for this a single equation was used.

number of independent variables identified earlier. The education variable is designated as E_j, and the non-educational variables by the set X. The single equation takes the form:

$$I = \alpha X + \beta_j E_j + \varepsilon_I \quad (1)$$

Where ε_I is the random error term for the infection equation, and the set α (which for convenience here includes the constant term) and the β_j (depending on the education variable specification) are fixed coefficients.

Two-equation recursive

In the attempt to apply the single equation framework for alternative specifications for the education variable it was found that E_j was a dominant variable. Most of the Over variables lost significance. The non-educational ones that retained significance can again be designated as X variables.[6] The ones that were so highly correlated with an education variable that one or other became insignificant were assigned to be Z variables. Since these Z variables could not be included with the E_j in the one equation, a second equation was specified to use these excluded variables as determinants (and consequently instruments) for the education variable. The two-equation system was formulated as:

$$I = \alpha X + \beta_j E_j + \varepsilon_I \quad (2a)$$

$$E_j = \gamma Z + \varepsilon_E \quad (2b)$$

where ε_E in Equation 2b is the random error term for the education variable and γ is the set of fixed coefficients for the Z variables. In this recursive framework, the Z variables have their impact on infection rates by first determining the education variables and these education variables subsequently influencing the infection rates.

Two equation simultaneous

In line with the World Bank group that focused on the relationship between education and HIV/AIDS, one needs to allow for the possibility in testing that infection rates can also affect the education variables. The estimation framework becomes simultaneous:

$$I = \alpha X + \beta_j E_j + \varepsilon_I \quad (3a)$$

$$E_j = \gamma Z + \delta I + \varepsilon_E \quad (3b)$$

Equation 3a has the same form as Equation 2a. Equation 3b differs from Equation 2b by the inclusion of the infection rate in the former equation. In Equation 3b, δ is the fixed coefficient indicating the feedback effect of infection rates on education.

Differences in types of education as independent variables

In order to allow for the policy option whereby one education level is substituted for another, the education equation for each of the three estimation frameworks just outlined can be respecified with an education difference replacing an education level or gap, such that the education equation becomes:

$$I = \alpha X + \beta_{jk}(E_j - E_k) + \varepsilon_I$$

Here β_{jk} is the coefficient attached to the particular education difference $E_j - E_k$.

IV. The Estimation Equations

First the dependent and independent variables are specified. Because a quadratic specification is used for one of the independent variables, the paper explains how an estimate for the turning point for the epidemic can be obtained. Then it indicates how the equations are to be estimated. The final part summarizes the data used and lists the sample.

Dependent variable

Over based his dependent variable on actual, average infection rates. As it is female education that is to be examined, it makes sense to define the infection rates pertaining to this gender only. Also the study departs from Over by using expected rather than actual infection rates. The dependent variable is based on:

$$\text{Female infection rate } (I_f) = \frac{\text{Expected number of adult females infected}}{\text{Total number of adult females}}$$

As emphasized by Over, the path of infection rates is likely to be nonlinear and take an S shape. So the logistic transformation L is also used, such that the dependent variable becomes:

$$L = \log_e\left[\frac{I_f}{(1 - I_f)}\right]$$

[6] Only six of Over's eight significant variables could be tested because there was no data on the percentage foreign born, and the variable male-female gender gap in urban areas was not relevant for the study, as the dependent variable was the national infection rate and not just that in urban areas. The national urbanization percentage and the rate of growth of urbanization were tried but both these measures were insignificant.

In a dynamic model of an infectious disease as formulated by Gersovich and Hammer (2001) it is important to specify how infections spread from individual to individual. Their methods are followed by using the random matching mechanism.[7] The expected number of adult female infections (on the numerator of I_f) would then be given by the product of the number of females susceptible S_f and the probability of a female getting infected from a male partner π, i.e., $S_f * \pi$, where the probability is simply the male infection rate I_m:

$$\pi = \frac{\text{Number of infected males}}{\text{Total number of adult males}} = I_m$$

Independent variables

The only X variable of those found significant by Over that could be included with the education variables in the sample was the age of epidemic A. However, because this variable was thought to affect infection rates non-linearly (increasing at a decreasing rate), it was added as an additional independent variable with a negative sign the square of A. The infection rate equation therefore took the form:

$$L = \alpha_0 + \alpha_1 A - \alpha_2 A^2 + \alpha_3 E_j + \varepsilon_I \tag{4a}$$

The quadratic nature of the age of the epidemic variable in Equation 4a can be utilized to find the implied turning point for the infection rate in the sample of countries. The derivative of the infection rate with respect to the age of the epidemic is given by:

$$\frac{\partial \hat{I}f}{\partial A} = \hat{I}_f(1 - \hat{I}_f)(\alpha_1 - 2\alpha_2 A) \tag{5}$$

Because $\hat{I}_f(1 - \hat{I}_f)$ is simply a number that depends on the particular values for the independent variables chosen and the magnitudes of the regression coefficient estimates, when Equation 5 was set equal to zero, this number will simply be divided out. Hence, the turning point for the infection rate I_f is the same as the turning point for the logit L in Equation 4a, i.e., $A = \alpha_1/2\alpha_2$. The logit coefficient estimates for age of infection and age of infection squared can therefore be used to estimate the epidemic's turning point. The second derivative is:

$$\frac{\partial^2 \hat{I}_f}{\partial A^2} = \hat{I}_f(1 - \hat{I}_f)(-2\alpha_2) + (\alpha_1 - 2\alpha_2 A)\frac{\partial \hat{I}_f(1 - \hat{I}_f)}{\partial A} \tag{6}$$

This is clearly negative at the turning point $\alpha_1 - 2\alpha_2 A = 0$. So the estimated turning point will always indicate a maximum infection rate.

There were two X variables that could not be included in the infection equation and they now become Z variables in the education equation. These two variables were: Z_1, the log of per capita income of a country, and Z_2, the percentage of the population that were Muslim. The second equation in the recursive framework was:

$$E_j = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon_E \tag{4b}$$

The only difference that the simultaneous estimation framework makes is that the logit infection rate is added to the right-hand side of Equation 4b to produce:

$$E_j = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_I L + \varepsilon_E \tag{4c}$$

Estimation methods

Four different estimation techniques will be applied in order to accommodate the three alternative estimation frameworks outlined in Section II. It is the infection Equation 4a that one is chiefly interested in estimating. The education Equations 4b and 4c come into play only when one is considering the estimation of Equation 4a as a pair with an education equation.

In a single-equation framework, OLS estimation is appropriate to estimate the coefficients in Equation 4a. In a recursive set up, Two-Stage Least Squares (2SLS) provides an unbiased estimator. The variables Z_1 and Z_2 in Equation 4b would be the relevant instruments for E_j in Equation 4a. To allow for the possibility that the two error terms ε_I and ε_E are correlated, Equations 4a and 4b will be estimated jointly using Zellner's Seemingly Unrelated Estimator (SURE). Finally, when the infection and education equations are considered simultaneously, Three-Stage Least Squares estimation (3SLS) will be used. This estimates the two equations (Equations 4a and 4c)

[7] Philipson and Posner (1993) consider that assortative matching is more plausible, whereby people choose a partner from the same risk class as they themselves are in. Given the aggregate nature of the data that are used, it is prefered to adopt the simpler random matching assumption to form the dependent variable as the expected number of female infections. Dow and Philipson (1996) find that, for the USA, an HIV-positive person is twice as likely as an HIV negative person to have an HIV-positive partner, which leads to infection predictions being one-third higher with random than assortative matching. It would then be likely that the expected prevalence figures over-estimate the true number. However, the results if one uses the actual number of infections are very similar to those found using the expected infection rates, so the results are not affected by the choice of matching assumption.

jointly and allows for the correlation between the two error terms. Both Equations 4a and 4c exclude two predetermined variables and include two endogenous variables minus 1, so the order condition for identification is satisfied (i.e., the two equations are overidentified).[8]

Data and sample

There are 48 Sub-Saharan African countries on the World Bank list. UNAIDS provided fact sheets for all of them as of the end of 1999. However, only for 40 of the countries did UNAIDS give information on female adult infection numbers. These numbers were converted into infection rates by dividing by the total number of females aged 15–64 as given in the World Bank's (2001) World Development Indicators tables for 2000. The UNAIDS fact sheets were the main source for most of the variables examined by Over that were tried. The GNP per capita figures were for 1997. The literacy rates for males and females were for 1995, and the secondary school enrolment rates were for 1996.

Primary school enrolment rates had to come from a different source, the World Bank's African Development Indicators (2002). Unfortunately these data had many deficiencies. Only for 22 of the countries were there primary enrolment figures for all four categories (male, female, gross and net) for exactly the same period, 1994–1997, in order that one could calculate the education gap and difference variables on a consistent basis. This study was able to match up a further nine countries provided that one were willing to go back to earlier periods, some for 1993 and the others for 1985. Note that all nine of these countries did turn out to have gross enrolment rates higher than net rates when the years were matched. So data on non-standard primary students that were positive were ensured. In all then, there were 31 countries in the final sample. For this sample, the percentage of the total population that were Muslim in 1998 was obtained from the web.[9] The 31 Sub-Saharan African countries in the sample are listed in Table 1. Table 2 gives a summary of the data for the dependent variable, the 12 education variables and the four non-educational independent variables.

Table 1. List of countries

Benin	Botswana	Burkina Faso	Burundi
Cameroon	Chad	Congo	Cote d'Ivoire
Dem. Rep. Congo	Djibouti	Ethiopia	Gambia
Guinea	Kenya	Lesotho	Madagascar
Malawi	Mali	Mauritania	Mozambique
Namibia	Niger	Rwanda	Senegal
Sierra Leone	South Africa	Swaziland	Tanzania
Togo	Uganda	Zambia	

V. Results

In every one of the 3SLS pair of estimates, the infection variable L was insignificant as a right-hand side variable in the second equation with education as the dependent variable. Thus the relevant estimation framework was found to be recursive rather than simultaneous. Consequently, most of the efforts are directed to reporting the infection equation results.

This study starts with the results for the education variables in the infection equation and then proceeds to the findings for the non-educational variables, which include the estimates of the turning points for the infection rate. For reference purposes, Tables 3–6 give estimates using OLS, 2SLS, SURE and 3SLS, respectively, for the nine level and gap educational variables. Tables 5 and 6 relate to estimation of the two-equation system, and therefore are split into (a) and (b) parts, with the first part related to Equation 4a and the second related to Equation 4b or 4c depending on whether one is dealing with the SURE or 3SLS estimates. Tables 7 and 8 give the results for the remaining three educational difference variables. Table 7 presents the OLS and 2SLS results, and Table 8 the SURE and 3SLS results. Table 9 has the turning point estimates that correspond to every education equation estimated. All tables have the absolute values of t-statistics in parentheses.

The educational variables

All four estimation methods produce the result that the level of female education is positively related to HIV/AIDS infection levels. This result is confirmed by the male-female gap coefficient estimates. These are all negative, which means that, holding male

[8] The rank condition for identification is easily satisfied as there is required to be at least one non-zero determinant of order one from the coefficients excluded from one equation and included in the other. So as long as one of β_1 and β_2 are significant, Equation 4a will satisfy the rank condition; and one of α_1 and α_2 being significant will satisfy the rank condition for Equation 4b.

[9] The website address was: http://faculty.stcc.tn.edu/bmcclure/Links/musliompop.htm.

Table 2. Description of the data ($N = 31$)

Variable	Mean	Standard deviation	Minimum	Maximum
Dependent variable: I				
Logit of infections (expected)	−2.95	1.22	−6.75	−1.34
Education variables: E_j				
Adult female literacy (%)	43.31	22.41	7	82
M–F literacy gap (% points)	20.29	8.03	0	35
Female secondary (%)	21.00	21.63	4.10	91.40
M–F secondary gap (% points)	6.07	8.36	−14.90	25.30
Female gross primary (%)	73.32	33.28	23.00	133.50
M–F gross primary gap (% points)	13.94	12.56	−12.00	41.00
Female net primary (%)	57.13	25.29	19.00	104.00
M–F net primary gap (% points)	8.58	10.19	−12.00	33.00
Non-standard primary (% points)	16.19	10.20	3.00	40.50
Gross primary minus secondary (% points)	52.33	22.80	18.20	114.80
Net primary minus secondary (% points)	36.13	17.51	12.00	91.80
Secondary minus non-standard primary (% points)	4.80	17.81	−16.30	69.40
Other variables: X and Z				
A: Age of epidemic (years)	13.87	2.17	6	18
A^2: Age of epidemic squared (years^2)	196.97	4.71	36	324
Z_1: GNP per capita (log)	5.98	0.90	4.70	8.10
Z_2: Percentage Muslim	45.56	34.21	1	100

Table 3. Education as a determinant of the logit of infections (OLS, $N = 31$)

Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X variables									
Constant term	−17.37	−12.57	−13.30	−13.94	−14.58	−14.06	−14.34	−14.72	−14.17
	(6.535)	(4.368)	(5.440)	(5.253)	(5.595)	(5.355)	(5.562)	(6.062)	(5.033)
Age of epidemic	1.787	1.283	1.230	1.464	1.315	1.494	1.289	1.607	1.286
	(4.483)	(2.768)	(3.266)	(3.495)	(3.352)	(3.595)	(3.305)	(4.183)	(3.038)
Age of epidemic^2	−0.059	−0.038	−0.037	−0.046	−0.039	−0.046	−0.039	−0.051	−0.037
	(3.744)	(2.099)	(2.495)	(2.773)	(2.557)	(2.833)	(2.539)	(3.380)	(2.237)
E variables									
Adult female literacy	0.029								
	(4.046)								
M−F literacy gap		−0.031							
		(1.334)							
Female secondary			0.049						
			(3.624)						
M−F secondary gap				−0.055					
				(2.889)					
Female gross primary					0.015				
					(3.270)				
M–F gross primary gap						−0.039			
						(3.033)			
Female net primary							0.020		
							(3.300)		
M–F net primary gap								−0.059	
								(3.990)	
Non-standard primary									0.039
									(2.366)
$\overline{R}^2$	0.597	0.393	0.565	0.506	0.537	0.518	0.539	0.593	0.464

Table 4. Education as a determinant of the logit of infections (2SLS, $N = 31$)

Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X variables									
Constant term	−20.15	−13.61	−13.53	−16.66	−16.24	−17.07	−15.91	−17.35	−17.13
	(6.916)	(3.975)	(5.856)	(4.701)	(5.948)	(4.649)	(5.845)	(5.507)	(5.070)
Age of epidemic	2.158	1.834	1.257	2.044	1.467	2.137	1.429	2.140	1.570
	(5.016)	(3.126)	(3.536)	(3.477)	(3.626)	(3.471)	(3.523)	(4.108)	(3.190)
Age of epidemic^2	−0.074	−0.062	−0.038	−0.070	−0.045	−0.071	−0.045	−0.072	−0.047
	(4.348)	(2.657)	(2.736)	(2.986)	(2.870)	(2.961)	(2.825)	(3.523)	(2.446)
E variables									
Adult female Literacy	0.044								
	(4.651)								
M–F literacy gap		−0.124							
		(2.761)							
Female secondary			0.030						
			(4.207)						
M–F secondary gap				−0.144					
				(3.031)					
Female gross Primary					0.026				
					(4.187)				
M–F gross primary gap						−0.103			
						(2.996)			
Female net primary							0.035		
							(4.183)		
M–F net primary gap								−0.121	
								(3.626)	
Non-standard primary									0.097
									(3.421)
$\overline{R}^2$	0.526	0.027	0.556	0.117	0.446	0.078	0.437	0.328	0.206
Sargan value	0.320	2.784	5.185	2.352	1.136	2.099	0.916	2.121	1.279

education levels constant, an increase in female education will increase the infection rates. Of all the nine education variables, only the gap in male-female literacy fails to have a coefficient that is significant at least the 1% level (and then just for the OLS estimates).

For the three education difference variables, which one was significant depended on the estimation technique used, though the signs were the same throughout Tables 7 and 8. All three educational differences revolved around a substitution involving the female secondary school variable. For the OLS and 2SLS estimates in Table 7, only a substitution from secondary school to non-standard enrolments would significantly lower infection rates; substitutions from gross primary, or net primary, to secondary school were both insignificant. All three differences were significant in the SURE estimates in Table 8. A substitution away from secondary to non-standard lowers infection rates as before. However, it would be a substitution towards secondary and away from primary (either gross or net) that would lower infections in this case. The direction of these substitutions was confirmed by the 3SLS estimates in Table 8, except that this time the net primary minus secondary difference was not significant.[10]

The non-educational independent variables

The coefficients for A and A^2 were significant above the 1% level in every equation estimated. A always had a positive sign, and A^2 always had a negative sign. The infection rate thus had the expected rising and then falling shape with respect to the age of the epidemic. The coefficients for A and A^2 estimated by all four techniques are collected and presented in Table 9. The last column of the table reveals that the turning point for infections corresponding to the A and A^2 coefficient estimates lie in the narrow band of 16 ± 1. The mean age of the epidemic in the sample is around 14 years. On average, there is predicted to be two more years before country infection rates start to fall. The highest A for any country was 18 years, so some countries should already be experiencing reduced infection rates.

[10] In the 2SLS estimates for all 12 education variables, the Sargan values are sufficiently low that one cannot reject the null that the four independent variables A, A^2, Z_1 and Z_2 were not good instruments.

Table 5a. Education as a determinant of the logit of infections (SURE, Equation 1, $N = 31$)

Independent variables	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(7a)	(8a)	(9a)
X variables									
Constant term	−16.85	−10.75	−14.08	−13.91	−16.05	−13.53	−16.18	−14.41	−15.78
	(7.717)	(5.561)	(6.638)	(7.175)	(7.866)	(7.064)	(8.033)	(7.883)	(7.690)
Age of epidemic	1.647	1.373	1.345	1.619	1.479	1.650	1.488	1.714	1.478
	(5.054)	(4.382)	(4.108)	(5.260)	(4.797)	(5.417)	(4.858)	(5.902)	(4.765)
Age of epidemic^2	−0.055	−0.043	−0.042	−0.053	−0.048	−0.054	−0.048	−0.057	−0.048
	(4.235)	(3.470)	(3.241)	(4.391)	(3.042)	(4.530)	(3.994)	(4.971)	(3.941)
E variables									
Adult female literacy	0.041								
	(6.507)								
M–F literacy gap		−0.135							
		(7.647)							
Female secondary			0.032						
			(4.939)						
M–F secondary gap				−0.158					
				(10.43)					
Female gross primary					0.027				
					(6.533)				
M–F gross primary gap						−0.115			
						(11.27)			
Female net primary							0.036		
							(6.686)		
M–F net primary gap								−0.130	
								(10.68)	
Non-standard primary									0.110
									(7.974)
$\bar{R}^2$	0.525	<0	0.549	<0	0.414	<0	0.417	0.220	0.021

Table 5b. Determinants of the education variables (SURE, Equation 2, $N = 31$)

Independent variables	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)	(7b)	(8b)	(9b)
Z variables									
Constant term	17.17	33.15	−63.26	13.59	10.91	23.61	13.29	16.70	1.493
	(1.108)	(6.715)	(5.461)	(2.426)	(0.491)	(2.992)	(0.781)	(2.393)	(0.226)
GNP per capita (log)	7.558	−2.943	16.25	−2.073	14.60	−2.757	10.537	−2.388	3.506
	(3.119)	(3.834)	(8.960)	(2.406)	(4.173)	(2.280)	(3.939)	(2.213)	(3.397)
Per cent Muslim	−0.419	0.104	−0.285	0.107	−0.548	0.150	−0.422	0.135	−0.138
	(6.240)	(5.303)	(6.126)	(4.648)	(5.770)	(4.582)	(5.768)	(4.638)	(4.924)
$\bar{R}^2$	0.570	0.325	0.806	0.208	0.553	0.174	0.541	0.237	0.327

The two independent variables in the education equation were always highly significant no matter which education variable was used. The percentage Muslim had a negative effect on all nine of the education level and gap variables. Given that it was also found that education increased infection rates, the negative sign for Z_1 was consistent with the Over result that countries with a higher share of the population that was Muslim had lower HIV/AIDS infection rates. What was not consistent with the Over work was the positive sign found attached to the per capita income of a country Z_2 in the education equation. Over found that the higher a country's income level, the lower would be its infection rate. In the present case though, higher income levels leads to more education which, in turn, would then increase infection rates.

VI. Summary and Discussion

The answer to the question as to whether female education prevents the spread of HIV-AIDS in Sub-Saharan Africa is: No, not at the present time. The results do not therefore support previous empirical evidence and is in contradiction to the strong a priori expectation that female education will lead to improed empowerment in helping to control sexual relations.

In a large part this unexpected result is due to the fact that the impact of education on HIV/AIDS

Table 6a. Education as a determinant of the logit of infections (3SLS, Equation 1, $N = 31$)

Independent variables	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(7a)	(8a)	(9a)
X variables									
Constant term	−19.38	−12.31	−13.75	−15.78	−16.83	−16.60	−16.32	−16.82	−18.40
	(6.851)	(3.882)	(6.062)	(4.832)	(5.331)	(4.860)	(6.249)	(5.715)	(5.505)
Age of epidemic	2.013	1.527	1.305	1.851	1.581	2.035	1.509	2.026	1.811
	(4.879)	(2.993)	(3.810)	(3.591)	(4.140)	(3.716)	(4.033)	(4.332)	(3.873)
Age of epidemic^2	−0.068	−0.048	−0.040	−0.061	−0.051	−0.067	−0.049	−0.067	−0.058
	(4.174)	(2.486)	(3.066)	(3.108)	(3.513)	(3.250)	(3.472)	(3.778)	(3.319)
E variables									
Adult female literacy	0.043								
	(4.570)								
M–F literacy gap		−0.117							
		(2.678)							
Female secondary			0.304						
			(4.256)						
M–F secondary gap				−0.139					
				(3.017)					
Female gross primary					0.026				
					(4.302)				
M–F gross primary gap						−0.101			
						(3.012)			
Female net primary							0.036		
							(4.299)		
M–F net primary gap								−0.118	
								(3.642)	
Non-standard primary									0.103
									(3.523)
$\overline{R}^2$	0.536	0.055	0.554	0.152	0.541	0.103	0.425	0.348	0.142

Table 6b. Determinants of the education variables (3SLS, Equation 2, $N = 31$)

Independent variables	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)	(7b)	(8b)	(9b)
Z variables									
Constant term	−4.885	38.28	−59.33	20.88	−4.859	38.75	10.48	28.12	−11.71
	(0.216)	(4.245)	(4.109)	(1.905)	(0.155)	(2.331)	(0.445)	(2.121)	(1.028)
GNP per capita (log)	9.259	−3.542	15.96	−2.851	16.06	−4.221	10.83	−3.472	4.736
	(3.075)	(2.885)	(8.226)	(1.924)	(3.847)	(1.882)	(3.479)	(1.954)	(3.117)
Per cent Muslim	−0.506	0.122	−0.278	0.138	−0.593	0.212	−0.431	0.182	−0.175
	(5.346)	(3.526)	(4.777)	(3.158)	(4.446)	(3.169)	(4.232)	(3.337)	(3.652)
Logit of infections	−5.373	0.797	0.835	1.369	−3.089	3.125	−0.509	2.397	−2.551
	(1.584)	(0.548)	(0.387)	(0.771)	(0.620)	(1.163)	(0.132)	(1.125)	(1.393)
$\overline{R}^2$	0.533	0.329	0.798	0.135	0.542	0.065	0.525	0.108	0.341

is very much tied up with income and religion in Sub-Saharan Africa, at a time when: countries with higher incomes have higher infection rates, smaller Muslim populations and greater education participation; and countries with a larger Muslim population have lower infection rates, smaller incomes and lower education participation. Of course, if one regards religion as an education variable, rather than a general cultural proxy, then one would view the results differently. Either way, the main policy conclusion of the study is that, in order to evaluate particular education policy interventions in Sub-Saharan African countries, one needs to have samples of infection rates with and without the intervention that include eight categories: low and high income individuals split into Muslims and non-Muslims, and for each of these categories those who are educated and those who are not.

One possible reason why Over found a positive sign for income and a negative sign for education in his infection equation, while the present study found the opposite, may be due to the difference in the sample of countries used. Over used a mix of Latin American and African countries, and the present study concentrated on Sub-Saharan African countries. Some Latin American countries, such as Brazil and Argentina, have much higher income levels and could thus be classed as middle-income developing

Table 7. Difference in type of education as a determinant of the logit of infections (OLS, 2SLS)

Independent variables	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
X variables						
Constant term	−13.31	−12.44	−12.05	−20.87	−14.20	−11.92
	(4.427)	(4.117)	(4.598)	(3.053)	(3.722)	(4.578)
Age of epidemic	1.187	1.115	1.110	1.792	1.234	1.117
	(2.638)	(2.433)	(2.744)	(2.033)	(2.446)	(2.778)
Age of epidemic^2	−0.034	−0.031	−0.033	−0.056	−0.036	−0.034
	(1.915)	(1.732)	(2.064)	(1.633)	(1.809)	(2.173)
E differences						
Female gross primary minus	0.010			0.077		
Female secondary	(1.245)			(1.962)		
Female net primary minus		0.003			0.032	
Female secondary		(0.332)			(0.900)	
Female secondary minus			0.025			0.042
Female non-standard primary			(2.733)			(3.493)
$\bar{R}^2$	0.388	0.356	0.494	0.094	0.234	0.457
Sargan value	na	na	na	0.788	10.66	5.354

Table 8. Difference in type of education as a determinant of the logit of infections (SURE, 3SLS)

Independent variables	SURE (1)	SURE (2)	SURE (3)	3SLS (4)	3SLS (5)	3SLS (6)
X variables						
Constant term	−18.60	−20.26	−14.00	−16.34	−15.56	−12.09
	(9.109)	(9.832)	(6.820)	(6.106)	(4.125)	(4.883)
Age of epidemic	1.546	1.547	1.409	1.512	1.459	1.154
	(5.002)	(4.984)	(4.416)	(3.872)	(3.083)	(3.217)
Age of epidemic^2	−0.051	−0.051	−0.044	−0.049	−0.046	−0.036
	(4.155)	(4.130)	(3.540)	(3.262)	(2.651)	(2.743)
E differences						
Female gross primary minus	0.008			0.036		
Female secondary	(14.71)			(4.268)		
Female net primary minus		0.162			0.039	
Female secondary		(23.26)			(1.072)	
Female secondary minus			0.050			0.043
Female non-standard primary			(6.125)			(3.533)
$\bar{R}^2$	0.000	0.000	0.334	0.425	0.052	0.423

countries. Middle-income countries have both greater education levels and smaller HIV-AIDS infection rates than low-income countries. So the difference in findings might simply be that, in Over's study, raising education levels means moving from a low-income to a middle-income group, while, the present study is estimating the impact of raising education levels within a low-income group of countries.

None the less, this study has the common weakness with the Over study that it measures HIV/AIDS infection rates using prevalence data and not that based on incidence. There may well be lag effects that the cross-section analysis was not able to pick up. For example, Parkhurst (2002) reports that although prevalence rates eventually will reflect declines in incidence, they will do so after a time lag of seven years. This point needs to be especially borne in mind when one examines the time dimension of the results.

Is the result of a positive relation between education and HIV/AIDS infection rates really new? In the early days of the epidemic, when the transmission mechanism was not well understood, there was evidence that education and HIV/AIDS was positively related. But, this was thought not to be the case once educated people came more knowledgeable about the disease.[11] So the new part is that

[11] World Bank (2002) state: 'Data for the late 1980s and early 1990s, when the HIV/AIDS pandemic was just emerging, mostly showed a positive correlation between level of education and rates of infection. This was perhaps because the higher socioeconomic status and greater mobility of better educated people enabled encounters with a greater number and range of sexual partners, but also because at that time education seldom included HIV/AIDS prevention or behavioral change programs, and the level of knowledge about the disease was generally low. However, once the ways to avoid infection became better known, better educated people were more likely to adopt safer behavior (World Bank 1999), and later studies show a reversal in the trend.'

Table 9. Estimates of the peak of the epidemic

Equation	Coefficient on age of epidemic: α_1	Coefficient on age of epidemic squared: α_2	Year epidemic reaches peak: $A = \alpha_1/2\alpha_2$
OLS 1	1.787	0.059	15
OLS 2	1.283	0.038	17
OLS 3	1.230	0.037	17
OLS 4	1.464	0.046	16
OLS 5	1.315	0.039	17
OLS 6	1.494	0.046	16
OLS 7	1.289	0.039	17
OLS 8	1.607	0.051	16
OLS 9	1.286	0.037	17
2SLS 1	2.158	0.074	15
2SLS 2	1.834	0.062	15
2SLS 3	1.257	0.038	17
2SLS 4	2.044	0.070	15
2SLS 5	1.467	0.045	16
2SLS 6	2.137	0.071	15
2SLS 7	1.429	0.045	16
2SLS 8	2.140	0.072	15
2SLS 9	1.570	0.047	17
SURE 1a	1.647	0.055	15
SURE 2a	1.373	0.043	16
SURE 3a	1.345	0.042	16
SURE 4a	1.619	0.053	15
SURE 5a	1.479	0.048	15
SURE 6a	1.650	0.054	15
SURE 7a	1.488	0.048	16
SURE 8a	1.714	0.057	15
SURE 9a	1.478	0.048	15
3SLS 1a	2.013	0.068	15
3SLS 2a	1.527	0.048	16
3SLS 3a	1.305	0.040	16
3SLS 4a	1.851	0.061	15
3SLS 5a	1.581	0.051	16
3SLS 6a	2.035	0.067	15
3SLS 7a	1.509	0.049	15
3SLS 8a	2.026	0.067	15
3SLS 9a	1.811	0.058	16
Average	1.618	0.052	16

the old positive relation has returned to Sub-Saharan Africa and public policies must be adjusted accordingly. This does not, of course, mean that the development goal of raising female education should be abandoned; it means only that the type of education that is being expanded must be carefully evaluated.

The research in this study highlighted two considerations as they relate to the issue of the type of education that can impact HIV/AIDS infection rates. First, there is the age-old controversy concerning whether to promote primary or secondary education in developing countries. The results come out on the side of secondary education, as raising enrolment in this sector at the expense of enrolments in primary education (gross or primary) would tend to lower HIV/AIDS infection rates. Second, it is suggested that attention be focused, and resources be directed to, an alternative category of female student, i.e., the non-standard primary enrollees. These students may be older and, the repeaters, less able than the conventional student of the official school age. Education as a private income investment may be less important for this category of student and thus the confounding effect of income of education less severe. The social return on recruiting these students may be higher than for conventional ones if attendance at school means that such females are 'kept out of harm's way'. Support for this view comes from the finding that a substitution from secondary to primary non-standard enrollees would lower country HIV/AIDS infection rates. Lastly, the methods enable one to estimate that female HIV/AIDS prevalence rates in Sub-Saharan countries peak within 15–17 years of the first case reported.

The main weakness of the research reported in this paper lies in the data that were available. Apart from the very high degrees of correlation among the education variables and the other independent variables, the study was limited by the type of education data that could be used. Many African countries had no education data. When data were available, they were only for particular years that were not common for all countries. The end result was that the study was able to estimate only the effect of the *level* of education on HIV/AIDS and could not analyse *changes* in education. It is the dynamic effects that would be the most useful to know. A country study could best reveal this kind of relationship.[12]

Acknowledgements

This paper was completed with the financial assistance provided by the Fulbright Research award for 2003 as part of the project on the Cost-Benefit Analysis of HIV-AIDS interventions programmes in Tanzania when the author was a visiting Professor in the Department of Economics at the University of Dar Es Salaam in Tanzania. The paper benefited from detailed comments during seminars at Fordham University in the USA and the University of Dar es Salaam. I wish to thank Daryl McLeod for suggesting that I analyse the turning point for the epidemic. Helpful comments were also received from Mary Beth Combs, Dereck Reagle and Parantap Basu.

[12] See for example, De Walque (2004).

References

De Walque, D. (2004) How does the impact of an HIV/AIDS information campaign vary with educational attainment? Evidence from rural Uganda, Policy Research Working Paper No 3289, World Bank, Washington, DC.

Dow, W. H. and Philipson, T. (1996) An empirical examination of the implications of assortative matching on the incidence of HIV, *Journal of Health Economics*, **15**, 735–49.

Gersovitz, M. and Hammer, J. S. (2004) The economical control of infectious diseases, *Economic Journal*, **114**, 1–27.

Over, M. (1998) The effects of societal variables on urban rates of HIV infection in developing countries, in *Confronting AIDS: Evidence from the Developing World* (Eds) M. Ainsworth, L. Fransen and M. Over, European Communities, UK.

Parkhurst, J. O. (2002) The Ugandan success story? Evidence and claims of HIV-1 prevention, *Lancet*, **360**, 78–80.

Philipson, T. and Posner, R. (1993) *Private Choices and Public Health: An Economic Interpretation of the AIDS Epidemic*, Harvard University Press, Cambridge, MA.

Seftor, N. S. and Turner, S. E. (2002) Back to school: federal student aid policy and adult college enrollment, *The Journal of Human Resources*, **37**, 336–52.

World Bank (1999) *Confronting AIDS: Public Priorities in a Global Epidemic*, Report, (Revised edn), Oxford University Press, Oxford.

World Bank (2002) *Education and AIDS*, World Bank, Washington, DC.