# A Cost-Benefit Analysis of Female Primary Education as a Means of Reducing HIV/AIDS in Tanzania

*[Forthcoming in Applied Economics]*

**Robert J. Brent***
**Department of Economics**
**Fordham University**
**Bronx, NY 10458**

## Abstract

This article uses panel data related to 20 Tanzania regions and 8 years to estimate the direct and indirect effects of female primary education on HIV/AIDS rates. A recursive framework for education, income and infections is employed, based on two autoregressive equations that allow us to obtain dynamic estimates of effectiveness. We find that the indirect effect working through changes in income outweighs the direct positive effect of education on infections, implying that female education can be effective as an intervention to lower the disease in Tanzania. The estimates of effectiveness are then utilized to carry out a cost-benefit analysis of the education expenditures. The human capital approach is used to measure the benefits. Irrespective of the timing of the benefits and costs, and the discount rate alternatives we consider, our best estimates result in positive net-benefits, with benefit-cost ratios in the range 1.3 to 2.9.

**Keywords:** Cost-Benefit Analysis; Female Primary Education; Preventing HIV/AIDS; Tanzania.

**JEL Classifications**: D61: Cost-Benefit Analysis. O22: Project Analysis. I12 : Health Production. I28 : Government Education Policy.

---

• This paper was completed with the financial assistance provided by the Fulbright Research award for 2003. It was part of the project on the Cost-Benefit Analysis of HIV-AIDS interventions programs in Tanzania carried out when the author was a visiting Professor in the Department of Economics at the University of Dar Es Salaam in Tanzania. I particularly wish to thank Dr. Longinus Rutasitara for all his support during my stay at the University and Dr. Nester Sekwao at the Ministry of Health and Culture for helping me collect and understand the education data. Dr. Bennet Fimbo of TACAIDS provided valuable advice concerning the health data.

## 1. Introduction

HIV-AIDS has affected Sub-Saharan Africa more than any other region, with the December 2003 estimate by UNAIDS being 26.6 million out of a world total of 40 million. Also disturbing is the fact that in 2003 there were more new cases in the region, 3.2 million, than persons who died during the year, 2.3 million. The disease is continuing to rise. One estimate of the consequences of this pandemic has it lowering life expectancies by a quarter, from 64 to 47 years.[1] Much of the development progress in the region in the last half-century is in danger of being reversed. Intervention is clearly required. The problem is to identify the type of interventions that are likely to be most beneficial. Many working in the field believe that all the existing HIV-AIDS interventions are necessary. However, this ignores the fact that many of the Sub-Saharan African countries are some of the poorest in the world, and their resources are particularly limited. Choices need to be made. If hospitals beds have half the patients in them with HIV-AIDS, then those suffering from other diseases must be denied care. Even when resources are donated from outside, it is not obvious that antiviral therapies are always going to have the greatest impact, as opposed to simply supplying vitamins and other nutritional supplements.[2] In order to establish priorities, cost-benefit analysis (CBA) is required. Using cost-effectiveness analysis, the evaluation technique most utilized by health care professionals is not sufficient, for it assumes that there is a fixed budget. There is a world-wide movement seeking greater funding for HIV-AIDS programs.[3] How best to spend these additional funds is the *raison d'être* for CBA.

[1] See Logie (1999).
[2] See Fawzi et al. (2004).

The issue as to which type of interventions to finance is especially important given the gender and age dimension of the Sub-Saharan African HIV-AIDS epidemic. The December 2003 UNAIDS update points out that African women are 1.2 times more likely to be infected with HIV than men. Among young people aged 15-24 (a good index of the number of new cases) women were 2.5 times more likely to be infected. Some of the gender differential is due to biological factors, as HIV is more easily transmitted from men to women than the other way round. But, some of the cause is due to the sociological power structure between the sexes in Sub-Saharan Africa. Interventions that seek to empower females, especially young females, would seem to be especially worth evaluating.

In this article we carry out a CBA of female primary education in Tanzania based on panel data for 20 Mainland regions over the period 1994-2001. Education is a central element in any country's development process. At a time when Sub-Saharan Africa's development itself is being threatened, it is important to confirm that one of its pillars is functioning as required. Female education has been identified by the World Bank (2002) as one of most promising ways of combating HIV-AIDS because of the empowerment it provides, in terms of delayed first-sex and income independence. Schooling clearly targets the young. In Tanzania, primary schooling is the predominate form. 61% of females in 2000/1 had been through some primary education, while only 4% has been to secondary school, and 1% had a diploma or degree.[4]

The center-piece of our analysis involves establishing the effectiveness of female primary enrollments to reduce HIV-AIDS and this takes place in section 2. This is the main contribution

[3] For example, the Commission on Macroeconomics and Health (2001) called for the assistance from developed nations to Sub-Saharan Africa and other nations to rise from the current levels of US$ 6 billion to US$ 27 billion by 2007 and US$ 35 billion by 2015.
[4] See Household Budget Survey 2000/01 (2002), table 4.2.

of the paper as there is an extensive literature (which we discuss below) which finds a positive relation between education and infection rates, contrary to most expectations. We treat this positive relation as the direct effect of schooling and regard it as coming from a static estimation framework. We then identify an indirect effect of education working through changes in income that has the opposite effect. Using a dynamic estimation approach developed by Arellano and Bond (1991) we are able to show that the indirect effect outweighs the direct effect, producing the net effect that education lowers infection rates. The CBA that follows in section 3 involves valuing the estimated number of HIV cases averted by their earnings, i.e., using the human capital approach, and comparing this with the costs for the enrollment numbers that generated the reduced number of infections. While the magnitude of the net-benefits depends crucially on the size of the discount rate used, our best estimate is that the benefits are between 1.3 to 2.9 times the costs. In the final section outlining the summary and conclusions we highlight the role of external effects of female education as being the mechanism generating the favorable cost-benefit outcome.

The data on infections that we use to make our effectiveness estimates is based on information on voluntary blood donors supplied to the Ministry of Health in Tanzania. Unlike the main alternative data source coming from pregnant women in antenatal clinics (ANCs), blood donor infection rates are collected for every region. Blood donor data has many well known defects. For example, in Tanzania, blood donors are typically family members of those who are going to receive the blood transfusion. So they may not be representative of the Tanzanian population as a whole. A nationally representative estimate of regional infection rates based on blood samples has recently been completed, i.e., the Tanzania HIV/AIDS Indicator Survey (2005). But this information is available just for a one year period during 2003-04 and so can only be the benchmark for future dynamic HIV studies. It should be noted that even this nationally representative sample does not include data on incomes and so would not be able to test the strength of the indirect effect of education on infections that is central for this article.[5]

## 2. Estimating Education Effectiveness.

In this section we lay out a recursive framework for determining education effectiveness and then express it in terms of estimation equations. Then we show the results of the estimation and proceed to transform the regression coefficients into estimates of the number of HIV cases averted.

We can think of the HIV-AIDS infection rate $H$ being determined by a production function that contains education $E$, income $Y$ and other factors $Z$: $H = H(E,Y,Z)$. Education is considered a policy variable in a developing country that is determined by the government and so is exogenous to $H$.[6] Income has its own production function with education and other factors $Z$ as determinants: $Y = H(E,Z)$. Let us assume that the $Z$ variables in both equations are independent of education.[7] The marginal effect of education on infections would be:

$$\frac{dH}{dE} = \frac{\partial H}{\partial E} + \frac{\partial H}{\partial Y}.\frac{dY}{dE}$$

The first term on the right-hand-side is the direct effect. The second term is the indirect effect, whereby education first affects income and, through this, it impacts infections. Linear approximations for the two production functions produce:

$$H = \alpha_0 + \alpha_E E + \alpha_Y Y \qquad (1a)$$

[5] One drawback of our blood donor data is that there exists large annual fluctuations in regional infection rates. However, this problem would not be removed if we relied on the ANC data instead as large annual fluctuations are present in this data source as well. For example, in the Iringa region (Regional Hospital site) the infection rate among clinic enrollees rose from 24.9% in 1998 to 40.1% in 2000 and then shrunk to 4.6% in 2000.

[6] It is illegal not to go to primary school in Tanzania.

[7] As we shall see in the empirical work, the only significant Z variables are time dummies and they are not functions of education.

$$Y = \beta_0 + \beta_E E \quad (2a)$$

and the marginal effect is given by:

$$\frac{dH}{dE} = \alpha_E + \alpha_Y \beta_E$$

The effectiveness of education can be determined by estimating the three parameters, $\alpha_E$, $\alpha_Y$ and $\beta_E$.

Although we will be adopting a modified version of the marginal effect expression, we can still use the simple version to explain the main issues that arise in determining education effectiveness. There are many empirical studies related to developing countries that have found a positive sign attached to $\alpha_E$, the direct effect of education on HIV-AIDS. In a survey of the literature by Hargreaves and Glynn (2002), only 1 of the 27 studies (related to Ethiopia) had a significant negative relation between education and infection. This positive relation was confirmed by a recent cross-section study of 31 Sub-Saharan African countries using 9 different education measures.[8] There are a number of plausible explanations for this positive sign, seeing that education is associated with higher socio-economic status, and the more educated are likely to live in urban areas, have more leisure time, travel more, and have greater mobility. Although there is some evidence that the magnitude of the positive effect may be declining over time,[9] at this point in history the direct effect can be expected to be positive.

However, effectiveness is not just determined by the direct effect. In this paper we identify and estimate an indirect effect. Education can be predicted to raise income ($\beta_E > 0$). So

[8] See Brent (2006a).

[9] The World Bank (1999), p130, cites research for Brazil that shows that in 1985, about 3/4 of those newly diagnosed with AIDS had a secondary or university education, and by 1994 this proportion had decreased to 1/3. The World Bank predicted that "eventually" the "positive relation in Africa" would be reversed. But, this is not to say that this has already occurred, hence Brent's (2006a) positive results for 2001. The first research to find that the

if income lowers infection rates ($\alpha_Y < 0$), it can be expected that the indirect effect is negative.[10] The indirect effect may outweigh the direct effect and thus lead to an overall negative effect of education on infection rates. We will now explain how we plan to estimate both the direct and indirect effects for Tanzania to see what the net effect is.

Our statistical analysis involves a panel, where $i$ ($i$ = 1, …, 20) is a Tanzanian region and $t$ ($t$ = 1, … , 8) is the year involved. The initial panel data estimation equations for the production functions are:

$$H_{it} = \alpha_0 + \alpha_E E_{it} + \alpha_Y Y_{it} + \alpha'_X X_i + \alpha'_Z Z_t + u_{it} \tag{1b}$$

$$Y_{it} = \beta_0 + \beta_E E_{it} + \beta'_X X_i + \beta'_Z Z_t + w_{it} \tag{2b}$$

where: the $\alpha$ and $\beta$ coefficients are constants to be estimated; the vector $X_i$ are region-specific characteristics; the vector $Z_t$ are time-specific dummy variables; and $u_{it}$ and $w_{it}$ are random error terms (assumed to be normally distributed). The next step is to convert this into a dynamic framework. Define $H_{it}^*$ as the long-run, equilibrium value for $H_{it}$ and use this to replace $H_{it}$ in equation (1b). We will assume a partial adjustment mechanism for $H_{it}$ to approach $H_{it}^*$: $H_{it} = \delta_H H_{it}^* + (1-\delta_H)H_{i,t-1}$, where $\delta_H$ is the partial adjustment coefficient. The dynamic version of the infection equation becomes:

$$H_{it} = \varphi_0 + \varphi_E E_{it} + \varphi_Y Y_{it} + \varphi_X X_i + \varphi_Z Z_t + \varphi_H H_{i,t-1} + \delta_H u_{it} \tag{1c}$$

where: $\varphi_0 = \delta_H \alpha_0$; $\varphi_E = \delta_H \alpha_E$; $\varphi_Y = \delta_H \alpha_Y$; $\varphi_X = \delta_H \alpha_X$; $\varphi_Z = \delta_H \alpha_Z$; $\varphi_H = (1-\delta_H)$.

---

decline over time in Africa is such that it reverses in sign and becomes negative is by De Walque (2004) related to Uganda.

[10] The sign of $\alpha_Y$ is an empirical issue as it may not be negative. Income in Sub-Saharan Africa acts much like education, as Brent (2006a) has found. It is the richest countries in Southern Africa that have the highest infection rates. In Tanzania, the level of income of a region was also positively related to the contemporaneous rate of infection. But, using the dynamic estimator we obtained the result that it is *increases* in income (not the levels themselves) that lower infection rates and we use this to produce a negative indirect effect.

Similarly, with $\delta_Y$ as the partial adjustment coefficient in $Y_{it} = \delta_Y Y_{it}^* + (1-\delta_Y)Y_{i,t-1}$, and $Y_{it}^*$ replacing $Y_{it}$ in equation (2b), the dynamic version of the income equation is:

$$Y_{it} = \gamma_0 + \gamma_E E_{it} + \gamma_X X_i + \gamma_Z Z_t + \gamma_Y Y_{i,t-1} + \delta_Y w_{it} \qquad (2c)$$

with: $\gamma_0 = \delta_Y \beta_0$; $\gamma_E = \delta_Y \beta_E$; $\gamma_X = \delta_Y \beta_X$; $\gamma_Z = \delta_Y \beta_Z$; and $\gamma_E = (1-\delta_Y)$.

The marginal effect of education now appears as:

$$\frac{dH}{dE} = \frac{\varphi_E}{\delta_H} + \frac{\varphi_Y}{\delta_H} \cdot \frac{\gamma_E}{\delta_Y}$$

By having an autoregressive term in them, equations (1c) and (2c) are in a form to be estimated by the Arellano-Bond, Method of Moments technique. This is a fixed effects approach that uses first differences to eliminate the individual regional effects $X_i$. Defining first differences in the usual way, for example, $\Delta H_{it} = H_{it} - H_{i,t-1}$ and $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$, the two production functions are estimated using:

$$\Delta H_{it} = \varphi_E \Delta E_{it} + \varphi_Y \Delta Y_{it} + \varphi_Z \Delta Z_t + \varphi_H \Delta H_{i,t-1} + \delta_H \Delta u_{it} \qquad (1d)$$

$$\Delta Y_{it} = \gamma_E \Delta E_{it} + \gamma_Z \Delta Z_t + \gamma_Y \Delta Y_{i,t-1} + \delta_Y \Delta w_{it} \qquad (2d)$$

In the estimation, instruments are used for all the right-hand side variables in (1d) and (2d). The lagged dependent variables are treated as endogenous and lagged values are used as instruments for them. The number of instruments for the dependent variable is *T*–*p*–2, where *T* is the number of years and *p* is the number of lags for the dependent variable. For exogenous variables, the first differences serve as the instruments themselves. This clearly applies to the time dummies. In the first instance we will test all the right-hand side variables (other than the lagged dependent variables) as if they were exogenous. When variables are regarded as predetermined, the number of lagged values that can be used as instruments is $T$–$p$–1. If a variable is treated as endogenous, the number of instruments again is $T$–$p$-2. So the difference between treating a variable as predetermined or endogenous is that in the latter case there are one fewer lagged values available as instruments.

To see explicitly what Arellano-Bond estimation involves, let us focus on the education variable $E_t$ as a determinant of HIV infections.[11] The dependent variable is lagged one period so $p = 1$ and $T = 8$. If education is regarded as exogenous then only lagged values of $H_t$ are used as instruments. There are a different set of instruments with each time period used. Since we must have $t - 2 > 0$, the first usable $t$ is $t = 3$. At $t = 3$, the one instrument is $H_1$. In year 4, there are two instruments $H_1$ and $H_2$, and so on until year 8 when there are six instruments $H_1$ to $H_6$. When education is considered to be predetermined, the instruments in year $t = 3$ would include $E_1$ as well as $H_1$. In year 8 there would be twelve instruments ($E_1$ to $E_6$ together with $H_1$ to $H_6$). Finally, with education treated as endogenous, there would be the same number of lagged $H_t$ variables as instruments, but there would be one fewer lagged education instruments. Thus, $E_1$ would only appear in year 4 and in year 8 the instrument set would be $E_1$ to $E_5$ and $H_1$ to $H_6$.

There are two main implications of using the Arellano-Bond method for our estimation that we wish to highlight. There is the whole issue of the possible endogenously of education in the context of HIV infections. Using alternative lagged instruments allows us to check for

[11] In the estimation $E_{t-1}$ gave a better fit, so it is the lagged version of the education variable that is relevant. But, for convenience of exposition, we explain the choice of instruments in terms of $E_t$.

simultaneous equation bias. Secondly, because it is a fixed effects estimator, the Arellano-Bond technique enables us to isolate education and income from cultural effects, such as the tribe and/or religion of a region. Cultural variables do not change much over time. They would virtually be the same in every year. So when we first difference these variables they drop out from the estimation equations. This is important because in our study previously in this journal we tried a large number of variables to explain HIV infection rates in 31 Sub-Saharan African countries.[12] Only three were statistically significant (other than the year of outbreak). They were: education, income and religion. Education and income are both measured and analyzed in depth in this study. There is an estimation equation for each of them and each appears in both equations. The third statistically significant variable was religion (the percentage of the population that was Moslem) and this had a negative impact on infections. The problem was that the three variables were so highly correlated that it was difficult to isolate their separate effects. In this study by first differencing, the religion variable drops out (seeing that religion does not change much in a region over time). So the effects of income and education have been more reliably estimated in this study.

Arellano-Bond provided two estimators. In the one-step procedure estimation assumes that the errors are homoscedastic, and this can be applied with and without robust standard errors. In the two-step procedure, which allows for heteroscedasticity, estimation cannot be applied with robust standard errors. The results of applying the Arellano-Bond technique to the two production functions for the Tanzania panel are presented in tables 1 and 2.

---

[12] See Brent (2006a).

*Table 1: Dynamic Estimates (GMM): Dependent Variable is Change in Infection Rates (t-statistics in parentheses; p-values in square brackets)*

| Independent Variables | 1-Step Exog. 1 | 1-Step Robust 2 | 1-Step Predet. 3 | 1-Step Robust 4 | 1-Step Endog. 5 | 1-Step Robust 6 | 2-Step Exog. 7 | 2-Step Predet. 8 | 2-Step Endog. 9 |
|---|---|---|---|---|---|---|---|---|---|
| Infection Rate: $\Delta H_{t-1}$ | −0.437 | −0.437 | −0.298 | −0.298 | −0.371 | −0.371 | −0.419 | −0.289 | −0.334 |
| | (4.99) | (4.91) | (3.65) | (2.44) | (4.45) | (3.44) | (19.8) | (47.5) | (24.8) |
| Education: $\Delta E_{t-1}$ | 0.072 | 0.072 | 0.138 | 0.138 | 0.249 | 0.249 | 0.072 | 0.182 | 0.116 |
| | (0.78) | (3.06) | (1.56) | (2.93) | (2.23) | (1.59) | (6.33) | (8.34) | (1.65) |
| Income: $\Delta Y_{t-1}$ | −0.148 | −0.148 | −0.092 | −0.092 | −0.137 | −0.137 | −0.153 | −0.085 | −0.128 |
| | (3.27) | (2.76) | (2.27) | (1.14) | (3.12) | (2.01) | (8.27) | (14.9) | (21.0) |
| Year Dummy: $\Delta Z_{01}$ | −3.405 | −3.405 | −2.985 | −2.985 | −3.289 | −3.289 | −2.782 | −2.647 | −2.162 |
| | (2.68) | (2.95) | (2.27) | (2.94) | (2.49) | (2.86) | (5.12) | (10.2) | (5.25) |
| Constant term | 4.968 | 4.968 | 3.761 | 3.761 | 4.806 | 4.806 | 4.831 | 3.505 | 4.123 |
| | (5.07) | (3.85) | (4.23) | (2.20) | (4.92) | (3.19) | (9.02) | (19.0) | (12.2) |
| Sargan Test (Chi-square) | 39.67 | | 81.15 | | 62.55 | | 15.13 | 13.75 | 9.47 |
| | [0.013] | | [0.036] | | [0.077] | | [0.769] | [1.000] | [1.000] |
| Wald Test (Chi-square) | 33.81 | 45.56 | 20.14 | 21.84 | 29.46 | 16.23 | 8188.7 | 57898 | 8100 |
| Arellano-Bond Test | –0.64 | –1.58 | 0.24 | 0.41 | –0.23 | –0.53 | –1.32 | 0.20 | –0.02 |
| Autocorrelation Order 2 | [0.521] | [0.114] | [0.809] | [0.685] | [0.817] | [0.598] | [0.186] | [0.845] | [0.983] |

The results in columns 1-6 of the tables are for the one-step procedure and those in 7-9 are for the two-step method. Column 1 has the coefficient estimates assuming that the right-hand side variables are exogenous. Column 2 makes the same assumption, but uses robust standard errors, and column 7 is the two-step counterpart. Column 3, 4 and 8 assume the right-hand side variables are predetermined, column 4 using the robust standard errors and column 8 applying the two-step procedure. Finally, columns 5, 6 and 9 treat the right-hand side variables as endogenous, with column 6 having the robust standard errors and column 9 containing the two-step estimates.

*Table 2: Dynamic Estimates (GMM): Dependent Variable is Change in Income (t-statistics in parentheses; p-values in square brackets)*

| Independent Variables | | 1-Step Exog. 1 | 1-Step Robust 2 | 1-Step Predet. 3 | 1-Step Robust 4 | 1-Step Endog. 5 | 1-Step Robust 6 | 2-Step Exog. 7 | 2-Step Predet. 8 | 2-Step Endog. 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Income: | $\Delta Y_{t-1}$ | 0.922 | 0.922 | 0.921 | 0.921 | 0.925 | 0.925 | 0.921 | 0.921 | 0.925 |
| | | (23.9) | (23.6) | (25.9) | (24.5) | (25.7) | (26.5) | (864) | (704) | (618) |
| Education: | $\Delta E_{t-1}$ | 153.5 | 153.5 | 208.3 | 208.3 | 338.2 | 338.2 | 145.6 | 198.5 | 336.9 |
| | | (0.96) | (2.03) | (1.35) | (1.55) | (1.51) | (1.20) | (5.37) | (4.51) | (6.80) |
| Year Dummy: | $\Delta Z_{99}$ | 3530 | 3530 | 3519 | 3519 | 3763 | 3763 | 3501 | 3487 | 3758 |
| | | (2.17) | (2.39) | (2.17) | (2.57) | (2.26) | (2.76) | (46.8) | (32.8) | (34.1) |
| Constant term | | 882 | 882 | 900 | 900 | 873 | 873 | 880 | 900 | 873 |
| | | (1.00) | (0.85) | (1.11) | (1.00) | (1.06) | (0.99) | (30.9) | (32.0) | (31.7) |
| Sargan Test (Chi-square) | | 53.71 | | 90.61 | | 80.33 | | 210000 | 19.81 | 19.77 |
| | | [0.000] | | [0.000] | | [0.000] | | [0.000] | [0.997] | [0.975] |
| Wald Test (Chi-square) | | 589.2 | 598.2 | 687.2 | 615.3 | 675.6 | 716.9 | 769613 | 636827 | 648189 |
| Arellano-Bond Test | | –3.75 | –1.32 | –3.64 | –1.31 | –3.40 | –1.28 | –1.35 | –1.32 | –1.30 |
| Autocorrelation Order 2 | | [0.000] | [0.186] | [0.000] | [0.190] | [0.001] | [0.199] | [0.178] | [0.185] | [0.195] |

Before discussing the implications of the results for effectiveness, we need to identify which of them will be regarded as the best set of estimates. We will use the summary statistics at the bottom of the tables to guide us. In order for the estimates in the Arellano-Bond technique to be consistent, there must be no second-order autocorrelation. All the equations for the infection rate in table 1 cannot reject the null of no second order autocorrelation. So this is not an issue there. In the equations for the income production function shown in table 2, autocorrelation is a problem only for the one-step results, and only if we do not use the robust standard-errors. Since the coefficient estimates are the same with or without the robust standard errors, we can obtain estimates for all the income versions where the estimates are consistent.

Arellano and Bond recommend that only the one-step estimates be relied on for statistical inference purposes (as the two-step standard errors tend to be biased downwards in small samples). The main purpose of listing the two-step results is in order to apply the Sargan test of over-identifying restrictions (i.e., to see whether the right-side variables are exogenous or not).

The one-step Sargan test over-rejects in the case of heteroscedasticity. On the basis of the Sargan test using the two-step procedure, we cannot reject the null that the variables are exogenous in the infection equation. So columns (1) and (2) have the most reliable estimates in table 1. But, we can reject exogeneity in the income equation. This would suggest that the estimates in columns (3) – (6) would be better in table 2. Unfortunately, the estimates for the education variable in the income equation are not statistically significant at the 5% level in these columns (unlike all the other variables in either table). So we will place greater reliance on the results in columns (1) and (2) in table 2 as well as in table 1. Nonetheless, we will report all the results whether they be exogenous, predetermined, or endogenous. One can then see what difference it makes to the cost-benefit outcomes in the next section.

The estimates in tables 1 and 2 are reassembled in table 3 so that the results in terms of direct, indirect and total effects can be clearly seen. For simplicity we summarize results using the same estimation approach in both equations. Thus, we match columns in the two tables. So for example, we take the exogenous case of table 1 (column 1) and pair that with the exogenous case of table 2 (column 1). There are only three distinct sets of estimates in columns (1) to (6) of tables 2 and 3. So there are three results recorded in table 3 from these six equations. Column (a) of table 3 comes from columns (1) or (2) of tables 2 and 3, and this is regarded as the best set of estimates. Similarly, column (b) comes from (3) or (4), and column (c) from (5) or (6). Results (d), (e) and (f) correspond to the estimates for columns (7), (8) and (9) respectively. We see in all cases, that the direct effects are positive, the indirect effects are negative, and the indirect effects are stronger than the direct effects, which make the total effects negative. The net effect is therefore that increasing enrollment rates lowers HIV-AIDS infection rates.

*Table 3: Direct and Indirect Effects of Changes in Enrollments*

| Regression Coefficients | (a) | (b) | (c) | (d) | (e) | (f) |
|---|---|---|---|---|---|---|
| $\varphi_E$ | 0.072 | 0.138 | 0.249 | 0.072 | 0.182 | 0.334 |
| $\varphi_Y$ | –0.000148 | –0.000092 | –0.000137 | –0.000153 | –0.000085 | –0.000116 |
| $\gamma_E$ | 153.5 | 208.3 | 338.2 | 145.6 | 198.5 | 336.9 |
| $\delta_H$ | 1.437 | 1.298 | 1.371 | 1.419 | 1.289 | 1.334 |
| $\delta_Y$ | 0.078 | 0.079 | 0.075 | 0.079 | 0.079 | 0.075 |
| Direct effect | 0.050 | 0.106 | 0.182 | 0.051 | 0.141 | 0.250 |
| Indirect effect | –0.203 | –0.187 | –0.451 | –0.199 | –0.166 | –0.391 |
| Total effect | –0.153 | –0.081 | –0.269 | –0.148 | –0.024 | –0.140 |
| HIV Cases Averted | 1,408 | 743 | 2,481 | 1,365 | 226 | 1,294 |

Both the enrollment rate and the infection rate were measured as percentages in the estimation. So the regression results for the total, net effect show the effect of a 1% point rise in the enrolment rate on the percentage number of HIV-AIDS cases. To carry out a CBA it is necessary to covert these percentage effect changes into absolute numbers. In our sample, the average number of female enrollments in primary school was 2,050,672. The change in enrollments that defines the education expansion being evaluated is therefore 1% of this, i.e., 20,507. The number of HIV cases that were averted by this 20,507 expansion depends on the regression equation, which estimates the total effect in percentage terms, and knowing how many HIV cases a 1% drop amounts to. The average number of female HIV-AIDS cases for the period 1994 to 2001 was 922,537. 1% of this total is 9,225. So multiplying the total effects in each equation by 9,255 produces the estimates of the number of HIV cases averted, and these results appears as the last row in table 3. We see that the best estimate of the number of cases averted was 1,408 with a range of 226 to 2,481.[13]

[13] If one matches the exogenous case for the infection equation (column 1 or 2) with the predetermined case for the income equation (column 3 or 4) then the number of averted cases rises to 2,297. It would be even higher at 4,231 if one matches the exogenous case of the infection equation with the endogenous case for the income equation. However, we again point out the fact that the estimates for the education variable in the predetermined and

## 3. The Cost-Benefit Analysis

We have just seen that increasing Tanzanian primary school enrollments lowered the number of those infected with HIV-AIDS. The CBA is primarily involved with finding out whether the income gains from lowering the infections can outweigh the education costs. The plan is to estimate the benefits per person not infected $B_t$ and the costs per person educated $C_t$ and then multiply these averages by the number of beneficiaries $N_B$ and the number educated $N_C$ to obtain the total benefits and costs. The per person benefits and costs in any year will have to be discounted by the social discount rate $r$, so the cost-benefit criterion will be the net present value given by:

$$\text{Net Benefits} = \sum_{t=0}^{t=T} \frac{N_B B_t - N_C C_t}{(1+r)^t}$$

Our assumptions concerning the timing of the benefits and costs are summarized in figure 1 below. This summary is for the benchmark case (profile 1) that assumes that the HIV-AIDS cases that are averted are for the students themselves. Since the students are younger than the average female infected in Tanzania, we regard infection as occurring as soon as schooling has been completed. Benefits do not start accruing until ten years later as this is the median interval between infection and AIDS/death in Sub-Saharan Africa.[14] In order to ensure that we do not in any way overstate the benefits, we assume that for these ten years those infected by HIV are symptom free and they can work the same as those unaffected. We will also use a second time-line (profile 2) that differs from the benchmark by assuming that it is an average-aged female,

endogenous cases in the income equation are not statistically significant using the one-step procedure, so we do not rely on these high numbers.

perhaps someone other than the person who is being educated, whose infection is being averted. The main difference this makes to the calculations is that we assign infection as taking place five years later. The symptom free period is increased from ten to fifteen years. Benefits start at age 30 rather than 25, which mean that there are five fewer years of earnings. Since the women involved are now in the general population we use average female earnings for profile 2 rather than just average female primary school earnings as in profile 1. Benefits in both profiles end at death in year 50.[15]

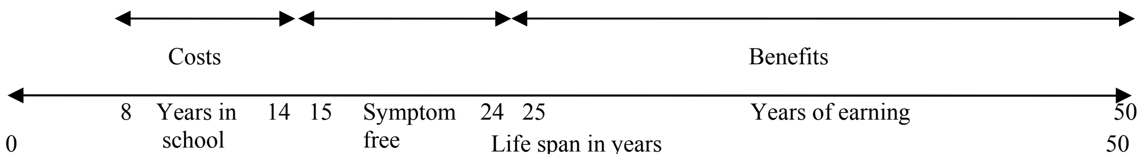


***Figure 1: Time-line for benefits and costs under profile 1***

The estimates of effectiveness were obtained in the previous section using a long-run, equilibrium framework. From this perspective it is logical to regard the education costs for generating the HIV-AIDS cases averted as coming from the entire primary schooling period and not just for one year. In Tanzania, primary education lasts for 7 years. We assign year 8 to be the first year of schooling and this becomes $t = 0$.[16] All years subsequent to year 8 will have values discounted. The standard assumption in the health care literature is to use a 3% discount rate.[17] We also use a 5% rate for discounting in an alternate set of results and these will act as the main sensitivity analysis. The seven years of schooling for the students end at the age of 14. The Ministry of Education and Culture's (2002) report on basic statistics in education for 2001 (table

---

[14] See Morgan et al. (2002).

[15] 50 years was the life expectancy at birth for 1988 stated in table 13 of the poverty report by the Research and Analysis Working Group (2002).

[16] We start at 8 and end at 14, rather than at 7 and 13 (which is the legal primary school ages), because in practice primary school covers the age range 6 to 17 and we want to better represent this range.

[17] See Brent (2003), ch.7 on discount rate theory and practice in health care evaluations .

8.3, p70) gives the primary school expenditures for the seven years 1994/95 to 2000/2001. We use these amounts and divide them by enrollments to give the government's cost per person educated for each of the seven years. To obtain the total cost we added a small amount for private costs.[18] The present value for all seven years of costs was 0.128 million TSHS (Tanzanian shillings) per person using the 3% discount rate and was 0.120 million TSHS with the 5% rate. This produced a per-person cost estimate of 0.162 million TSHS with the 3% rate and 0.153 million TSHS using the 5% rate. Multiplying these per person costs by the 20,507 enrollments produces the total cost figures that appear in table 4.

The benefits per person, based on the human capital approach, are a lower bound estimate determined as the present value of lifetime earnings. Annual earnings are assumed to rise at a 3.1% rate over time to reflect productivity gains.[19] For 25 years of earnings, the present value using the 3% rate was 5.344 million TSHS, and it was 4.291 million TSHS for the shorter 20 year benefit profile.[20] The corresponding amounts with the 5% discount rate are 3.059 and 2.333 million TSHS respectively. Because the number of infections averted varies by the regression equation from which they were estimated, the total benefits are different in table 4 in each of the columns (unlike the costs).

**Table 4: Cost-Benefit Outcomes (m: millions of TSHS)**

[18] The Household Budget Survey 2000/01 (2002), table 6.3, gave the mean monthly education expenditure per capita as 227 TSHS in 2000/01 and 25 TSHS in 1991/2. We averaged these figures and multiplied by 12.

[19] The real growth rate in Tanzania over the period 1992-2000 was 3.1%, see table 15 of the Research and Analysis Working Group (2002).

[20] The Household Budget Survey 2000/01 (2002), table B9.2, gives the monthly income for women with primary education as 19,990 TSHS, or 239,880 TSHS per annum. We adjust this amount upwards for each year after 2001 by the growth rate of 3.1% and then apply discount rates of 3% and 5% to obtain the net present value figures for earnings given in the text for profile 1. In the same table, monthly income for all women (with and without education) was slightly lower at 19,798 TSHS per month (237,880 TSHS per annum) and this is the earnings base for profile 2.

| | (a) | ( b) | (c) | (d) | (e) | (f) |
|---|---|---|---|---|---|---|
| **Benefits** | | | | | | |
| Cases averted | 1,408 | 743 | 2,481 | 1,365 | 226 | 1,294 |
| Benefits Profile 1 | | | | | | |
| Cases @ 5.344m (3% discount rate) | 7,522m | 3,972m | 13,260m | 7,295m | 1,208m | 6,913m |
| Cases @ 3.059m (5% discount rate) | 4,305m | 2,273m | 7,590m | 4,175m | 691m | 3,957m |
| Benefits Profile 2 | | | | | | |
| Cases @ 4.291m (3% discount rate) | 6,039m | 3,189m | 10,647m | 4,427m | 970m | 5,551m |
| Cases @ 2.333m (5% discount rate) | 3,284m | 1,734m | 5,789m | 3,185m | 527m | 3,018m |
| **Costs** | | | | | | |
| Enrollments | 20,507 | 20,507 | 20,507 | 20,507 | 20,507 | 20,507 |
| Enrollments @ 0.128m (3% discount rate) | 2,620m | 2,620m | 2,620m | 2,620m | 2,620m | 2,620m |
| Enrollments @ 0.120m (5% discount rate) | 2,455m | 2,455m | 2,455m | 2,455m | 2,455m | 2,455m |
| **Net Benefits** | | | | | | |
| Benefits Profile 1 | | | | | | |
| Cases @ 5.344 m – Enrollments @ 0.128m | 4,902m | 1,353m | 10,641m | 4,675m | –1,412m | 4,294m |
| Cases @ 3.059 m – Enrollments @ 0.120m | 1,850m | – 182m | 5,135m | 1,720m | –1,764m | 1,502m |
| Benefits Profile 2 | | | | | | |
| Cases @ 4.291 m – Enrollments @ 0.128m | 3,420m | 569m | 8,027m | 3,238m | –1,650m | 2,931m |
| Cases @ 2.333 m – Enrollments @ 0.120m | 829m | – 721m | 3,334m | 730m | –1,928m | 563m |
| **Benefit / Cost ratio** | | | | | | |
| Benefits Profile 1 | | | | | | |
| Cases @ 5.344 m / Enrollments @ 0.128m | 2.9 | 1.5 | 5.1 | 2.8 | 0.5 | 2.6 |
| Cases @ 3.059 m / Enrollments @ 0.120m | 1.8 | 0.9 | 3.1 | 1.7 | 0.3 | 1.6 |
| Benefits Profile 2 | | | | | | |
| Cases @ 4.291 m / Enrollments @ 0.128m | 2.3 | 1.2 | 4.1 | 2.2 | 0.4 | 2.1 |
| Cases @ 2.333 m / Enrollments @ 0.120m | 1.3 | 0.7 | 2.4 | 1.3 | 0.2 | 1.2 |

On the basis of the calculated benefits and costs the net benefits are displayed in table 4. In addition, we present the results as benefit-cost ratios as these facilitate comparisons across programs (where the scale of operations may be different) and across countries (where the local currencies are different).[21] We see from our best estimates in column (a) that, no matter the benefits profile or the discount rate, education enrollments always have positive net-benefits. Benefit-cost ratios vary between 1.3 and 2.9. Positive net-benefits throughout are also seen in columns (c), (d) and (f). Column (e) has the lowest number of cases averted and net-benefits are negative throughout. Lastly, columns (b), which has half the number of cases averted as in the best estimates, the sign attached to the net benefits depends on the discount rate, being positive with the lower rate and negative with the higher rate. Although benefit profile 2 greatly reduced

[21] For a comparison among CBA criteria, which includes a discussion of the NPV versus the benefit-cost ratio, see Brent (1998), ch.2. As explained in the introduction, we do not assume that the funds for HIV-AIDS are fixed - an assumption which would make the benefit-cost ratio the appropriate criterion. We include both criteria outcomes for the reasons explained in the text (i.e., to facilitate comparisons with other evaluations).

the net-benefits, at no time did it make the sign of the net-benefits different from that obtained under benefit profile 1.

The importance of the size of effectiveness is obvious. The significance of the choice of rate for discounting for our education CBA mirrors that often found when evaluating HIV-AIDS interventions geared at prevention. The long duration between being infected with HIV and dying with AIDS associated diseases means that benefits from interventions come well into the future where discounting has its potentially largest impact.

## 4. Summary and Conclusions

We undertook a cost-benefit analysis of female primary education in Tanzania over the period 1994-2001 and found that, on the basis of our best estimates, the net-benefits were strongly positive. Our cost-benefit analysis was carried out in two stages. First we had to establish whether increasing enrollments was effective in lowering HIV-AIDS infection rates. Then we tried to see whether the number of HIV-AIDS cases avoided was worth more than the costs incurred in financing the added enrollments. Both stages involved estimating a number of crucial parameters and relationships and we identify and discuss them here.

We were able to establish the effectiveness of primary school enrollments because we found that the indirect effect that lowered infection rates was able to offset the direct effect that raised rates. There were two forces that had to combine to obtain the negative indirect effect. First enrollments had to raise incomes. The human capital approach to education has been well established in both developing and developed countries, so this relationship should be expected. But, the second force, whereby raising incomes would lower infection rates could not be taken for granted in developing countries. There is a lot of evidence pointing to the fact that it is the richer countries in sub-Saharan African that have the higher HIV-AIDS rates, not the poorer ones. This was also the case in our data set that related to richer and poorer regions of Tanzania. However, these findings relate to a static framework where one is comparing levels of income and levels of infection. Once a dynamic estimation approach was adopted, we were able to find that it was *increases* in levels, and not the levels themselves, that lowered infection rates. So the second force behind the indirect effect was in evidence in the Tanzanian data.

Given that the indirect effect was negative, what explains its relative magnitude, such that it could outweigh the direct positive effect of enrollments? We give a statistical and an economic explanation. The statistical answer seemed to lie in the different dynamics of regional infection rates and income determination. The key regression coefficients in the two equations, with HIV-AIDS rates and per-capita income levels as dependent variables, had to be divided by their partial adjustment coefficients $1-\delta$ in order to derive their long-run impacts. In turned out that in the income equation $\delta_Y$ was around 0.922 in all the equations, which made $1-\delta_Y$ only 0.078. Thus, dividing the education coefficient in the income equation by this small amount greatly enhanced its long-run value and this helped produce a large indirect effect. While in the other equation, infection rates over-reacted to lagged infection rates, actually making $\delta_H$ negative in all the equations. So $1-\delta_H$ exceeded unity. Its reciprocal was less than one, meaning that the education coefficient in the infection rate equation was being reduced in forming its long-run impact and it is this that resulted in the lower direct effect.

Note that our use of the partial adjustment mechanism means that our estimates of effectiveness can be viewed as steady-state results. It is not the case that we are suggesting that, on an annual basis, one year of current education leads to large current infection decreases and large current income increases (i.e., we are not claiming that effectiveness is immediate). Rather the point is that we have established how much infections will fall and income will rise with a year's primary education in the steady state. Then we have converted this to a planning period where 7 years of primary school expenses are being compared with a future 20 year stream of future income (which is being discounted).

This statistical explanation is more in terms of why we *observed* such a large relative indirect effect rather than what *caused* it to be so large. On this issue it seems that $\gamma_E$, the size of the human capital effect, was crucial. It is important to recognize that our estimates are derived from a *regional* unit of analysis. The regional income gains proceeded from the increased regional female primary school enrollments. These income gains include the external effects and not just the primary effects of education on the females' incomes accruing to themselves. These external effects of female education can raise income through many different channels. From the work of Basu and Foster (1998) there is the idea that education by one person can raise the effective education level of illiterate household members by making them "proximate-literate". The authors cite evidence that this effect is likely to be larger for female household members that are educated than for males. Not only is there this education-to-education externality, Gibson (2001) has shown that there is a health externality for family members when one person in a household is educated (i.e., child height increases). Recently, Alderman, Hentschel and Sabates (2003) have found that the scope of influence is even wider. Female education raises the height-for-age levels of neighborhood children in addition to the impact on the household's children. These external effects are in terms of education and nutrition, but there is no reason to doubt that they would translate (via productivity effects) into larger long-run incomes for all. Even though the reality may be that, at this time in Sub-Saharan Africa, education may not raise the incomes of the females themselves by much, it may well be the case, as we have found, that the external effects on others is large, justifying greater government investment in female education.

The cost-benefit analysis basically came down to the issue as to whether it was worthwhile incurring the expenses involved with educating a cohort of 20,507 females in primary schools in order, in our best estimates, to avert 1,408 HIV-AIDS cases. Obviously if effectiveness had been lower, the net-benefit would have been less, and if the number of cases were larger, then the investment would have been even more beneficial. So program effectiveness is a necessary condition for any intervention to be worthwhile. Given the level of program effectiveness in Tanzania, the CBA then required valuing the reduced infections and costing the enrollments. We used the net-present value of the future earnings of females to price the health benefits. The use of the human capital approach in health care evaluations is always controversial.[22] The Report of the Commission of Macroeconomics and Health (2001, p32) cites a number of studies where a valuation three times the per capita income has been used rather than a conservative estimate that relies on just the per capita income itself. Since we are focusing on females lives saved, the use of the human capital approach is even more controversial in a developing country context where gender discrimination takes place. In Tanzania, males with primary education earned 1.8 times their female counterparts.

[22] For a comprehensive analysis of the strengths and weaknesses of the human capital approach relative to willingness to pay (WTP) in health care evaluations, see Brent (2003). While we basically endorse WTP as the superior methodology from the point of view of Welfare Economics, we need to point out that this approach assumes consumer sovereignty and that there are grave misgivings in applying this for expenditure decisions related to 7 – 14 year olds as in our study here. The human capital approach has the practical advantage of ready data availability. There are very few WTP CBAs for Africa. Examples are Forsythe et al.'s (2002) evaluation of voluntary HIV counseling and testing in Kenya, which used Contingency Valuation (CV) techniques, and Brent's (2006b) CBA of the condom Social Marketing Program in Tanzania, which relied on the revealed preference approach via estimated market demand curves. Primary schooling is now free in Tanzania (tuition before it was abolished in 2000 was about US $2 per year). So there is no scope to estimate market demand curves for primary education in Tanzania. There are no published CV studies of the WTP for education in Tanzania. CBA as applied

Nonetheless, even with a conservative methodology, we were still able to show that female primary education had large positive net-benefits in Tanzania. This is because of the other side of the evaluation equation for any low income country, i.e., costs and not only benefits are low when measured in monetary terms. Even though we used a valuation of around US $7,500 per case averted in our best estimate using profile 1, when upwards of US $5 million might be the number used in an US evaluation based on the statistical life approach, one needs to be aware that the cost of 7 years of primary school education in Tanzania was only about $213.[23] The cost in the US in many school districts would be hundreds of times larger. It is the difference between these scaled-down benefits and scaled-down costs that determines evaluation outcomes for developing countries. The desirability of a public project is not necessarily predetermined by the benefit methodology one adopts for making the evaluation. Problems arise when costs involve resources from high-income donor countries and are thus expressed in large monetary values. Very few interventions will generate comparable benefit sums no matter the efficiency-based valuation method one uses for the benefits.[24]

---

Welfare Economics is covered in Brent (1996), which includes a number of CV applications (none related to Africa).

[23] We are using an exchange rate of 600 TSHS = 1 US$ for the monetary comparisons in this paragraph, the rate used by Sweat et al. (2000) in his evaluation of Voluntary Counseling and Testing in Tanzania.

[24] The CBA of the condom socially marketing program in Tanzania by Brent (2006b) is an exception to this rule as it found that the program broke-even at existing subsidy levels, and it would have a benefit-cost ratio in the range 1.7-2.3 if the subsidy were reduced to its most efficient size. In this evaluation the condom costs were in US $ and the benefits were given by the area under the domestic demand curve for condoms.

## References


Alderman, H., Hentschel, J and Sabates, R. (2003): "With the Help of One's Neighbors: Externalities in the Production of Nutrition in Peru." *Social Science and Medicine*, 56: 2019-31.

Arellano, M. and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an application to Employment Equations", *The Review of Economic Studies*, 58: 277-297.

Basu, K. and Foster, J.E. (1998): "On Measuring Literacy." *The Economic Journal*, 108: 1733-1749.

Brent, R.J. (2006a): "Does Female Education Prevent the Spread of HIV-AIDS in Sub-Saharan Africa?" *Applied Economics*, 38: 491-503.

Brent, R.J. (2006b): "A Cost-Benefit Analysis of a Condom Social Marketing Program in Tanzania." *Department of Economics*, Fordham University.

Brent, R.J. (2003): *Cost-Benefit Analysis and Health Care Evaluations*, Massachusetts: Edward Elgar Publishing Company.

Brent, R.J. (1998): *Cost-Benefit Analysis for Developing Countries*, Cheltenham: Edward Elgar Publishing Company.

Brent, R.J. (1996): *Applied Cost-Benefit Analysis*, Cheltenham: Edward Elgar Publishing Company.

Commission on Macroeconomics and Health (2001): *Investing in Health for Economic Development*, Geneva: World Health Organization.

Creese, A., Floyd, K., Alban, A. and Guinness L. (2002): Cost-effectiveness of HIV/AIDS Interventions in Africa: A Systematic Review of the Evidence." *The Lancet*, 359: 1635-42.

De Walque (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, D.C.

Fawzi, W.W., Msamanga, G.I., Spiegelman, D., Wei, R., Kapiga, S., Vilamor, E., Mwakagile, D., Mugusi, F., Hertzmark, E., Essex, M. and Hunter, D.J. (2004): "A Randomized Trial of Multivitamin Supplements and HIV Disease Progression and Mortality." *The New England Journal of Medicine*, 351: 23-32.

Forsythe S., Arthur G., Gilbert N, Mutemi R., Odhiambo J. and Gilks C. (2002): "Assessing the Cost and Willingness to Pay for Voluntary HIV Counselling and Testing in Kenya." *Health Policy and Planning*, 17: 187-195.

Gibson, J. (2001): "Literacy and Intrahousehold Externalities." *World Development*, 29: 155-166.

Household Budget Survey 2000/01 (2002): *Final Report*. Dar es Salaam: National Bureau of Statistics Tanzania.

Hargreaves, J.R. and Glynn J.R. (2002): "Educational Attainment and HIV-1 Infection in Developing Countries: A Systematic Review." *Tropical Medicine and International Health*, 7: 489-498.

Logie D. (1999): "AIDS Cuts Life Expectancy by a Quarter", *British Medical Journal*, 319: 806.

Ministry of Education and Culture (2002): *Basic Statistics in Education 2001, Regional Data.* Dar es Salaam: United Republic of Tanzania.

Ministry of Health (2002): *HIV/AIDS/STI Surveillance Report Number 16, January – December 2002*. Dar es Salaam: National AIDS Control Program.

Morgan, D., Mahe, C., Mayanja, W., Okongo, M.J., Lugeba, R. and Whitworth, J.A.G. (2002); "HIV-1 Infection in Rural Africa: Is there a Difference in Median Time to AIDS and Survival Compared with that in Industrialized Countries?" AIDS 6: 597-603.

National Bureau of Statistics (2001): *National Accounts of Tanzania* 1989-2001, National Accounts Department, The President's Office, Planning and Privatization. Dar es Salaam: United Republic of Tanzania.

Research and Analysis Working Group (2002): *Poverty and Human Development Report 2002*, Dar es Salaam: United Republic of Tanzania.

Sweat M., Gregorich S., Sangiwa G., Furlonge C., Balmer D., Kamenga C., Grinstead O. and Coates T. (2000): "Cost-Effectiveness of Voluntary HIV-1 Counselling and Testing in Reducing Sexual Transmission of HIV-1 in Kenya and Tanzania", *Lancet*, 356, 113-121.

Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS), and ORC Macro (2005): *Tanzania HIV/AIDS Indicator Survey* 2003-2004. Calverton, Maryland, USA

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

World Bank (2002): "*Education and AIDS*", World Bank, Washington, D.C.

## Appendix. Summary of the data

There were three variables that feature in the dynamic statistical analysis for which we had panel data for 20 regions in Mainland Tanzania over the period 1994-2001, i.e., the infection rate, the enrollment rate and income levels. Here we present the detailed specifications for the variables, the data sources and statistical summary.

*Infection rate*: This is the percentage prevalence HIV-AIDS rate among female blood donors aged 15 and over. Data for the period 1991-2001 were consolidated in the Surveillance Report Number 16, table 8, issued by the Ministry of Health for the year January to December 2001 as reported to the National AIDS Control Programme (NACP) – see Ministry of Health (2002).

*Education*: We used the gross and net primary school enrollment ratios for girls by region as they appear in the annual Basic Statistics in Education regional data reports issued by the United Republic of Tanzania - see Ministry of Education and Culture (various years). The annual reports for the years 1997-2001 gave the actual enrollment ratios directly. The annual reports for 1994-1996 gave only the number of female enrollments. We calculated the ratios from our estimates of the relevant female populations, which we obtained by applying the total population growth increases for each region for the years 1996-1998 to the known 1998 school age population (i.e., extrapolating backwards). The official primary school ages were from 7-13 years and these were the populations for the net enrollment ratios; while all females ages 5-17 are the populations for the gross enrollment ratios.

*Income*: The National Bureau of Statistics (2001) gave regional income figures for the years 1989-2001. So this covered all the years 1994-2001 in our analysis. Table 19 listed the data as Regional Per Capita GDP at Current Prices, which makes the data nominal GDP.

The Arellano-Bond technique utilizes lagged values as instruments for variable differences, so we summarize the data for the variables in current, difference and lagged forms. The best estimates were for gross enrollment ratios so we list only their data values in the table below.

| Variable | Number of Observations | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| **Current Values** | | | | | |
| Infection Rate | 142 | 11.7 | 8.0 | 0.1 | 55.0 |
| Education | 160 | 79.0 | 12.3 | 49.4 | 130.4 |
| Income | 160 | 151,243 | 80,248 | 50,898 | 554,287 |
| Year Dummy 99 | 160 | 0.125 | 0.332 | 0 | 1 |
| Year Dummy 01 | 160 | 0.125 | 0.332 | 0 | 1 |
| **Lagged Values** | | | | | |
| Infection Rate | 122 | 11.3 | 8.0 | 0.1 | 55.0 |
| Education | 140 | 78.4 | 12.6 | 49.4 | 130.4 |
| Income | 140 | 141,413 | 74,290 | 50,898 | 518,351 |
| **Differences** | | | | | |
| Infection Rate | 116 | 0.4 | 7.3 | – 40.1 | 22.9 |
| Education | 140 | 0.5 | 6.4 | – 42.2 | 38.9 |
| Income | 140 | 20,459 | 10,705 | – 7,556 | 64,048 |
| **Lagged Differences** | | | | | |
| Infection Rate | 96 | 0.5 | 7.7 | – 40.1 | 22.9 |
| Education | 120 | –0.4 | 6.1 | – 42.2 | 38.9 |
| Income | 120 | 20,907 | 9,834 | 6835 | 64,048 |