HIV Does Matter for Fertility:

Human Capital, Mortality and Family Size

Günther Fink

Harvard School of Public Health

Sebastian Linnemayr[1]

Harvard School of Public Health

This version: May 2009

Abstract

In this paper we provide new evidence regarding the long-term impact of HIV on fertility and economic development. We develop a theoretical framework where parents optimally allocate their resources between child-rearing and consumption, and incorporate both infant and adult mortality in their fertility decision. The model predicts an ambiguous overall effect of HIV on fertility, but suggests that the optimal fertility adjustment to HIV is larger for more educated parents than for parents with little or no formal education. We test this prediction using a novel data set combining historical individual level data from World Fertility Surveys (WFS) with recent data from the Demographic and Health Surveys (DHS) including nationally representative HIV-testing. The result that more educated women reduce fertility more than uneducated mothers in the presence of HIV appears to hold both in the longitudinal and the cross-sectional analysis. Our results imply that HIV is unlikely to have a significant effect on population size, but will negatively affect countries' long term economic prospects through an adverse shift in the population's human capital composition.

[1] Corresponding author: Center for Population and Development Studies, 9 Bow Street, Cambridge, MA 02138. Email: slinnema@hsph.harvard.edu.

The authors are grateful to David Canning, Jocelyn Finlay, Michael Moore, Younghwan Song and the participants of the Population Economics Program (PEP) Lunch Seminar Series, the participants of Session 3 at the PAA 2009 in Detroit, and participants at the Economics Department Seminar at Union College for their comments and suggestions.

1. Introduction

With its first case diagnosed only 25 years ago[2], HIV/AIDS has grown to become one of the most salient issues in global health today. By 2007, 40 million people were estimated to be HIV-positive worldwide, and an estimated 12 million people had lost their lives to the disease (UNAIDS 2008). The epidemic scale of HIV reached within this short period of time has eroded the improvements in life expectancies experienced in the first half of the 20th century by many Sub-Saharan African countries, and substantially altered actual and perceived mortality risk in the most affected areas (Kohler, Behrman et al. 2007).

From an economic perspective, the effects of HIV are not obvious; a series of theoretical and empirical studies has yielded mixed, and often ambiguous, results (Cuddington 1993; Bloom and Mahal 1997; Dixon, McDonald et al. 2001; International Labour Office 2004; Young 2005; Werker, Ahuja et al. 2006). As Young (2005) argues, increases in mortality may be tragic from a human perspective, but may at the same time be beneficial from a aggregate welfare perspective in settings where natural resources and capital are scarce. With limited capital, a smaller workforce can earn higher wages in equilibrium, and income per capita will rise. This conjecture, however, rests on the assumption that there are no behavioral responses in fertility in the face of increased mortality, so that HIV leads to a population decline. Given the close to perfect correlation between fertility and infant mortality observed across countries and time (Schultz 1997), this hypothesis appears questionable at least ex-ante.

Young (2007) and Kalemli-Ozcan (2006) use national level time series data on HIV prevalence in a first attempt to estimate the fertility response to HIV, coming to strikingly different conclusions. With rising concerns regarding the quality of international HIV time series data, more recent research has shifted to using the newly available individual level population based HIV data collected in the Demography and Health Surveys (DHS). Juhn et al. (2008) and Fortson (2009) use a number of DHS surveys to investigate whether HIV prevalence can predict

[2] HIV was first recognized and officially listed in the early 1980s CDC (1981). "Pneumocystis Pneumonia --- Los Angeles." CDC Report **30**(21): 1-3.

fertility changes between regions and across DHS waves, and unanimously conclude that the average fertility response is very small, and not significantly different from zero.

In this paper, we extend the existing literature on HIV and fertility in two ways. First, we develop a simple theoretical framework that highlights the key mechanisms driving the parental fertility response to HIV. In the model, rational and risk-neutral parents optimally allocate their resources between consumption and child care in the presence of mortality risk. Rather than assuming risk to be age-invariant, we make an explicit distinction between infant and early adulthood mortality risk in the model. This distinction is important in the context of HIV. As opposed to most other infectious diseases, the death burden of HIV is not concentrated among infants, but mostly affects young adults (Oster 2009). This is illustrated in Figure 1, which compares mortality and survival rates for 1990 and 2006 in Senegal, a country with relatively low HIV burden, with the corresponding rates in Zambia, a typical East African country with high HIV prevalence. While mortality rates have slightly fallen for all age groups in Senegal over this period, mortality rates in Zambia have been essentially constant for children under age 10, but have increased substantially for the age group of 15 to 40 year olds.

[Figure 1 here]

From a human capital investment perspective, the cost of losing an adult child is much higher than the cost of losing a child during the very early stages of life. While it may be optimal to respond to higher infant mortality with higher fertility from a parental perspective, the same is not necessarily true for adult mortality. In our model, the overall effect of HIV on fertility is ambiguous and depends on the relative magnitudes of infant and adult mortality. However, since the (relative) cost of child rearing increases with parents' labor market opportunities, the model predicts that more educated parents respond more strongly to the increased mortality risk generated by HIV, so that the interaction between HIV and parental labor market opportunities is negative.

To test the predictions of the model, we construct a novel data set combining all currently available Demographic and Health Surveys (DHS) with nationally representative HIV-testing with micro-level data collected in the World Fertility Surveys (WFS) during the pre-HIV era. This dataset allows us to explore cross-sectional variation in the DHS as well as long-term

changes in fertility induced by HIV. We divide our empirical analysis in three parts. In the first part, we take a long-term perspective, and focus on the 32 regions in five Sub-Saharan countries that allow a direct comparison between WFS and DHS. In the second part, we combine all 21 currently available DHS data sets with nation-wide HIV testing to contrast the time-series results of the first section with evidence from the most recent micro surveys. Last, we test the prediction of our model in a short regional panel for Mali, the only country where two independent DHS surveys with nationally representative HIV testing are available.

Overall, our empirical results appear highly consistent with the main predictions of our theoretical model. Similar to the results presented by Fortson (Fortson 2009), we find little correlation between long-term changes in HIV and fertility at the regional level. These results, however, mask important differences across educational groups. Our long term results indicate that populations with no education or merely primary education increase their fertility in response to HIV, but that the opposite is true for populations with secondary or higher education, who reduce their fertility in the presence of HIV. Very similar results emerge from the extended cross-sectional sample of DHS surveys with HIV measurement. More educated mothers appear to respond to the HIV epidemic by reducing fertility more than their less educated peers do. Exploring the larger set of available variables in the DHS sample, we also test an alternative specification, where we take the asset index rather than mother's education as our proxy for the labor market opportunities of the parents. Both specifications yield the predicted negative interaction between HIV and labor market opportunities. To test the robustness of our results, we split the pooled DHS sample by geographic location and HIV prevalence – the results change little. The results are also consistent with the difference-in-differences estimates from two independent surveys from Mali. Similar to our long-term results, mothers with some education reduce their fertility by more than mothers who have never gone to school in the presence of HIV.

The rest of the paper is organized as follows: we introduce the theoretical framework in Section 2 of the paper, and present the data and main analysis in Section 3. We discuss our main findings and conclude in section 4.

2. A model of fertility, education and mortality

2.1 Derivation of the model

Consider a single-period model where parents have homothetic preferences over consumption *c* and the number of children *n* such that their utility *u* is given by

$$u = \frac{c_t^{\alpha}}{\alpha} + \rho(T, \beta)\frac{n^{\alpha}}{\alpha}. \quad (1)$$

Assume that $0 < \alpha < 1$, so that the marginal utility of both consumption and children is positive and decreasing. The relative utility ρ parents derive from their children depends on an individual taste or altruism parameter $\beta > 0$, as well as the children's life expectancy. Specifically, we assume that parents are better off if their children live longer, i.e. $\frac{\partial \rho}{\partial T} > 0$. This assumption is primarily based on the notion that parents altruistically care about their children in the tradition of Becker (1965; 1981). It is, however, also consistent with non-altruistic parents who maximize the probability of survival of their own genetic lineage, and thus directly benefit from the longevity of their descendants (Bergstrom 1996). Similarly, if parents invest in children to have support during their own old age, longer life expectancies likely increase the insurance value of, and relative returns to, child investment (Kotlikoff and Spivak 1981).

Similar to Soares (2005), we make an explicit distinction between early childhood mortality and mortality during later stages of the offspring's life. As discussed in the introduction and illustrated in Figure 1, this distinction between child and adult mortality is crucial in the context of HIV AIDS. While HIV significantly increases mortality among *infected* children (Walker, Schwartländer et al. 2002), the average effects on infant mortality are small compared to the effects HIV has on adult mortality. Countries with high HIV prevalence appear to have witnessed smaller (or no) improvement in infant or child mortality than similar countries with low HIV prevalence (Adetunji 2000). The main burden of HIV mortality, however, is highly concentrated among young adults as shown in Figure 1.

For the purpose of our model, we assume that children die either right at the beginning of their life or at some point during adulthood. As customary in fertility models (Schultz 1997), we assume that childrearing has a time cost t_n, so that parents have to divide their time (of length 1) between working and caring for (or educating) their children, i.e.,

(2) $$1 \geq l + nt_n$$

Parents are heterogeneous in their labor marked productivity θ and earn a wage *w* which is strictly increasing in their productivity, i.e. $\frac{\partial w}{\partial \theta} > 0$. The consumption budget constraint is thus given by

(3) $$w(\theta)l \geq c$$

Given the non-satiation property of the utility function, equation (2) and (3) hold with equality at the optimum, so that we can restate them as

(4) $$w(1 - nt_n) - c = 0$$

We abstract from the emotional or financial burden associated with pregnancy and child birth, so that the cost of childrearing is zero for any child dying right at the beginning of its life in the model. Accordingly, the number of children *n* parents raise (and draw utility from) in the model is different from the number of births given. Since the cost of child replacement is zero by assumption, the total number of birth *b* chosen in equilibrium is inversely proportional to the probability of childhood survival and the optimal number of (adult) children, that is

(5) $$b = \frac{n}{(1 - p)}$$

where *p* is the probability of dying at the beginning of life, our proxy variable for infant mortality. Equation (5) implies that parents (on average) fully anticipate future mortality of their children and adjust their reproductive behavior accordingly. This is consistent with the close to perfect inverse correlation between infant mortality and fertility observed across countries and

time (Schultz 1997)[3]. Maximizing equation (1) with respect to constraint (4), and re-arranging the first order conditions, we get

$$c = n\left(\frac{wt_n}{\rho}\right)^{\frac{1}{1-\alpha}} \tag{6}$$

Plugging into (4) we can derive the optimal number of children n^* as

$$n^* = \frac{1}{t_n + \left(\frac{t_n}{\rho}\right)^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}} = \frac{1}{t_n\left[1+\rho^{\frac{1}{\alpha-1}} t_n^{\frac{\alpha}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}\right]} \tag{7}$$

Equation (7) posits that the optimal number of children decreases with the time requirement for each child, t_n, and with parental wage w, which implies that substitution effects strictly dominate income effects in the framework chosen. This feature of the model is important, as it generates the negative correlation between parental education and fertility empirically observed. Somewhat less obviously, equation (7) implies that the optimal number of children increases with the relative utility ρ parents get from their children. Deriving the optimal number of children with respect to ρ we get

$$\frac{\partial n^*}{\partial \rho} = \frac{1}{1-\alpha}\rho^{\frac{2-\alpha}{1-\alpha}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}\left[t_n + \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}\right]^{-2} > 0. \tag{8}$$

The main question we want to address in this paper is the effect of HIV on fertility. Fertility is defined as the number of live births given by a woman over her fertile years, and is thus directly determined by the number of birth b chosen by each woman in the model. The number of birth b chosen is a function of the optimal number of children n^* and infant mortality p. HIV affects the number of births chosen by a women both through a decrease in adult life expectancy T, which lowers ρ, and through an increase in infant mortality p. Given that we allow for differential wages and incomes in the model, one key questions is whether the absolute

[3] This does not imply that parents directly "replace" each child that dies ex post; it only means that parents plan to give a number of births such that the desired family size can be achieved given perceived mortality rates.

increase in mortality risk generated by HIV differs across socioeconomic groups. Richer parents can theoretically provide better access to health care for their children; on the other hand, one could also argue that wealthier children might be more mobile and have more exposure to HIV/AIDS. The empirical evidence on this issue is mixed. While Fortson (2008) find a strongly positive correlation between socioeconomic status and HIV infection risk, Baker et al. (2009) find a negative relation between education and HIV infection for the younger cohorts. Given this mixed prior, we assume uniform increases in mortality risks across socioeconomic groups.

Proposition 1:

i) *The net average effect of HIV on fertility as measured by the number of birth given by a woman is ambiguous, and depends on the relative magnitudes of the shocks to infant and adult mortality.*

ii) *The fertility response to HIV depends on the labor market opportunities of the parent. As long as parental wage is below some critical threshold* $\bar{w}$*, the fertility response to HIV increases with the labor market opportunities of the parent: The higher the labor market opportunities of the parent, the larger will be their fertility reduction in response to HIV.*

The first part of Proposition 1 follows directly from equations (5) and (8). HIV increases infant mortality, which by (5) leads to a mechanical adjustment in the total number of births given[4]. With costless replacement, this first effect is unambiguously positive. This aspect has been highlighted in most of the recent work on HIV and fertility (Kalemli-Ozcan 2006; Kalemli-Ozcan, Juhn et al. 2008; Fortson 2009) The child mortality effect is, however, counteracted by

[4] The partial derivate with respect to infant mortality is given by $\frac{n^*}{(1-p)^2} > 0$

the effect of adult mortality derived in (8). Since HIV increases the risk of losing children during their adulthood ($\Delta T < 0$), parents shift their optimal allocation towards consumption, and thus lower the number of births given.

The second part of Proposition 1 follows from equation (8). Taking the (cross-)derivative of the optimal fertility response to the discount factor ρ with respect to the wage rate we get an expression which is larger than zero as long as fertility is above a threshold given by[5]

$$n(\overset{)}{w}) = \frac{1}{t_n + \left(\frac{t_n}{\rho}\right)^{\frac{1}{1-\alpha}} (\frac{\rho^{\frac{1}{\alpha}}}{t_n})^{\frac{\alpha}{1-\alpha}}} = \frac{1}{2t_n}. \qquad (9)$$

The intuition for this result is that as long as parents' earnings potential is close to zero, the optimal number of children is large, and the elasticity with respect to changes in the relative utility of children ρ is small. As parents' earning potential rises, parents increasingly reduce child rearing in favor of labor market income; as a result, the optimal number of children decreases, and the elasticity with respect to ρ increases. The marginal effect of wage, however, is not constant, and reverses as optimal number of births goes towards zero.

Given the low income country sample we are analyzing, most variation comes from parents with high fertility and relatively low wage rates, so that we expect the fertility response to increase with wages. To illustrate this point, we calibrate our simple model to the data used in our empirical work. The DHS sample shows a large dispersion of both income (as proxied by education or assets) and fertility. On average, the total number of children is about 6 in the lowest decile and slightly below 3 children in the highest decile. We normalize the base wage to

[5] See Appendix for a full derivation of this expression.

1 and choose (ρ, α, t_n) such that the curvature of the optimal fertility curve (blue line) fits the observed patterns. In Figure 2, we plot both the optimal fertility rate and the fertility response to a marginal increase in the relative utility (discount factor) ρ as a function of income. As Figure 2 shows, the marginal fertility response to changes in ρ initially increases rapidly with the wage rate, and only slowly declines as fertility rates get very low.

[Figure 2 here]

2.2 Model Implications and Empirical Specification

Proposition 1 has two main implications: first, the individual response to HIV depends on the relative burden of infant and adult mortality. Increased adult mortality (of the children) leads to a reduction in the demand for children, while increases in infant mortality mechanically increase fertility as parents adjust the number of births to achieve the desired number of children. Empirically, this point is hard to estimate as both infant and adult mortality likely have increased in areas affected by HIV. Even if detailed statistics on child and adult mortality at the regional level were available, disentangling these two effects at the regional level would be difficult, leaving us with estimating the average fertility response to HIV with an ambiguous prior.

The second part of Proposition 2 can more readily be tested empirically: our model implies that the magnitude of the fertility response is a function of the wage rate w, which defines the opportunity cost of child-rearing. We choose two variables as proxies for parents' labor market opportunities in the absence of wage data in the household survey data used: first, we use mother's education in years as the most obvious proxy for labor market opportunities. As alternative, we employ a variable that may be more directly related with income status: the wealth index provided by DHS. Our main empirical specification can be stated as follows:

$$(10) \quad birth_{ijkt} = \alpha + \beta hiv_{ijkt} + \gamma hiv_{jkt} + \lambda wage_{ijkt} * hiv_{jkt} + \phi X_{ijkt} + post + \delta_j + \delta_j * post + \varepsilon_{ijkt}$$

where $birth_{ijkt}$ is the number of births by woman *i* in region *j* of country *k* in period *t*. Our main dependent variable is the number of births given by the respondent over the five year period preceding the interview[6]. hiv_{jkt} is the regional HIV prevalence rate[7], *wage* stands for our wage proxy (education and asset index), and *X* is a matrix of additional controls. δ_j are regional fixed effects and *post* is an indicator equal to 1 if the data come from the DHS surveys, and zero otherwise. To control for the highly heterogeneous political and economic experiences over the last decades, we also use a regional or country specific time trend $(\delta_j * post)$ in the long run specifications. Since HIV has a direct biological effect on fertility, we control for the woman's own HIV status hiv_{ijkt} in our baseline specification. We also test alternative specifications where we exclude HIV positive women from the analysis. Since the model implies a non-linear interaction between HIV prevalence ($\Delta\rho$) and earning probabilities, we also include interaction terms with squared wage in some of our specifications.

3. Data and Results

3.1 Long-term Comparison: WFS and DHS

To investigate the long term relation between HIV and fertility we combine data from those Demographic and Health Surveys (DHS) for which nationally representative HIV-data are available with data from the World Fertility Surveys (WFS). The World Fertility Surveys are the predecessor of the DHS surveys and were conducted in 41 countries between 1975 and 1982, i.e. in a period when HIV prevalence was virtually zero (UNAIDS 2004).

[6] As a robustness check, we also looked at the more short-term measure of whether the woman is currently pregnant and number of births over the last year. The results look very similar and can be obtained from the authors upon request.

[7] Juhn, Kalemli-Ozcan, and Turan (2007) use the cluster HIV rate for this purpose. We prefer the regional HIV rate as on average there are only about 10 women with HIV testing per cluster, resulting in very noisy measure of HIV prevalence.

Currently, 38 WFS surveys are publicly available, 8 of which can be matched to DHS data sets with HIV measurements: Cameroon, Cote d'Ivoire, Dominican Republic, Ghana, Haiti, Kenya, Lesotho and Senegal. Unfortunately, there is a complete mismatch between the sampling framework used in the WFS and DHS surveys for the Dominican Republic, Haiti, and Lesotho, which makes a dynamic regional analysis for these three countries impossible. We include only those areas which were targeted in both the DHS and WFS surveys, leaving us with 32 geographical sampling areas in the 5 remaining countries. Figure 3 shows the HIV prevalence for all regions in our long-term comparison sample, with separate displays for rural and urban populations in each region.

[Figure 3 here]

As Figure 3 shows, the regions with the highest HIV prevalence rates in our sample are located in Cameroon and Kenya, while Senegal and Ghana have mostly low prevalence rates (1-2% prevalence). All countries show significant regional and urban/rural variation, with urban prevalence rates 5-10 times the rural rates in the most hard-hit areas in each country. The region with the highest prevalence rate in our sample is Nyanza (Kenya), with HIV prevalence rates of 19.6% (urban) and 12.7% (rural). Table 1 shows descriptive statistics for DHS and WFS samples. While surveyed women look very similar in terms of age, Table 1 highlights the large increases in education and urbanization witnessed in Sub-Saharan Africa over the last 30 years.

[Table 1 here]

Empirical Results

We present the main results from our investigation of the long-term effects of HIV on fertility in Table 2. As outlined in the previous section, the dependent variable used is the number of births in the five years preceding the interview. In the first column, we replicate the analysis presented in Juhn et al. (2008) and Fortson (2009) by regressing the number of births on the regional HIV prevalence rate. Similar to those two studies, we find a negative effect of individual level HIV status, but no significant effect of regional HIV prevalence on fertility. The results on the other control variables are as expected: fertility increases with age, but at a

decreasing rate, married women have more children, and women living in urban areas have fewer children than people living in rural neighborhoods.

[Table 2 here]

In column 2 of Table 2, we directly test the specification outlined in equation (11), which augments the basic model estimated in the first column with an HIV-education interaction term. Once we allow for a differential fertility response by education status, regional HIV prevalence has a *positive* and statistically significant effect on average fertility. A point estimate of 1.1 implies that a 10 percentage point increase in HIV prevalence in a region leads to an increase in the number of children born in the last five years by about 0.1. Given a mean number of births close to 1, this corresponds to a 10% increase in fertility. The effect, however, varies greatly with mother's education. As predicted by the model, the interaction between fertility is negative, and highly significant. A point estimate of -.23 on the interaction term implies that HIV lowers fertility for all women with 5 or more years of education, which is the case for about half of the women in the DHS sample.

In column 3 of Table 2, we restrict our sample to HIV-negative women to investigate the purely anticipatory, behavioral effect. For this restricted sample, we lose about 800 observations, since about 5% of the women in the DHS sample have tested positive for HIV. The results remain virtually of the same magnitude and statistical significance. In column 4 of Table 2, we further restrict our analysis to women of ages 18-40. Women over 40 have significantly lower fertility rates and may also have formed their fertility desires before the arrival of the HIV epidemic. We also exclude the under 18 year olds, as the 5-year birth variable becomes hard to interpret for this subgroup of women. This sample restriction leads to a significant increase in the magnitude of the coefficient for regional HIV prevalence, almost doubling in size. For this sample, a 10% regional HIV prevalence translates into an increase in the number of children born in the last five years by .25, or roughly a 25% increase compared to the mean value found in the sample. The interaction term of mother's education with regional HIV prevalence increases similarly in magnitude. The point estimate for the level of education where the fertility response turns negative moves to about 7 years of schooling.

In column 5 of Table 2, we restrict the sample to DHS observations only. While we lose the possibility to control for regional fixed effects in this specification, the purely cross-sectional analysis allows us to identify the degree to which the long-term results are driven by matching the two data sources over time. Even though we have a significantly smaller data set, the results change only slightly: the coefficient on regional HIV prevalence remains of the same magnitude, while the coefficient on the interaction term between mother's education and regional HIV prevalence is about 30% smaller than before. The finding that the long term difference estimates look relatively similar to the cross-sectional results suggests that the bias resulting from the potential correlation between HIV and regional unobservable (and time invariant) factors is rather small.

In Table 3, we present further robustness checks. In columns 1 and 2 we include a country specific time trend in the regression in order to control for heterogeneous country experiences in terms of policy and economic development. While the results remain very similar to those in the Table 3, the inclusion of a country time trend leads to large increases in the standard errors on the regional HIV prevalence rate, which becomes insignificant. The results change only marginally when own HIV status is omitted in column 2, which is intuitive given the relatively small fraction of women with positive HIV status in our sample. Columns 3 and 4 of Table 3 repeat the specification in columns 1 and 2 but exclude the regions with the lowest HIV prevalence rates (under 1%) and highest HIV prevalence rates (over 7%) to test whether the results are robust to the exclusion of extreme values of the regional HIV prevalence. When focusing on regions with moderate HIV prevalence rates, the magnitude of the coefficient on regional HIV prevalence more than doubles in magnitude, indicating that the results are not driven by regions with very high HIV prevalence rates, whereas the coefficient estimate for the interaction between education and HIV remains about the same size.

[Table 3 here]

One of the questions unanswered in Table 2 is the effect of different education levels on fertility in the context of HIV. To allow for non-linear effects of education, we split education by level of attainment in columns 5 and 6 of Table 4. The categories used are "no schooling", which serves as the omitted reference group, "some primary", "primary completed", "some secondary", "secondary completed", and "tertiary". When using educational attainment in categories rather

than years of schooling, we find that up to primary education there is little evidence of a negative impact on fertility. The estimated effect of education decreases monotonically, and becomes negative for women having some secondary education, and significantly negative for women with completed secondary or tertiary education.

More importantly, the specifications with educational attainment show that the negative interaction between years of education and HIV prevalence is primarily driven by the relatively highly educated subpopulation with secondary and higher education. In the DHS surveys used in the long-term analysis, about 17.5% of all women fall into this group; among women in the prime age group (20-39), about 20% of women are in this educational group, with a majority (2/3) in the upper secondary, and a minority (1/3, or 6% of the total population) in the tertiary education group. Kenya has the largest fraction of women in this group (30%), while Senegal significantly lags behind the other countries in our sample with only 6% of women of age 20-39 with completed higher secondary or tertiary education.

3.2 Cross-sectional Evidence from Extended DHS Sample

While the data used in Section 3.1 offer a novel long-term perspective on the effect of HIV on fertility, the relatively small sample size of 32 regions raises external validity concerns. To show that the results are not driven by our specific selection of countries and regions, we pool all currently available DHS data sets in this section for a cross-sectional analysis. Table 4 provides basic summary statistics for the 21 countries with a recent DHS survey including HIV testing. Most of the countries are in Sub-Saharan Africa; outside of Sub-Saharan Africa, DHS data with HIV testing is available only for Cambodia, the Dominican Republic, Haiti, and India.

[Table 4 here]

The main difference to the empirical specifications in section 3.1 is that we cannot control for regional fixed effects in the cross-sectional approach chosen here. In this section, we control for country fixed effects instead, and identify the effect of HIV at the regional level. The underlying identification assumption here is that there are no unobservable effects that are correlated with regional HIV prevalence rates and also affect individual fertility decisions. While

we cannot rule out the possibility of confounding unobservables, our findings from section 3.1 suggest that the omitted variable bias resulting from the exclusion of regional fixed effects is rather small.[8]

Results

Table 5 summarizes the main results from the cross-sectional DHS data. In column 1, we regress the number of births in the last five years on own HIV status as well as on the regional HIV prevalence. We also include further controls for marital status of the woman and whether she ever lived together with a man, as well as the wealth index imputed by the DHS.

[Tables 5 and 6 here]

In a first step, we investigate the impact of HIV on both HIV-positive and HIV-negative mothers. HIV has been shown to reduce fecundity (Gray, Wawer et al. 1998), which is consistent with our results in column 1 of Table 5. When looking at the impact of group-level HIV-prevalence in Table 5, we find that there is a negative and statistically significant association between district-wide HIV-prevalence and births in the last five years. This negative association persists in column 2 of Table 5, where we run our main specification outlined in equation (11). Our point estimates imply that a 10 percentage point increase in regional HIV prevalence decreases 5-year birth rates by about 0.08 for women with no education, and by about 0.16 for women with 10 years of education.

Since our model implies a non-linear interaction between HIV and education, we interact HIV with education and squared education in column 3 of Table 5. While the linear term becomes insignificant, the quadratic interaction term is estimated with a negative sign and highly significant. To further investigate these non-linearities, we create education categories (primary, secondary, and higher, with no education constituting the baseline) and interact them with HIV prevalence at the regional level in column 4 of Table 5. The results unveil interesting behavioral

[8] An alternative specification is to explore the variations in HIV prevalence rates between urban and rural area in a given region. While this approach would allow controlling for regional unobservables, it renders the interpretation of the estimated coefficients hard, as rural and urban populations often live close by and are likely to share risk perceptions. Preliminary analysis suggests that the results from such specifications look very similar to the ones presented in the following section.

differentials: whereas the fertility differences across educational groups are rather small, the interaction between HIV and education becomes increasingly negative, and is statistically significant for all women with secondary or higher education – very similar to the results from the long-term analysis.

As further robustness check, we restrict the sample to HIV-negative women in column 5 of Table 5. In the HIV negative group, the effect of regional HIV prevalence on fertility appears negative, but is not significant. On the other hand, we find similar interaction effects as before: the reduction in fertility induced by HIV increases with each year of education by about .1 births. In the last column, we restrict the sample further to regions with intermediate HIV prevalence rates between 1 and 7 percent. While the impact of the regional HIV prevalence rate remains statistically insignificant, its interaction with education becomes larger, indicating that our results are attenuated rather than driven by the regions with very high and very low prevalence rates.

In Table 6, we use the wealth index available in the DHS as an alternative proxy for parental wage or labor market opportunities. The wealth index published as part of the DHS surveys uses principal component analysis of country-specific asset lists to classify households into wealth quintiles. In column 1 of Table 6, we estimate our main equation of interest with a linear asset index specification. The asset index shows to be a highly significant predictor of birth rates, which mothers from the highest quintile having 0.18 fewer births than mother from the bottom quintile. While the fertility effect of regional HIV prevalence is negative but insignificant as before, the interaction between HIV prevalence and the asset index is negative and highly significant. The magnitude of the HIV effects is slightly smaller than in the specifications with education. Our point estimates imply that a 10 percentage point increase in HIV reduces the 5-year birth rate by about 0.07 for women from the lowest quintile, and by about 0.14 for the women in the highest quintile.

In the second column of Table 6, we enter the five wealth quintiles (with the poorest women representing the omitted baseline category) separately in our empirical model. We find that the fertility reduction increases almost linearly in wealth, but that the interaction with HIV shows pronounced non-linearities: similar to the results from section 3.1, we find that the negative interaction between HIV and education is mostly driven by the top quintiles. For the two wealthiest quintiles, the interaction coefficient estimate is highly significant and sizeable,

with an estimated reduction of 0.07 and 0.05 births for a 10 percentage point increase in HIV prevalence.

In columns 3 and 4 of Table 6, we split the sample into sub-Saharan Africa and the rest of the world. There are few non-African countries with relevant DHS surveys: Cambodia, Dominican Republic, Haiti, and India, with India and the Dominican Republic dominating the sample in terms of number of observations. The overall patterns look similar: wealthier women have fewer children, and the interaction between HIV and the asset index is negative. However, the estimated fertility effects of HIV are significantly larger and estimated much less precisely in the non-African sample. This result is likely driven by the very low average prevalence rates in the non-African sample: three out of the four countries in this sample have prevalence rates of less than 1%, which makes HIV measurement noisy even at the regional level, and results more difficult to interpret.

3.3. Country-specific Evidence: Mali 2001 and 2006

For one country, Mali, there are two DHS surveys including HIV testing, one from 2001, and one conducted five years later, in 2006. Although we cannot link the 2001 HIV data to individual women[9], we can impute HIV prevalence rates at the regional level and use tem to analyze the correlation between changes in regional prevalence and fertility rates. As opposed to the previous section, the repeated cross-section here allows us control for unobservables at the regional level, and base our identification solely on the change in the regional HIV prevalence rate over the relatively short period between 2001 and 2006.

The main drawback of this approach is that we cannot focus solely on the preventive effect by excluding HIV-positive women from the analysis as in the previous sections of the paper. Given the low prevalence rate of under 2% in Mali, and the relatively small "direct" effect of individual HIV status estimated in the previous section, pooling all women should have a rather small effect on our results. As a robustness check, we compare the results for the 2006

[9] The first round of DHS surveys removed individual identifiers from all HIV files to guarantee privacy for respondents. In later waves, individual identifiers were made available, while privacy protection was guaranteed through the "scrambling" of individual IDs in the sampling file and the addition of error terms to the geographic information collected.

cross-section only when using all women in the analysis as compared to those with HIV-negative status only – as expected, the differences are very small.

Country Background

Mali is a poor, land-locked country in Western Africa. The large majority of its population is Muslim, potentially contributing to its low level of HIV prevalence (Gray 2004). The majority of Mali's women are illiterate, with about 35% of the population living in urban areas. Table 7 presents summary statistics for the 2001 and 2006 DHS surveys.

[Tables 7 & 8 here]

There are 9 regions of which 8 have data for both rural and urban areas in 2001 and 2006. In order to control for the wealth level of the household, we employ the DHS asset index (linear and in quartiles) as in the previous analysis. The main difference from the previous empirical specifications is that we have two independent measures of regional HIV prevalence rates at different points in time. This allows us to control for regional fixed effects, as well as a general time trend to account for other country specific changes occurring over the 5 year sample period.

Results

Table 8 shows the main estimation results for Mali. In the first column of Table 8, we regress the number of births of a woman in the last five years on regional HIV prevalence without interaction terms; we find a negative but statistically insignificant effect of HIV on fertility. In the second column or Table 8, we interact women's education with regional HIV prevalence. While the interaction term with HIV is negative, it does not appear statistically significant.

[Table 8 here]

Given the pronounced non-linearities found in the section 3.2, we test a quadratic specification in column 3 of Table 8. The results of this specification match the patterns outlined in the theoretical model: the magnitude of the fertility response to HIV increases with socioeconomic status, but at a decreasing rate. Given the large probability mass at zero years of education, we also estimate an alternative specification with a binary education variable, that

broadly divides women into those who have no schooling and those with at least one year of schooling ('some education'). In this simplified specification, we find a strong and negative interaction between having some schooling and regional HIV prevalence rate. The negative coefficient on the interaction term (-3.6) combined with the positive direct effect of HIV prevalence (4.1) implies that a change in HIV prevalence from 0 to 10 percent increases the number of births for women without education by .4, while the response of women with at least one year of education is virtually zero. Overall, the magnitude of the estimated HIV effects in Mali appear large and more positive relative to the results from the extended DHS sample in section 3.2; this may partially reflect the low average prevalence rates seen in Mali, but may also be seen as evidence for significant heterogeneity in the fertility response to HIV across countries and regions.

4. Summary and Discussion of Results

In this paper we investigate the interactions between HIV/AIDS and fertility both from a theoretical and an empirical perspective. We argue that the increases in adult mortality triggered by HIV are likely to lower fertility and to offset the positive fertility response generally associated with increased infant mortality. To test these claims empirically, we create a novel dataset combining all currently available DHS data sets with HIV testing with historical fertility data from the World Fertility Surveys. This data set allows us to investigate a large cross-section of regions as well as long-term changes in HIV and fertility under the inclusion of regional fixed effects and country-specific time trends.

The main result emerging from this paper is that the weak and statistically insignificant correlation between fertility and HIV prevalence at the regional level hides important heterogeneity across socioeconomic groups. We find a *positive* effect of HIV prevalence on fertility of non-educated mothers and mothers with primary schooling, but a *negative* fertility response for mothers with completed secondary schooling and higher. We find the result of behavioral heterogeneity to be consistent both across time and countries. They also hold when we focus on the short-term dynamics in Mali as the only country with two subsequent DHS surveys with HIV testing.

One concern regarding the empirical approach chosen in this paper is the exogeneity assumption regarding HIV. While there is still a lot of debate about the relative importance of the specific factors driving the diffusion of HIV, geographical distance to the origin of the virus, regional prevalence of circumcision, differential sexual behavior and economic activity and trade are generally viewed as key drivers of the epidemic. (Halperin and Bailey 1999; De Walque 2006; Werker, Ahuja et al. 2006; Oster 2008). All these factors, and particularly economic development, could have a direct (negative) effect on fertility, and thus potentially lead to a downward bias in our estimated HIV effects on fertility. Our short- and long-run longitudinal analysis suggests that at least those factors that are constant over time (such as location) or country specific (such as government policy and openness to trade) play a rather subordinate role in the interactions between HIV and fertility; however, alternative IV estimates might be a useful extension of the work presented in this paper.

The relatively large effects found in this paper raise the question about their empirical interpretation. While we show that the negative interaction between mothers' education and the fertility response to HIV is consistent with a simple theoretical model, where parents optimally trade off children against consumption, several other interpretations are consistent with our results. de Walque (2007) argues that more highly educated women are more able to absorb the information generated in HIV information campaigns, and thus use condoms more frequently than their less-educated peers. Similarly, Glick and Sahn (2006) find that the education gradient for prevention knowledge is substantial and seems to have increased over time in nine African countries. These findings suggest that the interaction term between regional HIV prevalence and the woman's education may at least partially reflect differences in risk perception and HIV knowledge. Similarly, one could also interpret our results of a differing fertility response across educational groups as evidence for dissimilar planning horizons or discounting rates. Oster (2007) argues that members of higher socio-economic classes are more likely to adjust their risk behavior in the face of HIV due to the larger consumption loss implied by shorter life spans. If children are viewed as risky investment, more educated parents will reduce their fertility by more than less educated parents – a mechanism very similar to the one outlined in the theoretical part of this paper.

Independent of their interpretation, the results presented in this paper imply that HIV may not trigger a large fertility response, but nevertheless significantly alter the composition of a country's population by increases in fertility among the poorest, and decreases in fertility among the wealthiest and most educated sub-population. The point estimates presented in this paper imply that an HIV prevalence rate of 10 increases fertility by up to 1.7 children for women with no formal education, and reduces fertility by up to 1 child per women with high school education.[10] With very moderate income growth over the last decades, this does not only aggravate poverty problems, but also implies a lower human capital stock in the long run, further worsening the direct human capital loss generated by HIV mortality (Dixon et al., 2001). Fortson (2008) highlights the reduced human capital investment generated by increased orphanhood in areas of high mortality. Our results suggest that the educational investment will be lower on average even among those children who do not lose their parents to the epidemic. Even if one is willing to believe that HIV may generate positive (labor market) effects in the short run, our results suggest that the medium to long-run effects of HIV will be negative through the change in the human capital composition of future generations.

[10] These numbers are taken from column 4, Table 3. A point estimate on HIV prevalence (2.8) * 0.1 prevalence implies an increase in 5 year birth. Since total fertility rate (TFR) is the number of birth given over the interval 15-44, we can approximate TFR by multiplying the 5-year-birth rate by a factor of six.

Appendix: Additional Derivations for the Theoretical Model

Deriving expression (8) with respect to w we get

$$\frac{\frac{\partial n^*}{\partial \rho}}{\partial w} = \frac{\alpha}{1-\alpha}\frac{1}{1-\alpha}\rho^{\frac{2-\alpha}{1-\alpha}} t_n^{\frac{1}{1-\alpha}} w^{\frac{1}{\alpha-1}} \left[t_n + \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \right]^{-2} +$$

$$-2\frac{1}{1-\alpha}\frac{\alpha}{1-\alpha}\rho^{\frac{2-\alpha}{1-\alpha}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \left[t_n + \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \right]^{-3} \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{1}{\alpha-1}} = \quad \text{(A.1)}$$

$$= \frac{\alpha}{(1-\alpha)^2}\rho^{\frac{2-\alpha}{1-\alpha}} t_n^{\frac{1}{1-\alpha}} w^{\frac{1}{\alpha-1}} \left[t_n + \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \right]^{-3} \left[t_n - \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \right]$$

This is positive as long as

$$t_n - \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} > 0 \quad \text{(A.2)}$$

Since ρ and t are constants, this terms is positive at $w = 0$, and decreases in w. The partial derivate is zero at

$$t_n = \rho^{\frac{1}{\alpha-1}} t_n^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}, \quad \text{(A.3)}$$

which implies

$$w = \frac{\rho^{\frac{1}{\alpha}}}{t_n}. \quad \text{(A.4)}$$

Plugging in into (7), this implies that the optimal turning point is given by

(A.5) $$n(\overset{)}{w}) = \frac{1}{t_n + \left(\frac{t_n}{\rho}\right)^{\frac{1}{1-\alpha}} (\frac{\rho^{\frac{1}{\alpha}}}{t_n})^{\frac{\alpha}{1-\alpha}}} = \frac{1}{2t_n}.$$

Since $\frac{1}{t_n}$ is the maximum number of kids raised, this implies that the turning point occurs at half the maximum number of children positive. As long as the top group does not achieve a fertility level of zero, the correlation between the fertility response and education must thus be negative. If the fertility range covered is large, the relation would thus be best approximated with a quadratic form; with fertility rates around 3, a linear approximation is likely to describe the relation between wage and the fertility response better than a quadratic term.

References

Adetunji, J. (2000). "Trends in under-5 Mortality Rates and the HIV/AIDS Epidemic." Bulletin of the World Health Organization **78**: 1200-1206.

Baker, D. P., J. M. Collins, et al. (2009). "Risk Factor or Social Vaccine? The Historical Progression of the Role of Education in HIV and AIDS Infection in Sub-Saharan Africa." Prospects: Quarterly Review of Comparative Education **forthcoming**.

Becker, G. S. (1965). "A Theory of the Allocation of Time." Economic Journal **75**: 493-517.

Becker, G. S. (1981). A Treatise on the Family. Cambridge, Harvard University Press.

Bergstrom, T. C. (1996). "Economics in a Family Way." Journal of Economic Literature **34**(4): 1903-1934.

Bloom, D. E. and A. Mahal (1997). "Does the AIDS Epidemic Threaten Economic Growth? ." Journal of Econometrics **77**(1): 105-124.

CDC (1981). "Pneumocystis Pneumonia --- Los Angeles." CDC Report **30**(21): 1-3.

Cuddington, J. T. (1993). "Modeling the Macroeconomic Impact of the AIDS in Tanzania." World Bank Economic Review **7**(2): 173-189.

De Walque, D. (2006). "Who Gets AIDS and How? The Determinants of HIV Infection and Sexual Behaviors in Burkina Faso, Cameroon, Ghana, Kenya and Tanzania." World Bank Policy Research Working Paper **3844**.

Dixon, S., S. McDonald, et al. (2001). "AIDS and economic growth in Africa: a panel data analysis." Journal of International Development **13**(4): 411-426.

Fortson, J. (2008). "The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS." Demography **45**(2): 303-322.

Fortson, J. (2008). "Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa." Mimeo.

Fortson, J. (2009). "HIV/AIDS and Fertility." American Economic Journal: Applied Economics **forthcoming**.

Gray, R. H., M. J. Wawer, et al. (1998). "Population Based Study of Fertility in Women with HIV-l Infection in Uganda." Lancet **351**(9096): 98-104.

Halperin, D. T. and R. C. Bailey (1999). "Male Circumcision and HIV Infection: 10 Years and Counting." The Lancet **354**: 1813-1815.

International Labour Office (2004). HIV-AIDS and Work: Global Estimates, Impact and Response. Geneva, The ILO Programme on HIV/AIDS and the World of Work.

Juhn, C., S. Kalemli-Ozcan, et al. (2008). "HIV and Fertility in Africa: First Evidence from Population Based Surveys." NBER Working Paper **14248**.

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

Kalemli-Ozcan, S., C. Juhn, et al. (2008). "HIV and Fertility in Africa: First-Evidence from Population Based Surveys." NBER Working Paper **14428**.

Kohler, H.-P., J. R. Behrman, et al. (2007). "Social Networks and HIV/AIDs Risk Perceptions." Demography **44**(1): 1-33.

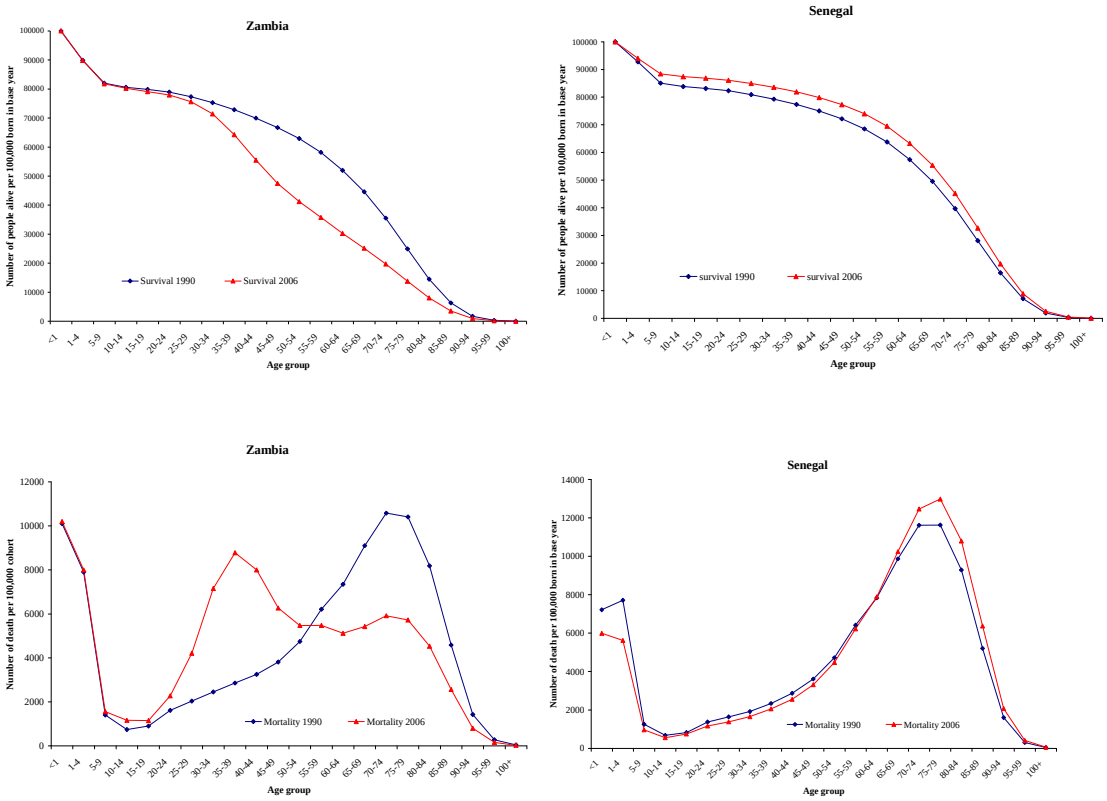
Kotlikoff, L. J. and A. Spivak (1981). "The Family as an Incomplete Annuities Market." Journal of Political Economy **89**(2): 372-391.

Oster, E. (2007). "HIV and Sexual Behavior Change: Why Not Africa?" NBER Working Paper **13049**

Oster, E. (2008). "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa." Mimeo.
Oster, E. (2009). "Estimating HIV Prevalence and Incidence in Africa from Mortality Data " Mimeo.
Schultz, T. P. (1997). Demand for Children in Low Income Countries. Handbook of Population and Family Economics. M. R. Rosenzweig and O. Stark. Amsterdam, North-Holland. **1A:** 643.
UNAIDS (2008). 2008 Report on the Global AIDS Epidemic, UNAIDS.
Walker, N., B. Schwartländer, et al. (2002). "Meeting International Goals in Child Survival and HIV/AIDS." The Lancet **360**(9329): 284-289.
Werker, E., A. Ahuja, et al. (2006). "Male Circumcision and HIV/AIDS: The Macroeconomic Impact of a
Health Crisis." Mimeo.
World Health Organization (2006). Life Tables for WHO Member States.
Young, A. (2005). "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations." Quarterly Journal of Economics **120**(2): 423-466.
Young, A. (2007). "In Sorrow to Bring Forth Children: Fertility amidst the Plague of HIV." Journal of Economic Growth **12**(4): 283-327.

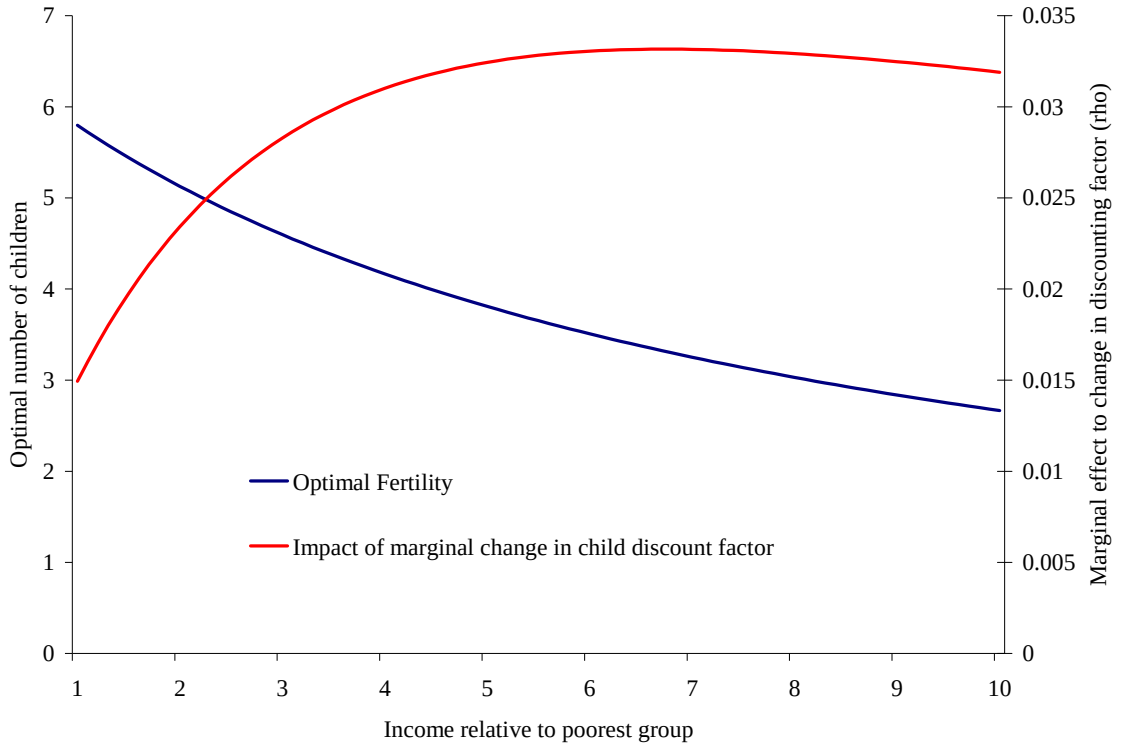
Figures

Figure 1: Changes in Survival and Mortality Rates 1990-2006: Zambia vs. Senegal



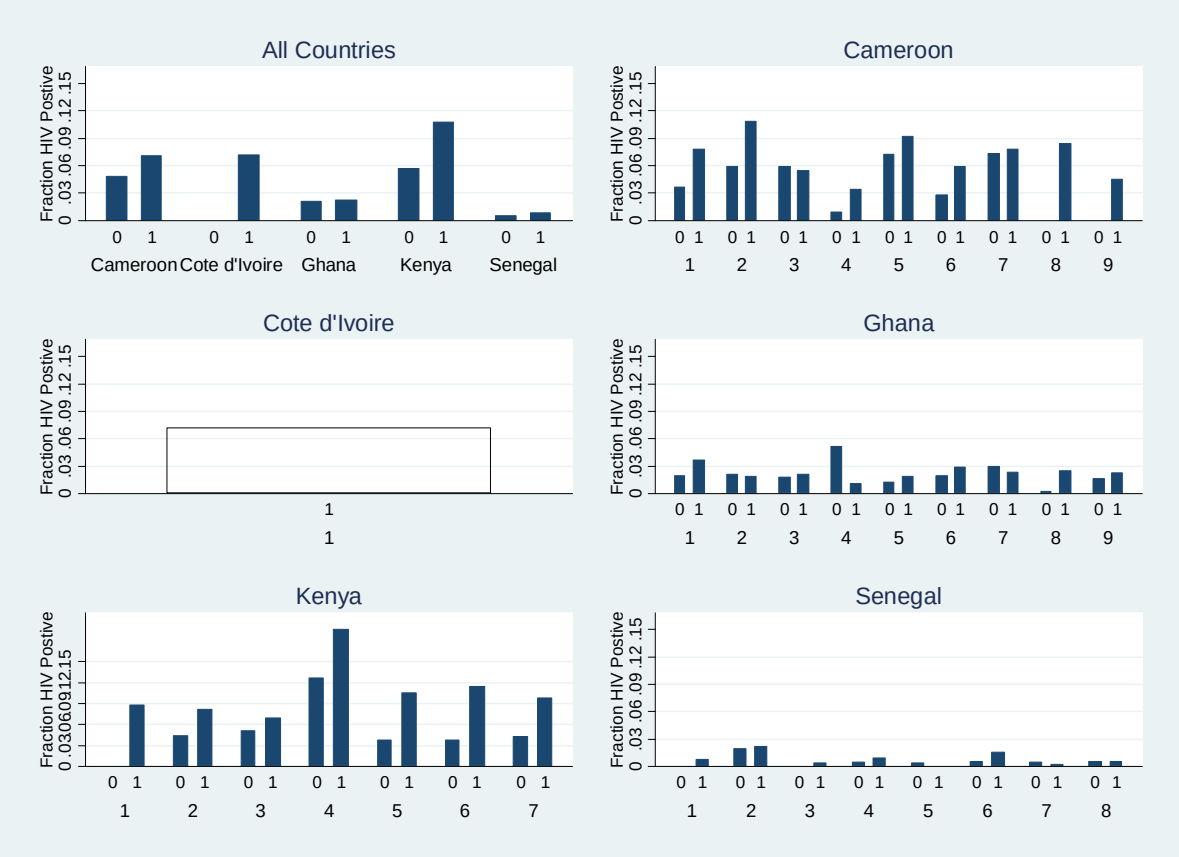
Source: Life Tables for WHO Member States (World Health Organization 2006)

Figure 2: Optimal Fertility and Marginal Effect of Increase in Adult Life Expectancy



Calibration parameters: $\alpha = 0.5$, $\rho = 1$, $t = 0.15$, baseline wage is 1.

Figure 3: HIV Prevalence in Urban and Rural Areas



Notes: 0 is rural, 1 is urban. Bottom line indexes the region number. For Cote d'Ivoire, we have only one (urban) region, the metropolitan area of Abidjan.

Tables

Table 1: Descriptive Statistics

	WFS		DHS	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Age	28.54	9.81	28.11	9.52
Years of education	2.92	3.83	5.26	4.47
Partner	0.75	0.43	.55	.50
Urban residence	0.30	0.46	.44	.50
HIV prevalence	0.00	0.00	.05	.04
Number of observations	27,319		17,861	

Notes: The HIV prevalence rates are assumed to be zero in the WFS.

Table 2: HIV, Education and Fertility: Long-term Analysis using WFS and DHS

	Dependent Variable: Number of children born in the last 5 years				
	(1)	(2)	(3)	(4)	(5)
Regional HIV prevalence	-0.516	1.113*	1.297*	2.798***	2.377***
	(0.64)	(0.62)	(0.66)	(0.94)	(0.82)
HIV * education		-0.225***	-0.242***	-0.375***	-0.254***
		(0.049)	(0.053)	(0.061)	(0.056)
Age	0.218***	0.220***	0.221***	0.313***	0.285***
	(0.0080)	(0.0081)	(0.0082)	(0.010)	(0.012)
Age squared	-0.00368***	-0.00369***	-0.00372***	-0.00552***	-0.00506***
	(0.00013)	(0.00013)	(0.00013)	(0.00017)	(0.00022)
Years of education	-0.0176***	-0.0130***	-0.0129***	-0.0128***	-0.0189***
	(0.0015)	(0.0019)	(0.0019)	(0.0022)	(0.0032)
Married	0.583***	0.583***	0.588***	0.636***	0.572***
	(0.026)	(0.027)	(0.027)	(0.033)	(0.041)
Urban	-0.104***	-0.104***	-0.101***	-0.133***	-0.224***
	(0.017)	(0.017)	(0.016)	(0.022)	(0.031)
DHS sample	-0.0804**	-0.0984***	-0.0963**	-0.134***	-
	(0.038)	(0.036)	(0.036)	(0.048)	-
Constant	-2.183***	-2.222***	-2.253***	-3.322***	-2.901***
	(0.11)	(0.11)	(0.11)	(0.15)	(0.16)
Sample restrictions	None	None	HIV neg.	HIV neg. 17< age< 41	HIV neg. 17< age< 41 DHS only
Observations	45180	45180	44367	32133	12209
R-squared	0.30	0.30	0.31	0.22	0.27

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parentheses are clustered at the regional level.

Table 3: Long-term Analysis using WFS and DHS: Robustness Checks

	Dependent Variable: Number of children born in the last 5 years					
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.220***	0.221***	0.219***	0.220***	0.223***	0.224***
	(0.00786)	(0.00789)	(0.0119)	(0.0124)	(0.00819)	(0.00825)
Age squared	-0.00370***	-0.00371***	-0.00371***	-0.00374***	-0.00372***	-0.00373***
	(0.000127)	(0.000128)	(0.000207)	(0.000214)	(0.000133)	(0.000134)
Years of education	-0.0128***	-0.0126***	-0.0171***	-0.0165***		
	(0.00174)	(0.00175)	(0.00304)	(0.00312)		
HIV status	-0.121***		-0.139***		-0.128***	
	(0.0264)		(0.0356)		(0.0257)	
HIV regional prevalence	1.351	1.384	3.186***	3.521***	0.317	0.343
	(1.095)	(1.070)	(0.585)	(0.581)	(0.555)	(0.591)
HIV * education	-0.220***	-0.235***	-0.198**	-0.226***		
	(0.0404)	(0.0437)	(0.0692)	(0.0714)		
DHS wave	0.0574	0.0612			-0.116***	-0.115***
	(0.0928)	(0.0903)			(0.0260)	(0.0255)
Schooling categories						
Some primary					0.0655	0.0654
					(0.0394)	(0.0393)
Primary completed					0.0251	0.0265
					(0.0224)	(0.0225)
Some secondary					-0.0288	-0.0271
					(0.0245)	(0.0244)
Secondary completed					-0.116***	-0.112***
					(0.0281)	(0.0278)
Tertiary					-0.323***	-0.324***
HIV interactions					(0.0372)	(0.0366)
Some primary					0.0572	0.292
					(0.964)	(1.033)
Primary completed					0.463	0.475
					(0.595)	(0.631)
Some secondary					-0.297	-0.398
					(0.590)	(0.622)
Secondary completed					-1.738**	-1.869**
					(0.801)	(0.835)
Tertiary					-1.540**	-1.661**
					(0.688)	(0.739)
Constant	-2.302***	-2.319***	-2.199***	-2.238***	-2.413***	-2.433***
	(0.104)	(0.105)	(0.129)	(0.136)	(0.0994)	(0.0997)
Sample	All	HIV neg.	All	HIV neg.	All	HIV neg.
Observations	45180	44367	10822	10405	45180	44367
R-squared	0.308	0.310	0.327	0.335	0.302	0.304

Notes: All columns control for urban residence and marital status. Column 1 and 2 include a country specific time trend. In columns 3 and 4, the sample is restricted to regions with prevalence rates above 1 and below 7% (DHS only). In columns 5 and 6, educational categories rather than years of schooling are used. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses are clustered at the regional level.

Table 4: Summary Statistics Extended DHS Sample

Country	Year	Births last 5 years	HIV prevalence rate	Education in years	Number of observations
Burkina Faso	2003	0.854	0.020	1.423	4421
Cambodia	2005	0.497	0.006	3.899	8187
Cameroon	2004	0.792	0.068	5.671	5148
Congo DR	2007	0.945	0.017	5.232	4609
Cote d'Ivoire	2005	0.706	0.056	2.300	4535
Dominican Republic	2007	0.418	0.008	8.450	25415
Ethiopia	2005	0.765	0.024	2.595	6812
Ghana	2003	0.686	0.026	5.386	5683
Guinea	2005	0.830	0.018	1.465	3843
Haiti	2005	0.563	0.025	4.887	5220
India	2005	0.409	0.004	6.172	52847
Kenya	2003	0.748	0.084	7.051	4039
Lesotho	2004	0.560	0.261	7.211	3517
Liberia	2007	0.829	0.023	3.451	6467
Malawi	2004	0.976	0.147	4.525	2864
Mali	2006	0.972	0.015	1.451	4743
Niger	2006	1.004	0.009	1.445	4441
Rwanda	2005	0.764	0.039	3.881	5663
Senegal	2005	0.741	0.011	2.456	4466
Swaziland	2006	0.576	0.314	7.931	4584
Zimbabwe	2005	0.599	0.207	7.795	7503

Table 5: Results for Extended DHS Data Set

	Dependent Variable: Number of children born in the last 5 years					
	(1)	(2)	(3)	(4)	(5)	(6)
Own HIV status	-0.088***	-0.090***	-0.089***	-0.088***	-	-
	(0.018)	(0.018)	(0.018)	(0.018)		
Regional HIV prevalence	-1.303***	-0.832*	-0.987**	-1.011**	-0.645	-0.771
	(0.446)	(0.463)	(0.464)	(0.477)	(0.505)	(0.971)
HIV*education	-	-0.082***	0.016	-	-0.108***	-0.234*
		(0.023)	(0.048)		(0.025)	(0.127)
HIV*education squared	-	-	-0.007***	-	-	-
			(0.002)			
Education	-0.004	-0.003	-0.012**	-	-0.002	-0.008*
	(0.003)	(0.003)	(0.005)		(0.003)	(0.005)
Education squared	-	-	0.001***	-	-	-
			(0.000)			
HIV* mother primary education				-0.003		
				(0.219)		
HIV* mother secondary education				-0.441		
				(0.283)		
HIV* mother higher education				-1.208***		
				(0.361)		
Mother primary education				-0.056***		
				(0.021)		
Mother secondary education				-0.044		
				(0.031)		
Mother higher education				-0.036		
				(0.041)		
Wealth index	-0.053***	-0.052***	-0.052***	-0.053***	-0.052***	-0.046***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
Constant	-0.914***	-0.930***	-0.898***	-0.887***	-0.934***	-1.366***
	(0.152)	(0.151)	(0.156)	(0.153)	(0.154)	(0.116)
Sample restrictions	None	None	None	None	HIV-neg.	HIV-neg. .01<p<.07
Observations	167664	167664	167664	167664	161223	54687
R-squared	0.342	0.343	0.343	0.343	0.351	0.374

Notes: All specifications control for age, age squared, urban residence, and marital and cohabitation status.
*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses are clustered at the regional level.

Table 6: Extended DHS Sample with Wealth Index Interactions

	Dependent Variable: Number of children born in the last 5 years			
	(1)	(2)	(3)	(4)
Own HIV status	-0.087***	-0.088***	-0.129***	-0.115***
	(0.018)	(0.018)	(0.017)	(0.027)
Regional HIV prevalence	-0.689	-1.002**	-0.211	-0.942
	(0.464)	(0.462)	(0.276)	(4.886)
DHS wealth index	-0.046***			
	(0.003)			
HIV * wealth index	-0.182***			
	(0.042)			
2nd asset quintile		-0.078***	-0.022*	-0.097***
		(0.009)	(0.012)	(0.016)
3rd asset quintile		-0.116***	-0.045***	-0.147***
		(0.011)	(0.013)	(0.017)
4th asset quintile		-0.142***	-0.061***	-0.160***
		(0.013)	(0.018)	(0.020)
5th asset quintile		-0.201***	-0.151***	-0.190***
		(0.012)	(0.022)	(0.024)
HIV * 2nd asset quintile		0.087	-0.195	-2.020
		(0.107)	(0.124)	(1.908)
HIV * 3rd asset quintile		-0.099	-0.398***	-1.575
		(0.105)	(0.115)	(2.360)
HIV * 4th asset quintile		-0.465***	-0.707***	-4.643*
		(0.138)	(0.154)	(2.750)
HIV * 5th asset quintile		-0.611***	-0.500***	-5.445*
		(0.172)	(0.189)	(3.117)
Sample Restrictions	None	None	Sub-Saharan Africa	Cambodia, DR, Haiti, India
Observations	167664	167664	80530	91669
R-squared	0.343	0.340	0.351	0.301

Notes: All specifications control for age, age squared, education, urban residence, and marital and cohabitation status.

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses are clustered at the regional level.

Table 7: Summary Statistics for Mali in 2001 and 2006

Year	2001			2006		
Region	HIV prevalence	Births last 5 years	Education in years	HIV prevalence	Births last 5 years	Education in years
Kayes	0.019	1.080	0.504	0.015	1.043	0.537
Koulikoro	0.021	1.111	0.702	0.012	1.087	1.249
Sikasso	0.009	1.149	0.691	0.011	1.111	1.027
Segou	0.013	1.146	0.596	0.018	1.017	1.462
Mopti	0.017	1.030	0.432	0.022	0.955	0.896
Tombouctou	0.011	0.984	0.785	0.003	0.934	0.912
Gao	0.004	0.953	1.428	0.003	0.895	1.225
Kidal	0.015	0.957	1.676	0.010	0.958	1.023
Bamako	0.028	0.660	4.413	0.022	0.707	4.144

Table 8: HIV, Education and Fertility: Mali 2001/2006

Dependent Variable	Number of children born in the last 5 years			
	(1)	(2)	(3)	(4)
Regional HIV prevalence	3.130	3.298	3.287	4.062
	(3.671)	(3.704)	(3.600)	(3.930)
Education * HIV		-0.130	-1.179**	
		(0.177)	(0.363)	
Education squared * HIV			0.130***	
			(0.028)	
HIV * Some education				-3.661**
				(1.156)
Some education				-0.048*
				(0.023)
Urban	-0.030	-0.031	-0.033	-0.035
	(0.029)	(0.029)	(0.029)	(0.030)
Age	0.229***	0.229***	0.230***	0.229***
	(0.008)	(0.008)	(0.008)	(0.008)
Age squared	-0.004***	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Education	-0.023***	-0.021***	0.013	
	(0.002)	(0.004)	(0.009)	
Education squared			-0.004***	
			(0.001)	
Married	0.729***	0.729***	0.733***	0.742***
	(0.045)	(0.045)	(0.045)	(0.044)
Formerly cohabitating	0.334***	0.334***	0.337***	0.345***
	(0.065)	(0.065)	(0.065)	(0.066)
Wealth index	-0.046***	-0.046***	-0.044***	-0.060***
	(0.006)	(0.006)	(0.006)	(0.006)
Constant	-2.586***	-2.590***	-2.607***	-2.606***
	(0.094)	(0.093)	(0.095)	(0.099)
Observations	27374	27374	27374	27374
R-squared	0.316	0.316	0.317	0.314

Notes: *** p<0.01, ** p<0.05, * p<0.1,
Robust standard errors in parentheses are clustered at the regional level.