

How Subjective Beliefs about HIV Infection Affect Life-Cycle Fertility: Evidence from Rural Malawi*

(Job Market Paper)

Gil Shapira[†]

November 14, 2010

Abstract

In this paper, I study the effect of subjective beliefs about HIV infection on fertility decisions in a context of high HIV prevalence and evaluate the impact of different policy interventions, such as HIV testing programs, informational campaigns, and antiretroviral therapy distribution, on fertility, child mortality and orphanhood. I develop a dynamic discrete-choice life-cycle fertility model in which expectations about the life horizon and child survival depend on a perceived HIV infection hazard process, which is allowed to differ from the actual hazard. In the model, women form and update beliefs about their HIV status and about their own and their children's survival in future periods. Women also update their beliefs when their HIV status is revealed by an HIV test. Model parameters are estimated by maximum likelihood with longitudinal data from the Malawi Diffusion and Ideation Change Project, which contain family rosters, information on HIV testing, and measures of subjective beliefs about own HIV status. I use the model to evaluate the effect of HIV on fertility by simulating behavior in an environment without HIV. Results show that the presence of HIV reduces the average number of births a woman has during her life-cycle by 0.24. I also find that HIV testing is effective at reducing fertility of infected women, leading to a reduction of child mortality and orphanhood.

*I thank my thesis advisors Jere Behrman, Áureo de Paula, Kenneth Wolpin, and especially my main advisor Petra Todd for their dedication, support and hours of their time. I am also grateful to Larbi Alaoui, Drew Griffen, Clement Joubert, Nirav Mehta, Michelle Poulin, Shalini Roy, Panos Stavrinides, and participants of the empirical microeconomics lunch and seminar at the University of Pennsylvania for helpful comments. I gratefully acknowledge financial support from the William and Flora Hewlett Foundation.

[†]Department of Economics, University of Pennsylvania. Email: gshapira@econ.upenn.edu

1 Introduction

Both fertility and HIV prevalence rates in Malawi are among the highest in the world, with the total fertility rate at 6 births per woman and the HIV prevalence rate at 12 percent.¹ Malawian women make fertility decisions in an environment characterized by high adult and child mortality, exacerbated by mother-to-child HIV transmission. Out of a population of about 15 million, it is estimated that 68,000 die annually from AIDS and that 560,000 children under the age of 17 have lost at least one parent to the disease.²

There are many policy interventions aimed at reducing HIV in Malawi and other Sub-Saharan African countries. These include HIV testing programs, information campaigns, and antiretroviral distribution programs. Evaluating the effects of such policies on outcomes such as number of births, child mortality, and orphanhood requires an understanding of how women's fertility decisions are affected by the presence of HIV.

An important aspect of the environment in Malawi is that women are typically uncertain regarding their own HIV status. An infected person can live for many years with no symptoms, and testing was not widely available until very recently. The median survival time after infection, without treatment, is about 10.4 years.³ During most of this time, an infected person is in a clinical latency stage and experiences few or no symptoms.⁴

In addition to being uncertain about own HIV status, women often express beliefs about HIV risk that differ substantially from actual risk. Studies using the Malawi Diffusion and Ideation Change Project data show that individuals in rural Malawi tend to overestimate both the the probability of being HIV-infected (Anglewicz and Kohler, 2009) and the HIV prevalence in their community (Anglewicz, 2007). Anglewicz and Kohler (2009) attribute these high risk assessments to overestimated probabilities of transmission. More than 95 percent of respondents believe that transmission from a single instance of unprotected in-

¹Malawi Demographic and Health Surveys (2004)

²UNAIDS/WHO/UNICEF Epidemiological Fact Sheets (2008)

³Todd et al. (2007). Estimate for infected adults in Eastern and Southern Africa

⁴Morgan et al. (2002), for example, find median time from infection to AIDS to be 9.4 years and median time from AIDS to death to be 9.2 months in rural Uganda.

tercourse with an infected person is highly likely or certain; however, studies estimate that it can be as low as 1 per 1,000 encounters in the absence of an increased viral load (Gray et al., 2001).⁵

Women’s perceptions of HIV risk and of their own HIV status affect beliefs about their own and their children’s life expectancy, which in turn may influence life-cycle fertility choices. In this paper, I study the determinants of women’s reproductive decisions in Malawi, taking into account uncertainty about HIV status and differences between perceived and actual HIV infection risk. I investigate how HIV affects fertility and evaluate the impact of different policy interventions, such as HIV testing programs and provision of antiretroviral therapy, on fertility, child mortality, and orphanhood.

To this end, I develop a dynamic discrete-choice life-cycle fertility model in which expectations about the life horizon and child survival depend on a perceived infection hazard. A woman makes annual pregnancy decisions from the time of marriage until she becomes infecund. She maximizes utility, which depends on her number of children, household consumption, and pregnancies, subject to a per-period budget constraint. The woman faces uncertainty regarding future income, HIV status, and the survival of herself and her children in future periods.

A woman’s perceived infection hazard is allowed to differ from her actual infection hazard to reflect the misperceptions about HIV risk observed empirically. The perceived hazard rate for each period is a function of a woman’s characteristics, such as her age, region of residence, marital status, and schooling level. To account for unobservable factors, the hazard rate also incorporates heterogeneity in the form of a discrete number of unobserved types.⁶

Given that HIV is initially asymptomatic, the model assumes that a woman does not

⁵The viral load is high in the few weeks following infection and increases again as an infected person develops AIDS. Infectivity increases with viral load as well as with other conditions, such as the existence of other STDs. Powers et al. (2008) review the studies estimating HIV infectivity and discuss different factors that increase infectivity. Note however that even a low transmission rate such as 0.001 can translate into a nontrivial probability of infection during a year of partnership. Gray et al. (2001) find average frequency of intercourse to be about 106 acts per year, implying about a 10 percent chance of transmission given this transmission rate.

⁶Heckman and Singer (1984)

observe realizations of the infection process and therefore does not know her HIV status. Assuming she knows the mortality process associated with HIV infection, however, survival during each additional period gives her information about her status. Specifically, she reduces her subjective probabilities of having become infected in each of the past periods based on the fact that she is still alive. According to these probabilities and given the mortality process, she updates her survival expectations. HIV infection also increases child mortality probabilities through mother-to-child transmission. In the model, the woman also updates expectations about the survival of each of her children depending on the probability assigned to their having been infected at the time of birth.

The dynamic fertility model is estimated using the Malawi Diffusion and Ideation Change Project (MDICP) dataset, a rich longitudinal dataset collected in rural areas of three different districts of the country. The data contain extensive information on more than 4,000 individuals at the individual and household level. The three sampled regions vary significantly in several aspects that are potentially relevant for the analysis, such as HIV prevalence rates, polygamy rates, and schooling levels. A unique feature of the MDICP data is that they include measures of subjective expectations regarding a range of outcomes, including the likelihood respondents assigned to being HIV-infected at the time of the interview. The expectations data were collected using a novel bean-counting method, further described below, which is appropriate for populations with low levels of numeracy. Delavande and Kohler (2009) find that that the reported subjective expectations follow basic properties of probabilities and that the assessments of HIV-infection vary meaningfully with observable characteristics associated with different levels of HIV prevalence. I use a subsample of 1006 married women who were interviewed at least once during the 2006 and 2008 rounds, when family rosters and subjective expectations were collected.

Since 2004, each round of data collection included HIV testing of respondents and the prevalence rate was found to be about seven percent.⁷ Although individuals who received

⁷HIV prevalence is lower in rural areas than in urban areas. The prevalence in the MDICP sample is lower than the rural DHS prevalence because the sample does not include peri-urban areas such as trading

positive test results assign significantly higher likelihood to being infected two years later relative to individuals who received negative test results, some individuals who tested positive later assign a probability of less than one to being infected. I therefore assume that a woman assigns a probability to the accuracy of the test. Given the test result, the woman updates the probabilities assigned to infection and the corresponding probabilities of survival.

I structurally estimate the model parameters using maximum likelihood and then use the estimated model to perform several counterfactual simulations. First, I simulate life-cycle fertility in an environment with no HIV exposure. The results indicate that the presence of HIV has a negative effect on fertility on average. Overall, women in the no-HIV environment have on average 0.24 more births over their life-cycle. The magnitude of the effect is higher for groups with higher perceived exposure to HIV infection. I also simulate the effects of HIV testing programs and of the distribution of antiretroviral therapy on fertility outcomes. Although they are not necessarily designed to influence fertility, they have the potential to affect reproductive choices by altering beliefs about HIV status, mother-to-child transmission rates, and life expectancy. Such policy interventions are already implemented in Malawi and are expanding in scope.

1.1 Related Literature

Several studies analyzed the response of fertility in Sub-Saharan Africa to the HIV/AIDS epidemic and reached mixed conclusions. Young (2005) finds HIV to have a negative effect on fertility in South Africa, which he attributes to an unwillingness to engage in unprotected sexual activity and increasing labor opportunities for women because of scarcity of labor. Using time series cross-country data on fertility and HIV prevalence rates, Young (2007) finds a similar effect. Kalemli-Ozcan (2006) uses similar types of data and finds a positive effect of the epidemic on fertility. Fortson (2009) and Juhn et al. (2008) use cross-country data from the Demographic and Health Surveys (DHS). These surveys are nationally representative centers (Obare et al., 2009).

and contain results of HIV testing of respondents. Both studies find lower fertility rates of HIV-infected women than of HIV-uninfected women; however, they find no effect of local prevalence rates on the fertility of uninfected women and an overall insignificant aggregate effect of HIV on fertility. Fink and Linnemayr (2009), linking historical data from World Fertility Surveys (WFS) with the DHS data, argue that while HIV does not have a significant effect on aggregate fertility levels, it affects differently women depending on their educational attainment. They find that, in the presence of HIV, more educated women reduce fertility more than uneducated women. My analysis differs from these studies both in the empirical approach and in the type of data used in the analysis. The data on beliefs allow me to exploit heterogeneity within communities, which the use of data on local HIV prevalence only does not allow.

Many economists have studied the determinants of fertility in different environments.⁸ My analysis is most closely related to studies that model fertility decision-making as a sequential process (Heckman and Willis, 1976) and studies of fertility in environments with non-negligible infant and child mortality risk (Wolpin, 1984; Sah, 1991; Mira, 2007). Wolpin (1984) presents an estimable dynamic discrete-choice fertility model in an environment where infant survival is uncertain and uses the model to study the response of fertility choices to experienced infant mortality. Mira (2007) uses a similar modeling framework and extends it by introducing heterogeneity in infant mortality risk across mothers. Parents in his model learn about a family-specific component of infant mortality risk throughout their life-cycle. Fertility choices are influenced by how the parents adapt to the information received from infant survival and mortality. My model builds on the Wolpin (1984) modeling framework and includes Bayesian updating of mortality expectations. It differs in that it includes mortality of both mothers and children and the learning process is different from that of Mira (2007). In my model, women update their beliefs about getting infected with HIV in different periods of their life-cycle. These beliefs about HIV status translate into expectations

⁸Joseph Hotz et al. (1997), Schultz (1997), and Wolpin (1997) survey the literature on fertility

about the survival of both mothers and children in future periods.

2 Data

2.1 Malawi Diffusion and Ideation Change Project

Malawi is a landlocked country located in Southeast Africa. It has a population of about 15 million, comprised of different ethnic and religious groups. 81 percent of the population lives in rural areas and relies mostly on subsistence agriculture.⁹ The Malawi Diffusion and Ideation Change Project (MDICP) data have been collected since 1998 in rural areas of three districts in Malawi: Balaka in the south of the country, Mchinji in the center, and Rumphi in the north.¹⁰ The different rounds of the longitudinal dataset contain extensive information on more than 4,000 men and women, at the individual, household and community levels.

A unique feature of the MDICP dataset is that it includes measures of subjective likelihoods respondents assign to being HIV-positive. The subjective expectations were collected using an elicitation methodology developed by Delavande and Kohler (2009) for a developing country context with low levels of literacy and numeracy.¹¹ Respondents were provided with ten beans and a plate. They were asked to allocate different number of beans on the plate to express the likelihood that different events will be realized. The respondents were instructed that zero beans reflect certainty that an event will not happen, more beans reflect higher likelihood that an event happens, and that ten beans imply certainty about the event happening. The likelihood assigned to being HIV-positive is measured by asking: “Pick the number of beans that reflects how likely you think it is that you are infected with HIV/AIDS now.” Delavande and Kohler (2009) find that reported subjective expectations follow basic properties of probabilities and that the assessments of HIV-infection vary meaningfully with

⁹data.worldbank.org

¹⁰Detailed information on the Malawi Diffusion and Ideation Change Project can be obtained at <http://www.malawi.pop.upenn.edu/>.

¹¹Delavande et al. (2010) review existing subjective expectations data from developing countries.

observable characteristics associated with different levels of HIV prevalence.

Since 2004, HIV testing has been offered to all respondents during data collection. The take-up rates of the tests were high, above 90 percent in all waves. The HIV prevalence rate in the sample was 6.9 percent (Obare et al., 2009). In 2004, test results were available after five to seven weeks. The testing component of the survey was linked to an experiment studying the incentives for voluntary consulting and testing (VCT) uptake, studied by Thornton (2008). Respondents were assigned vouchers for a small monetary reward, redeemable upon return to temporary VCT sites where results were provided. The VCT sites were set up such that all respondents' homes are within five kilometers distance from at least one site. Approximately 70 percent of those tested chose to pick up their results. In 2006 and 2008, rapid blood tests were adopted, eliminating the time delay between testing and provision of results.

For my analysis, I am using a subsample of married women in their first marriage, who did not become pregnant from a relationship with men other than their future husbands prior to marriage. I restrict the sample to women in unbroken first marriages because I abstract from modeling any decisions related to marriage or partnership. I also restrict the sample to include only women who were interviewed in at least one of the 2006 and 2008 rounds. Data collected in these rounds include family rosters, containing information on all children of respondents, and the elicitation of subjective expectations. I start with 2561 women who were interviewed in these rounds. I exclude 798 who were not in their first marriage. These women were either married more than once (736), never married (15), divorced (18), or widowed (29). I drop from the sample an additional 217 women who reported having children with different fathers. I also exclude 369 women who were born before 1960, because the last age of fertility in the model is assumed to be 44 and 2004 is the earliest year included for fertility outcomes. After excluding additional women for missing information, the estimation sample consists of 1006 women.

Tables 1 to 4 and Figures 1 and 2 provide descriptive statistics for the variables used

in the analysis. The average age of women in the sample is 26.6 in 2004. 29 percent of the sample resides in Balaka (south), 35 percent in Mchinji (center), and 36 percent in Rumphi (north). The median number of years of schooling is 5. The women in Balaka have the lowest schooling levels with 31 percent of women never having attended school and only three percent having attended some secondary school. The women in Rumphi have the highest schooling levels, with 99 percent of women ever having attended school and 26 percent some secondary school. The average age of marriage in the overall sample is 17.5 and is similar across regions. Polygamy is most prevalent in Rumphi, with 33 percent of the women in the sample from that region married to a polygamous husband. The share of polygamous women is 27 percent in Mchinji and 20 percent in Balaka. As shown in Table 2, HIV testing take-up rates were high in all the rounds: 87.9 percent in 2004, 93.3 percent in 2006, and 94.6 percent in 2008. The percentage of women who tested positive was 2.9 percent in 2004, 3.1 percent in 2006, and 3.7 percent in 2008.

Figure 1 depicts the distribution of the likelihood women assigned to being HIV-infected in the 2006 and 2008 rounds (pooled), measured by beans on a scale of zero to ten. 50 percent of the reports were of zero beans. The percentage of women who chose each category decreases with the number of beans, except for the 5-bean category. Fewer than one percent chose the 10-bean category. Table 3 shows the average number of beans. The overall sample average is 1.52 beans. The average number of beans chosen in each region conforms to the ranking of HIV prevalence in the general MDICP sample. The averages are 1.78 beans in Balaka, 1.69 in Mchinji, and 1.16 in Rumphi. The 2004 HIV prevalence rates for these regions are 7.9 percent in Balaka, 6.4 percent in Mchinji, and 4.4 percent in Rumphi. The table also shows that average beliefs decrease with schooling and are higher for women married to a polygamous husband.

Figure 2 depicts the distribution of the likelihood women assigned to be HIV-infected after receiving HIV test results. Because the testing component of the survey was conducted after the interview components, these are beliefs reported two years after the tests (the subsequent

round). Because of low HIV prevalence and high attrition of infected women, there are only 18 women who report their beliefs two years after receiving a positive test result. The average number of beans allocated by these women is 4.67 beans, which is significantly higher than the average of 1.5 beans reported by women who received a negative test result. Most of the women who received a positive test result assign at least some likelihood to not being HIV infected.

I obtain information about births from family rosters that were collected in 2006 and 2008. Respondents were asked to list all of the children ever born to them, but year of birth is missing for children who died more than two years before the interview. I therefore use only birth outcomes reported for the years 2005 to 2008. Depending on the year of marriage, year of birth, and the rounds in which the woman participated, I observe between 1 and 4 potential fertility years for women of ages 16 to 45. In total, I observe 3,148 years of potential fertility with births in 920 of them (29.2 percent). Table 4 shows the annual birth probabilities of married women by 5-year age groups. The annual birth probabilities decrease from 0.405 for ages 16 to 20 to 0.06 for ages 41 to 45.

2.2 Malawi Second Integrated Household Survey

The MDICP does not include detailed measures of household consumption. For this reason, I use the Malawi Second Integrated Household Survey (IHS-2) to impute a better measure of consumption of households in the MDICP sample. The dataset, gathered by the Malawi National Statistics Office in 2004-2005, is part of the World Bank's Living Standards Measurement Study program. It includes comprehensive data on consumption and expenditures of households, as well as on local prices. Specifically, the dataset includes a measure of annual consumption expenditure aggregates in real value. The survey was fielded in 26 out of Malawi's 27 districts, including all three MDICP districts. After restricting the sample to households in the districts covered by MDICP and excluding households without a woman as a head of the household or as a wife of the head, the estimation sample includes

612 households. 220 of these households are from Balaka, 188 from Mchinji, and 204 from Rumphu.¹²

2.3 Mortality Statistics

I supplement the use of the MDICP data with statistics about survival after infection and child mortality from other studies that use data from samples with repeated HIV testing and frequent follow-ups. These data provide more precise information on times of infection and death than MDICP does. Hallett et al. (2008) estimate probabilities of survival after infection by fitting a Weibull distribution to survival data presented by Todd et al. (2007) from 5 studies in Eastern and Southern Africa before highly active antiretroviral therapy. The probability of survival to year t conditional on getting infected at year τ is estimated as

$$S(\tau, t) = \exp\left(-\left[\frac{t-\tau}{\psi_\tau}\right]^2\right),$$

with ψ_τ reducing with age of infection. (Parameter estimates and median survival time are presented in Table 5.) The mortality hazard increases with both duration and age of infection. The median survival time of an individual infected at ages 15 to 19 is 13.3 years, while it is 8.4 years for an individual who gets infected at ages 40 to 44.

Data from different longitudinal studies with repeated assessments of HIV status of adults show higher mortality rates of children born to infected women. In these studies, the HIV status of the children was generally not available. Controlling for different background characteristics, mortality rates of children born to HIV-infected mothers were estimated to be about three times higher than those for children born to uninfected mothers, with the effect lasting throughout childhood years (Newell et al., 2004). Crampin et al. (2003) report child mortality rates by status of mother at birth from a retrospective cohort study with more than 10 years of follow-up in Karonga district in Northern Malawi. The rates are reported in Table 6.

¹²Detailed information on the Malawi Second Integrated Household Survey can be obtained from the World Bank's website at <http://econ.worldbank.org>.

3 Model

3.1 General setup

I develop a dynamic discrete-choice life-cycle model of woman's fertility decisions in an environment of exposure to HIV infection. Women maximize subjective expected utility by making sequential binary fertility choices in a framework similar to that of Wolpin (1984) and Mira (2007). Specifically, a woman makes annual decisions of whether to become pregnant beginning at the age of her marriage and ending when she becomes infecund at a fixed age F (assumed to be 45). A woman gives birth in the period following the one in which she became pregnant. If never infected, a woman survives with certainty to age T (assumed to be 60). An infected woman might die prior to reaching the terminal model period. HIV infection of a mother at time of birth also increases mortality probabilities of children of ages zero to three. Given that HIV infection is asymptomatic for the majority of the infection duration, women cannot observe their HIV status.

Women are heterogeneous with respect to a group of characteristics that are treated as exogenous determinants of their choices. These characteristics include region of residence, completed schooling level, the size of the household's land plot, age of marriage, and whether a woman is married to a polygamous husband. Women are also of different discrete unobserved types, which are incorporated in the model to account for unobservable permanent factors which might affect preferences as well as perceived exposure to HIV.

3.2 Preferences

Each period, a woman receives a utility flow from household consumption (C), her number of children (N), pregnancies (p), and a time-varying preference shock (ϵ_p) which is iid across

time and women. The per-period utility function is given by

$$U(t) = \frac{C(t)^\phi}{\phi} [1 + \exp(\lambda_1 N(t))] + \lambda_{2,r,e,m,\mu} N(t) + \lambda_{3,r,e,m,\mu} N(t)^2 - (\lambda_{4,t} + \epsilon_p(t)) p(t) - \lambda_{5,t} p(t) p(t-1),$$

$$\epsilon_p(t) \sim iidN(0, \sigma_p^2).$$

The utility function exhibits constant relative risk aversion (CRRA) in consumption and includes an interaction term between household consumption and the number of children to reflect consumption being divided among more individuals as the family size increases. The utility is quadratic in the number of children. The parameters related to preference for children, λ_2 and λ_3 , are allowed to vary with region of residence (r), schooling level (e), polygamy (m), and unobserved type (μ). The utility function also incorporates a nonpecuniary cost (or benefit) associated with being pregnant. This cost includes a stochastic preference shock as well as a deterministic age-dependant element ($\lambda_{4,t}$). The cost of pregnancy increases if a woman was pregnant in the previous period ($\lambda_{5,t}$). The added cost of consecutive births allows for spacing of births.

The specification of the model implies perfect control over conception and contraception. I could instead introduce a cost of contraceptives, a probability of conception conditional on not wanting to become pregnant and a probability of not conceiving conditional on trying. Using data on births only, however, I cannot separately identify these elements. Instead, these elements will be absorbed into the cost of pregnancy (both deterministic and stochastic). For example, having the cost of pregnancy change with age captures physiological factors which vary the propensity to conceive during different stages of a woman's life-cycle.

3.3 Income and Consumption

It is assumed that households cannot borrow or save, which implies household consumption (C) equals the household's income (Y). The household's income is exogenous and stochastic. I specify a parsimonious household income function that appropriate for the context of

subsistence agriculture. I assume that the logarithm of income is distributed as

$$\ln(Y(t)) = \theta_1 + \theta_2 \text{Balaka} + \theta_3 \text{Mchinji} + \theta_3 \text{Land High} + \theta_4 N(t) + \theta_5 t + \theta_6 t^2 + \epsilon_y(t), \quad (1)$$

$$\epsilon_y(t) \sim iidN(0, \sigma_y^2).$$

Income depends on region of residence (Balaka, Mchinji), size of household's land plot (Land High), the age of the woman (t), the number of children (N), and a time-varying income shock (ϵ_y).¹³ Realization of the income shock occurs after the fertility decision is made. Therefore, pregnancy decisions are based on expected income.

3.4 Perceived Infection Hazard

A woman's perceived exposure to HIV infection is modeled as a hazard process. Let $h(t)$ be the probability a woman assigns to getting infected at period t , conditional on being HIV-negative until then. The perceived hazard rate for period t is given by

$$h(t) = \frac{1}{1 + \exp(-x(t)'\beta)}, \quad (2)$$

where $x(t)$ is a vector containing the woman's characteristics, the duration of her marriage, her age and age squared, and a constant. The parameters related to the constant, age and age squared are allowed to differ for unobserved types. The perceived unconditional probability of getting infected at period t , $P(t)$, is given by

$$P(t) = h(t) \prod_{k=1}^{t-1} (1 - h(k)). \quad (3)$$

3.5 Life Horizon Expectations

The woman is assumed to know the mortality processes associated with HIV infection (for both adults and children). Survival to each additional period provides her with information

¹³The Malawi Multiple Indicator Cluster Survey 2006 estimates that 27.5 percent of children aged 5 to 11 are involved in at least one hour of economic work or 28 hours of domestic work per week. A more natural specification of the income function might take into account the ages of the children; however, I do not keep track of the ages of children above age 3 in the state space to reduce computational burden.

about her HIV status. Specifically, she reduces the probability of having gotten infected in each past period. Given these probabilities and the mortality process, the woman assigns probabilities to survival to each future period. The woman also updates expectations about the survival of children of ages 0 to 3 in future periods. Conditional on the probabilities she assigns to having been infected at times of giving birth, she assigns probabilities to the survival of each of her children to future periods.

Women's beliefs are also updated by receiving an HIV test result. Respondents of the MDICP did not anticipate being offered HIV testing, and almost all of the respondents agreed to get tested. Because of these features of the data, I abstract from modeling the decision to get tested. Instead, I treat the HIV testing as an unanticipated revelation of HIV status. I first present the updating of expectations about life horizon without testing and then proceed to present the updating with testing.

3.5.1 Updating Without Testing

Let $I(\tau, t)$ be the probability a woman assigns to getting infected at period τ , conditional on being alive at t . The probability assigned to having gotten infected at a *past or present* period τ is given by

$$I(\tau, t) = \frac{\Pr(\text{got infected at } \tau \text{ \& alive at } t)}{\Pr(\text{alive at } t)} = \frac{P(\tau)S(\tau, t)}{1 - \sum_{k=1}^t P(k)(1 - S(k, t))}, \quad t \geq \tau, \quad (4)$$

where $S(\tau, t)$ is the probability a woman who gets infected at period τ survives to period t . The probability assigned at period t to being HIV-positive is given by the summation of the probabilities assigned to infection in all periods up to t :

$$B(t) = \sum_{k=1}^t I(k, t). \quad (5)$$

The probability assigned at period t to getting infected at a *future* period τ is given by

$$\begin{aligned} I(\tau, t) &= \Pr(\text{HIV-negative at } t) \Pr(\text{get infected at } \tau \mid \text{HIV-negative at } t) \\ &= (1 - B(t)) h(\tau) \prod_{k=t+1}^{\tau-1} (1 - h(k)), \quad t < \tau. \end{aligned} \quad (6)$$

The probability a woman assigns at time t to being alive at a future period τ is given by

$$\begin{aligned}
\pi(t, \tau) &= \sum_{k=1}^t \Pr(\text{got infected at } k) \Pr(\text{alive at } \tau \mid \text{got infected at } k \& \text{alive at } t) \\
&+ \sum_{k=t+1}^{\tau-1} \Pr(\text{will get infected at } k) \Pr(\text{alive at } \tau \mid \text{will get infected at } k) \\
&+ \Pr(\text{not get infected by } \tau) \\
&= \sum_{k=1}^t \left(I(k, t) \frac{S(k, \tau)}{S(k, t)} \right) + \sum_{k=t+1}^{\tau-1} (I(k, t) S(k, \tau)) + \sum_{k=1}^{\tau-1} (1 - I(k, t)).
\end{aligned} \tag{7}$$

Let $S_c^-(\tau, t)$ ($S_c^+(\tau, t)$) be the probability that a child born in period τ to an HIV negative (positive) woman survives to period t . The probability a woman assigns at period t for a child of age a to survive to future period τ is given by

$$\begin{aligned}
\pi_c^a(t, \tau) &= \Pr(\text{mother was HIV-pos at birth}) \Pr(\text{child alive at } \tau \mid \text{mother was HIV-pos at birth} \& \text{child alive at } t) \\
&+ \Pr(\text{mother was HIV-neg at birth}) \Pr(\text{child alive at } \tau \mid \text{mother was HIV-neg at birth} \& \text{child alive at } t) \\
&= \left(\sum_{k=1}^{t-a} I(k, t) \right) \frac{S_c^+(\tau-t-a)}{S_c^+(a)} + \left(1 - \sum_{k=1}^{t-a} I(k, t) \right) \frac{S_c^-(\tau-t-a)}{S_c^-(a)}.
\end{aligned} \tag{8}$$

3.5.2 With Testing

HIV testing provides a woman information about her infection status at the time of the test; however, it does not provide her with any new information about when she might have gotten infected. I assume women do not use the test result to update their perceived infection hazard process.

Given that some of the women who received positive test results assigned some likelihood to not being infected after the test, I assume that women assign a probability to the accuracy of the test result. Specifically, she assigns probability p_{test} to the test result being her

actual status and probability $1 - p_{test}$ to the test result being uninformative. Therefore, the probability a woman assigns to being HIV-positive when receiving a test result at period t_{test} is

$$\hat{B}(t_{test}) = (1 - p_{test}) B(t_{test}) + p_{test} \mathbf{1}\{\text{positive test result}\},$$

where $B(t_{test})$ is the probability assigned at period t_{test} to being HIV-positive without having been tested.

To compute the subjective life expectancies in periods after t_{test} , it is necessary to recover the post-test probabilities women assign to infection in each period, $\hat{P}(t)$ (the post-test update of $P(t)$, defined in Equation 3). These probabilities can be recovered using the fact that a test result does not provide the woman with any new information about when she might have gotten infected. Let $G(\tau, t_{test})$ to be the probability assigned to having gotten infected at τ , conditional on being HIV-positive at t_{test} . This probability does not change after learning the test result. The after-test probability assigned at t_{test} to having gotten infected in a past period τ can be written as

$$\begin{aligned} \hat{I}(\tau, t_{test}) &= \Pr(\text{got infected at } \tau \mid \text{HIV-positive at } t) \Pr(\text{HIV-positive at } t) \\ &= G(\tau, t_{test}) \hat{B}(t_{test}) = \frac{I(\tau, t)}{B(t)} \hat{B}(t_{test}), \quad \tau = 1, \dots, t_{test}. \end{aligned} \tag{9}$$

In addition, similar to Equation 4, the after-test probability assigned at t_{test} to having gotten infected in a past period τ is also given by

$$\hat{I}(\tau, t_{test}) = \frac{\hat{P}(\tau) S(\tau, t)}{1 - \sum_{k=1}^t \hat{P}(k) (1 - S(\tau, t))}, \quad \tau = 1, \dots, t_{test}. \tag{10}$$

By (9) and (10), I get

$$\frac{\hat{P}(\tau) S(\tau, t)}{1 - \sum_{k=1}^t \hat{P}(k) (1 - S(\tau, t))} = \frac{I(\tau, t)}{B(t)} \hat{B}(t_{test}), \quad \tau = 1, \dots, t_{test}.$$

I can solve for $\hat{P}(1), \dots, \hat{P}(t_{test})$ by equating solving a system of t_{test} equations with t_{test} unknowns. The probability assigned to infection at a future period t is given by

$$\hat{P}(t) = h(t) \prod_{k=1}^{t-1} (1 - \hat{P}(k)), \quad t > t_{test}.$$

Given the vector \hat{P} , the updating in future periods is similar to that described in equations (4) to (8).

3.6 Model Solution

The woman's problem can be formulated as a discrete-choice discrete-time stochastic dynamic program. Let $\Omega(t)$ be the state space at time t , consisting of all of the information relevant to decision-making that the woman has available at that time. Specifically, it contains the realized preference shock, the number of living children, ages of young children at risk of dying, whether she was pregnant in the previous period, and her testing history (period and result of each test). It also contains her fixed characteristics: her region of residence, schooling level, age of marriage, polygamy status of her husband, and the land owned by the household. Note that life horizon expectations are fully determined by $\Omega(t)$; therefore, there is no need to include the expectations in the state space.

Let $V(\Omega(t), t)$ be the value function, that is, the present discounted value of lifetime utility. Let $V^f(\Omega(t), t)$ be the alternative-specific value function, that is, the value if choice f is taken, with f indicating pregnancy status. The Bellman equation of the optimization problem is

$$V(\Omega(t), t) = \begin{cases} \max [V^0(\Omega(t), t), V^1(\Omega(t), t)], & t = 1, \dots, F-1 \\ EU^0(t, \Omega(t)) + \delta\pi(t, t+1)E(V(\Omega(t+1), t+1 | p(t) = 0, \Omega(t))), & t = F, \dots, T-1 \\ EU^0(T, \Omega(T)), & t = T \end{cases}$$

$$V^f(\Omega(t), t) = EU^f(t, \Omega(t)) + \delta\pi(t, t+1)E(V(\Omega(t+1), t+1 | p(t) = f, \Omega(t))), \quad f = 0, 1, \quad t = 1, \dots, F-1,$$

where $U^1(t, \Omega(t))$ represents a utility flow at state $\Omega(t)$ with pregnancy, and $U^0(t, \Omega(t))$ represents the utility flow without pregnancy. The expectation associated with flow utility is

taken over the present income shock. The expectation associated with the next-period value function is taken over future income and preference shocks as well as over child survival.

$V^1(\Omega(t), t)$ is strictly increasing in $\epsilon_p(t)$; however, $V^0(\Omega(t), t)$ is constant in $\epsilon_p(t)$ because the preference shock enters the utility flow only if the woman becomes pregnant. Let $\Omega^d(t)$ be the set of deterministic elements of the state space, that is, the set without $\epsilon_p(t)$. For any $\Omega^d(t)$ there is a unique critical value $\epsilon^*(\Omega^d(t), t)$ such that $V^1(\Omega(t), t) = V^0(\Omega(t), t)$. The solution to the woman's optimization problem is to become pregnant only if the preference shock is bigger than the corresponding critical value.

4 Estimation

The main estimation sample contains data on 1006 women from the MDICP dataset. The information on the i th woman consists of

- Up to 4 years of pregnancy choices, $p_i(t)$, $t = \underline{t}_{pi}, \dots, \overline{t}_{pi}$
- Up to 2 reports of subjective assessments of HIV status, $b_i(t)$, $t = \underline{t}_{bi}, \dots, \overline{t}_{bi}$
- Up to 3 HIV test results, from the 2004, 2006 and 2008 rounds
- Vector of characteristics: region of residence, schooling level, year of birth, age of marriage, polygamy status of husband, household's land

I also use an auxiliary sample containing data on 612 households from the IHS-2 dataset.

The information on the j th household consists of:

- Household annual aggregate consumption expenditure (assumed to equal household's income), y_j
- Vector of household characteristics: region of residence, age of woman (head or wife of the head of the household), number of children, household's land

The first step of the econometric implementation involves estimating the parameters of the income function (Equation 1) by ordinary least squares regression using the auxiliary sample. The rest of the model parameters are estimated using maximum likelihood, taking the parameters of the income and survival functions as given. The likelihood function contains the following elements:

1. Belief reports probabilities
2. Fertility outcomes probabilities
3. Actual HIV status (HIV testing and survival)
4. Unobserved type distribution

I proceed by describing the contribution of each of these elements to the likelihood.

4.1 Beliefs

The perceived infection hazard is estimated using the data on subjective assessments of HIV-status. Let $B_i(t)$ be woman i 's perceived probability of being HIV-positive at period t and $b_i(t)$ be the number of beans she allocates to being HIV-infected in the expectations elicitation exercise. I make the following two assumptions regarding the relationship between a woman's belief and the recorded number of beans. First, I assume that beliefs are reported with some noise. Specifically, the reporting error, $\epsilon_{bi}(t)$, is assumed to be iid across time and women:

$$\epsilon_{bi}(t) \sim iidN(0, \sigma_b).$$

Second, I assume that each discrete "bean category" corresponds to a probability interval.¹⁴ A respondent reports the number of beans corresponding to the interval in which her belief added to the reporting error falls. The reports are assumed to be made according to the

¹⁴As in Delavande and Kohler (2009)

following rule:¹⁵

$$b_i(t) = 0, 1 \quad \text{if } B_i(t) + \epsilon_{bi}(t) \leq 0.15,$$

$$b_i(t) = 2 \quad \text{if } 0.15 < B_i(t) + \epsilon_{bi}(t) \leq 0.25,$$

⋮

$$b_i(t) = 10 \quad \text{if } 0.95 < B_i(t) + \epsilon_{bi}(t).$$

Let $\underline{\Omega}_i^d$ be the set of initial conditions for woman i . It contains the deterministic (and observable) elements of the state space. That is, it contains her permanent characteristics that enter the perceived hazard function. I define it also to contain the woman's testing variables (when she got tested and the test results). Although the woman does not forecast getting testing I treat this information as part of the initial conditions for the econometric implementation. The woman's sequence of life-cycle beliefs are determined given these initial conditions and her unobserved type and can be written as $B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)$. The probability of observing $b_i(t)$, conditional on the set of woman i 's initial conditions and her unobserved type is given by

$$\Pr(b_i(t) = 0, 1 \mid \underline{\Omega}_i^d, \text{type}_i = j) = \Phi\left(\frac{0.15 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right),$$

$$\Pr(b_i(t) = 2 \mid \underline{\Omega}_i^d, \text{type}_i = j) = \Phi\left(\frac{0.25 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right) - \Phi\left(\frac{0.15 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right),$$

⋮

$$\Pr(b_i(t) = 10 \mid \underline{\Omega}_i^d, \text{type}_i = j) = 1 - \Phi\left(\frac{0.95 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right).$$

As shown in Figure 1, the percentage of women reporting each bean category generally decreases with the number of beans. The percentage of women who report five beans, 8.78 percent, is higher than the two categories below (3.41 percent report four beans and 6.52

¹⁵Figure 3 shows the distributions of reported beliefs in the 2006 and 2008 rounds. There is a big drop in the percentage of women choosing the 0-bean category and the number of women who choose the 1-bean category doubles. My model would not be able to generate this shift. For my empirical analysis, I treat the 0 and 1-bean categories as a single category.

percent report three) and is also significantly higher than the next category (1.59 percent report six). My model is not likely to capture this pattern. I therefore assume that the probability interval corresponding to the five bean category is larger, implying that some of the women who would allocate 4 or 6 beans to the likelihood of being HIV-positive under the rule described above, report instead 5 beans. The probability of observing 5 beans is assumed to be

$$\Pr(b(t) = 5 \mid \underline{\Omega}_i^d, \mu_i = j) = \Phi\left(\frac{0.35 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right) - \Phi\left(\frac{0.65 - B(t \mid \underline{\Omega}_i^d, \text{type}_i = j)}{\sigma_b}\right).$$

4.2 Fertility

The parameters of the utility function are estimated using the data on pregnancy outcomes. The solution to a woman's optimization problem is to become pregnant at period t if the preference shock, $\epsilon_{pi}(t)$, is bigger than the critical value $\epsilon^*(\Omega_i^d(t), t, \text{type}_i)$. Conditional on the woman's type and the observable elements of the woman's state space at time t , the probability of the woman's choice to become pregnant is given by

$$\Pr(p_i(t) = 1 \mid \Omega_i^d(t), \text{type}_i = j) = 1 - \Phi\left(\frac{\epsilon^*(\Omega_i^d(t), t, \text{type}_i)}{\sigma_p}\right),$$

$$\Pr(p_i(t) = 0 \mid \Omega_i^d(t), \text{type}_i = j) = \Phi\left(\frac{\epsilon^*(\Omega_i^d(t), t, \text{type}_i)}{\sigma_p}\right).$$

Conditional on the set of initial conditions and the unobserved type, the probability of observing a sequence of belief reports is independent from the probability of observing a sequence of fertility outcomes. This is because the set of initial conditions and type map deterministically into the sequence of life-cycle beliefs.

4.3 Actual Infection Process

Given the assumption of no symptoms, the state space does not contain actual HIV status beyond any HIV test results that women have received. It is therefore not necessary to recover the actual hazard process for estimation of the decision-model parameters. For each

possible state, the estimated model generates a probability of becoming pregnant conditional on being alive at that state. I do need an estimate of the actual hazard process for my counterfactual analysis, however. To study the effect of different policy interventions on outcomes of life-cycle fertility, child mortality, and orphanhood, it is necessary to include the infection and mortality processes in the simulations.

Let $\tilde{h}(t)$ be the actual HIV infection hazard rate. The functional form is assumed to be similar to that of the perceived infection hazard described in equation (2). The probability of getting infected at t , conditional on being HIV-negative until then, is given by

$$\tilde{h}(t) = \frac{1}{1 + \exp(-x(t)' \tilde{\beta})}.$$

The unconditional probability of getting infected at period t is given by

$$\tilde{P}(t) = \tilde{h}(t) \prod_{k=1}^{t-1} (1 - \tilde{h}(k)).$$

Information about the hazard process is contained in the HIV test results, the age in which the tests were taken, and the ages in which a woman is last observed (regardless of testing histories). Let \bar{t}_i represent the age in which a woman is last observed (and is therefore known to be alive at that age). Let t_i^- be the *oldest* age in which a woman is observed to get a negative test result. Let t_i^+ be the *youngest* age a woman is observed to receive a positive test result. Let $H_i = (t_i^-, t_i^+, \bar{t}_i)$ be the vector of observed ‘‘HIV history’’ of woman i , with t_i^- (t_i^+) equaling zero if a woman is never tested negative (positive). Women’s ‘‘HIV histories’’ belong to one of the following 4 categories:

1. Never tested:

The probability of observing a woman who was never tested alive at \bar{t}_i , conditional on her initial conditions and unobserved type, is given by

$$\Pr(H_i = (0, 0, \bar{t}_i) \mid \Omega_i^d, \text{type}_i = j) = \Pr(\text{alive at } \bar{t}_i) = 1 - \sum_{k=1}^{\bar{t}_i} \tilde{P}(k \mid \Omega_i^d, \text{type}_i = j) (1 - S(k, \bar{t}_i)).$$

2. Tested only negative:

The probability of observing a woman who was only tested negative and is alive at \bar{t}_i ,

conditional on her initial conditions and unobserved type, is given by

$$\begin{aligned}
& \Pr(H_i = (t^-, 0, \bar{t}) \mid \underline{\Omega}_i^d, \text{type}_i = j) = \Pr(\text{negative at } t^-, \text{alive at } \bar{t}) \\
& = \prod_{k=1}^{t^-} (1 - \tilde{h}(k)) \left[1 - \sum_{k=t^-+1}^{\bar{t}} \tilde{h}(k \mid \underline{\Omega}_i^d, \text{type}_i = j) \prod_{j=t^-+1}^k (1 - \tilde{h}(j \mid \underline{\Omega}_i^d, \text{type}_i = j)) (1 - S(k, \bar{t})) \right] \\
& = \prod_{k=1}^{t^-} (1 - \tilde{h}(k)) - \sum_{k=t^-+1}^{\bar{t}} \tilde{P}(k \mid \underline{\Omega}_i^d, \text{type}_i = j) (1 - S(k, \bar{t})).
\end{aligned}$$

3. Tested only positive:

The probability of observing a woman who if infected by t^+ and alive at \bar{t} , conditional on her initial conditions and unobserved type, is given by

$$\Pr(H_i = (0, t^+, \bar{t}) \mid \underline{\Omega}_i^d, \text{type}_i = j) = \Pr(\text{got infected before } t^+, \text{alive at } \bar{t}) = \sum_{k=1}^{t^+} \tilde{P}(k \mid \underline{\Omega}_i^d, \text{type}_i = j) S(k, \bar{t}).$$

4. Seroconverter:

The probability of observing a woman who is known to be HIV-negative until t^- , positive by t^+ , and alive at \bar{t} ($t^- < t^+ \leq \bar{t}$), conditional on her initial conditions and unobserved type, is given by

$$\begin{aligned}
& \Pr(H_i = (t^-, t^+, \bar{t}) \mid \underline{\Omega}_i^d, \text{type}_i = j) = \Pr(\text{got infected between } t^- \text{ and } t^+, \text{alive at } \bar{t}) \\
& = \sum_{k=t^-}^{t^+} \tilde{P}(k \mid \underline{\Omega}_i^d, \text{type}_i = j) S(k, \bar{t}).
\end{aligned}$$

4.4 Type Distribution

A multinomial logit specification is used for the type probabilities. The probability that woman i is of type j is given by

$$\Pr(\text{type}_i = j \mid \underline{\Omega}_i^d) = \frac{\exp(w'_{\underline{t}i} \gamma_j)}{1 + \sum_{k=1}^J \exp(w'_{\underline{t}i} \gamma_k)}, \quad j = 1, \dots, J,$$

$$\Pr(\text{type}_i = 0 \mid \underline{\Omega}_i^d) = \frac{1}{1 + \sum_{k=1}^J \exp(w'_{\underline{t}i} \gamma_k)},$$

where $w_{\underline{t}i}$ is a vector of woman's initial conditions, including region of residence, schooling level, age of marriage, whether she is married to a polygamous husband, year of birth, and the number of children she had in the first period observed interacted with the age she was

when first observed. The last term is included to take into account the fact that not all the women are observed from the time of their marriage. The state in which they are first observed depends on past decisions and therefore on their unobserved type.

4.5 Likelihood Function

I estimate the model using maximum likelihood. The contribution of woman i to the sample likelihood is given by

$$\mathcal{L}_i = \sum_{j=0}^J \left(\prod_{t=\underline{t}_p}^{\bar{t}_p} \Pr(p_i(t) | \Omega_i^d(t), \text{type}_i = j) \right) \left(\prod_{t=\underline{t}_b}^{\bar{t}_b} \Pr(b_i(t) | \underline{\Omega}_i^d, \text{type}_i = j) \right) \left(\Pr(H_i | \underline{\Omega}_i^d, \text{type}_i = j) \right) \Pr(\text{type} = j | \underline{\Omega}_i^d).$$

5 Estimation Results

5.1 Parameter Estimates

Tables 7 to 11 report the parameter estimates and their standard errors. As can be seen in Table 7, the marginal utility of additional children is positive and varies with region of residence, schooling level, and polygamy status. The childbearing costs, which are assumed to depend only on women's age, are estimated to be positive for all ages. Childbearing cost is the lowest at ages 20 to 24, when it is lower by 16 percent than the cost in ages 15 to 19 and lower by 45 percent than the cost in ages 40 to 44.

The model is fit with 4 unobserved types.¹⁶ Recall that the types differ with respect to their perceived exposure to HIV, their preferences for children, and the assigned probability to the accuracy of HIV test results. The four types have different beliefs and behaviors. Table 12 shows the average probability assigned to being HIV-positive at age 30. Women of type 0, who represent 30.3 percent of the sample, assign the lowest average probability to

¹⁶There were significant improvements in model fit beyond three types. I did not attempt to fit the model with more types because of computational burden.

being infected. Women of type 3, who represent only 1.3 percent of the sample, assign an average probability of 0.89 to being infected when they are 30 years old. Type 1, representing 48 percent of the sample, values children the most and has the highest annual pregnancy probabilities for all age groups.

5.2 Model Fit

To assess model fit, I compare the model's prediction of the distribution of reported beliefs and pregnancy probabilities to the actual beliefs and and pregnancies observed in the data. I simulate each woman in the data 100 times. The simulation starts from the age a woman is first observed and takes as given the observable elements of the state space. For each simulation, I draw an unobservable type from the type distribution, preference shocks and realizations of child mortality.

Figures 4 and 5 and Table 13 compare the fertility and belief reporting outcomes predicted by the model to the actual outcomes observed in the data. Figure 4 shows that the model is able to generate the shape of the reported beliefs distribution. It under-predicts the proportion of women in the 0-bean category by about 4 percentage points and over-predicts the proportions in the 2 and 3-bean categories. As described in section 4, I assume that some of the respondents round and report 5 beans instead of 4 and 6. 13.8 percent of belief reports in the data are of the 4 to 6-bean categories; the model predicts 12.9 percent.

Figure 5 depicts the actual and predicted annual pregnancy probabilities for different age groups. The model captures the decline in pregnancy probabilities with age. This pattern is generated by the diminishing marginal utility of additional children and childbearing costs, which increase with age. Table 13 shows the actual and predicted annual pregnancy probabilities by region, schooling level, and polygamy status of the husband.

6 Counterfactual Analysis

Having estimated the structural model parameters, I use the model to perform counterfactual analysis to answer the following questions:

1. How would fertility outcomes differ if there were no HIV?
2. How did the HIV tests offered by the MDICP team affect fertility outcomes?
3. What would be the effect of treatments preventing mother-to-child transmission of HIV?

I answer these questions by simulating the women's life-cycle fertility decisions in the different environments. Each woman is simulated 100 times. In these simulations, I incorporate infection and mortality. That is, women are exposed to HIV infection according to the estimated actual infection hazard process, and, once infected, they are exposed to the mortality process.

To answer the first question, I simulate fertility in an environment with no HIV (and no beliefs about HIV). The results indicate that HIV has a negative effect on fertility. The average number of births during the life-cycle in the no-HIV environment is 7.26 in comparison to 7.02 in the baseline environment.¹⁷ The average difference of 0.24 births is due both to the shorter average life-span of women infected with HIV as well as to a behavioral effect. The magnitude of the effect is higher for groups with higher perceived exposure to HIV infection. In Balaka, for example, which is the region with the highest HIV prevalence and highest probabilities assigned to being infected, women in the no-HIV environment would have on average 0.28 more births during their life-cycle. Average annual pregnancy rates are higher by 0.8 percentage points without HIV. The percentage of newborns who do not survive to age four is 15.8, compared to 17.5 with HIV.

¹⁷The life-cycle number of births is higher than the national total fertility rate estimated at 6 births per woman because fertility rates are higher in rural areas of Malawi.

To study the impact of HIV testing on fertility outcomes, I simulate life-cycle fertility in a counterfactual scenario in which the respondents were not offered HIV tests by the MDICP team. The impact of testing on outcomes depends both on how women update their beliefs after learning their test result and on the behavioral response to the new beliefs. The updating depends on the discrepancy between the beliefs and actual HIV status and on the accuracy women assign to test results. I find that women who received positive HIV test results have on average 0.28 fewer births during their life-cycle. The average number of births falls by 0.34 births for women who tested positive by age 25. I find no effect on fertility for women who receive a negative test result. I also find no overall effect on child mortality, because the percentage of women who test positive is very low.

To assess the potential effect of provision of antiretroviral treatment, which prevents mother to child transmission, I simulate fertility outcomes in an environment with no transmission.¹⁸ That is, the child mortality probabilities of children born to HIV-infected women are equated to those of children born to HIV-negative women. I find a negligible change in the average number of births per woman. However, because of the reduced rate of child mortality, the average number of child deaths per woman falls from 1.22 to 1.11. The average number of children that survive beyond age four increases from 5.8 to 5.94.

7 Conclusion

In this paper, I specified and structurally estimated a dynamic model of fertility in an environment with high HIV prevalence. My analysis takes into account uncertainty about HIV status and the discrepancies between perceived and actual exposure to HIV infection. Women's perceptions about their exposure to HIV infection and about their own HIV status affect their beliefs about their own and their children's life expectancy. Beliefs about life expectancies, in turn, affect fertility choices by changing the profile of expected lifetime

¹⁸In high income countries, the rate of mother-to-child transmission has been reduced to less than one percent (unaids.org).

utilities associated with each choice. I estimate the model parameters by maximum likelihood with longitudinal data from the Malawi Diffusion and Ideation Change Project, which contain measures of subjective beliefs about own HIV status. The model fits well the fertility patterns in the data, as well as the distribution of reported beliefs about own HIV status.

Model simulations are informative about how HIV affects fertility and about the impact of policies aimed at reducing HIV, such as HIV testing programs and provision of antiretroviral therapy, on fertility and child mortality. Results show that the presence of HIV reduces fertility, both for women who are infected and who are not infected. The presence of HIV reduces the average number of life-cycle births by 0.24. I also find that HIV testing is effective at reducing fertility of infected women, leading to a reduction in child mortality and orphanhood; however, many women do not believe the test results, which mitigates the effectiveness of testing. I also find that prevention of mother-to-child transmission through dissemination of antiretroviral drugs would have a negligible effect on the average number of life-cycle births, although it can reduce the incidence of child mortality by about nine percent.

References

- ANGLEWICZ, P., "Migration, risk perception, and HIV infection in Malawi," *Dissertations available from ProQuest* (2007).
- ANGLEWICZ, P. AND H. KOHLER, "Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi," *Demographic research* 20 (2009), 65.
- CRAMPIN, A., S. FLOYD, J. GLYNN, N. MADISE, A. NYONDO, M. KHONDOWE, C. NJOKA, H. KANYONGOLOKA, B. NGWIRA, B. ZABA ET AL., "The long-term impact of HIV and orphanhood on the mortality and physical well-being of children in rural Malawi," *Aids* 17 (2003), 389.
- DELAVANDE, A., X. GINÉ AND D. MCKENZIE, "Measuring subjective expectations in developing countries: a critical review and new evidence," *Journal of Development Economics* (2010).
- DELAVANDE, A. AND H. KOHLER, "Subjective Expectations in the Context of HIV/AIDS in Malawi," *Demographic research* 20 (2009), 817.
- FINK, G. AND S. LINNEMAYR, "HIV Does Matter for Fertility: Human Capital, Mortality and Family Size," *Harvard Public School of Health, mimeo* (2009).
- FORTSON, J., "HIV/AIDS and fertility," *American Economic Journal: Applied Economics* 1 (2009), 170–194.
- GRAY, R., M. WAWER, R. BROOKMEYER, N. SEWANKAMBO, D. SERWADDA, F. WABWIRE-MANGEN, T. LUTALO, X. LI, T. VANCOTT AND T. QUINN, "Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda," *The Lancet* 357 (2001), 1149–1153.

- HALLETT, T., B. ZABA, J. TODD, B. LOPMAN, W. MWITA, S. BIRARO, S. GREGSON, J. BOERMA AND P. GHYS, “Estimating incidence from prevalence in generalised HIV epidemics: methods and validation,” *PLoS Med* 5 (2008), e80.
- HECKMAN, J. AND B. SINGER, “A method for minimizing the impact of distributional assumptions in econometric models for duration data,” *Econometrica: Journal of the Econometric Society* 52 (1984), 271–320.
- HECKMAN, J. AND R. WILLIS, “Estimation of a stochastic model of reproduction: An econometric approach,” *in: N.E. Terleckyj, ed., Household production and consumption (Columbia University Press, New York)* (1976).
- JOSEPH HOTZ, V., J. KLERMAN AND R. WILLIS, “The economics of fertility in developed countries,” *Handbook of population and family economics* 1 (1997), 275–347.
- JUHN, C., S. KALEMLI-OZCAN AND B. TURAN, “HIV and fertility in Africa: First evidence from population based surveys,” *NBER working paper* (2008).
- KALEMLI-OZCAN, S., “AIDS, " Reversal" of the Demographic Transition and Economic Development: Evidence from Africa,” *NBER Working paper* (2006).
- MIRA, P., “UNCERTAIN INFANT MORTALITY, LEARNING, AND LIFE-CYCLE FERTILITY,” *International Economic Review* 48 (2007), 809–846.
- MORGAN, D., C. MAHE, B. MAYANJA, J. OKONGO, R. LUBEGA AND J. WHITWORTH, “HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?,” *Aids* 16 (2002), 597.
- NEWELL, M., H. BRAHMBHATT AND P. GHYS, “Child mortality and HIV infection in Africa: a review,” *Aids* 18 (2004), S27.
- OBARE, F., P. FLEMING, P. ANGLEWICZ, R. THORNTON, F. MARTINSON, A. KAPATUKA, M. POULIN, S. WATKINS AND H. KOHLER, “Acceptance of repeat population-

- based voluntary counselling and testing for HIV in rural Malawi,” *Sexually transmitted infections* 85 (2009), 139.
- POWERS, K., C. POOLE, A. PETTIFOR AND M. COHEN, “Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis,” *The Lancet infectious diseases* 8 (2008), 553–563.
- SAH, R., “The effects of child mortality changes on fertility choice and parental welfare,” *Journal of Political Economy* 99 (1991), 582–606.
- SCHULTZ, T., “Demand for children in low income countries,” *Handbook of population and family economics* 1 (1997), 349–430.
- THORNTON, R. L., “The Demand for, and Impact of, Learning HIV Status,” *American Economic Review* 98 (December 2008), 1829–63.
- TODD, J., J. GLYNN, M. MARSTON, T. LUTALO, S. BIRARO, W. MWITA, V. SURIYANON, R. RANGSIN, K. NELSON, P. SONNENBERG ET AL., “Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy,” *Aids* 21 (2007), S55.
- WOLPIN, K., “Determinants and Consequences of the Mortality and Health of Infants and Children,” *Handbook of population and family economics* 1 (1997), 483–557.
- WOLPIN, K. I., “An Estimable Dynamic Stochastic Model of Fertility and Child Mortality,” *Journal of Political Economy* 92 (October 1984), 852–74.
- YOUNG, A., “The Gift of the Dying: The Tragedy of Aids and the Welfare of Future African Generations,” *The Quarterly Journal of Economics* 120 (May 2005), 423–466.
- , “In sorrow to bring forth children: fertility amidst the plague of HIV,” *Journal of Economic Growth* 12 (December 2007), 283–327.

Table 1: Descriptive Statistics

Variable		All	Balaka (South)	Mchinji (Center)	Rumphi (North)
Region:	Balaka	0.29	-	-	-
	Mchinji	0.35	-	-	-
	Rumphi	0.36	-	-	-
Education:	No school	0.14	0.31	0.15	0.01
	Primary	0.73	0.66	0.8	0.73
	Secondary	0.12	0.03	0.05	0.26
Land>1 hectare		0.44	0.19	0.53	0.55
Polygamy		0.27	0.2	0.27	0.33
Age of Marriage		17.54 (2.25)	17.02 (2.28)	17.56 (2.03)	17.93 (2.35)
Age 2004		26.63 (8.09)	25.99 (8.44)	25.75 (7.21)	28.01 (8.45)
Number of observations		1006	355	288	363

Table 2: Test Take-up and Percentage Tested Positive by Year of Test

	2004		2006		2008	
	%	N	%	N	%	N
Took Test	87.9%	620 ^a	93.28%	759 ^a	94.6%	741 ^a
Tested positive	2.94%	545 ^b	3.11%	708 ^b	3.71%	701 ^b

^a Number of women who were offered to take a HIV test for the given year

^b Number of women tested for the given year

Table 3: Reported Belief about Own HIV Infection, Measured in Beans: 2006 and 2008 pooled

		Mean	sd	N
All		1.52	2.13	1640
Region	Balaka	1.78	2.16	462
	Mchinji	1.69	2.24	565
	Rumphu	1.16	1.96	613
Schooling	No school	1.77	2.2	230
	Primary	1.55	2.18	1214
	Secondary	1.06	1.67	196
Polygamy	Mono	1.35	1.35	1188
	Poly	1.95	2.44	452
Age	≤26	1.41	2.02	658
	>26	1.59	2.21	982

Figure 1: Belief Distribution, 2006 and 2008 pooled

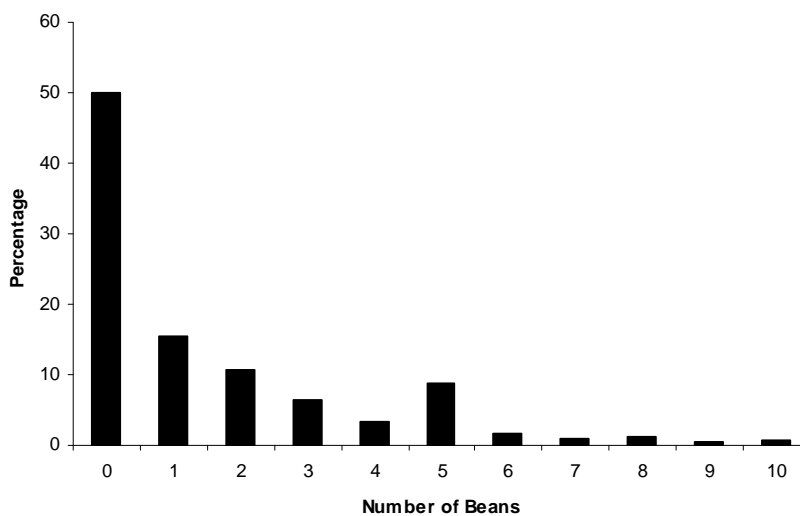


Figure 2: Beliefs by Test Result, 2006 and 2008 pooled

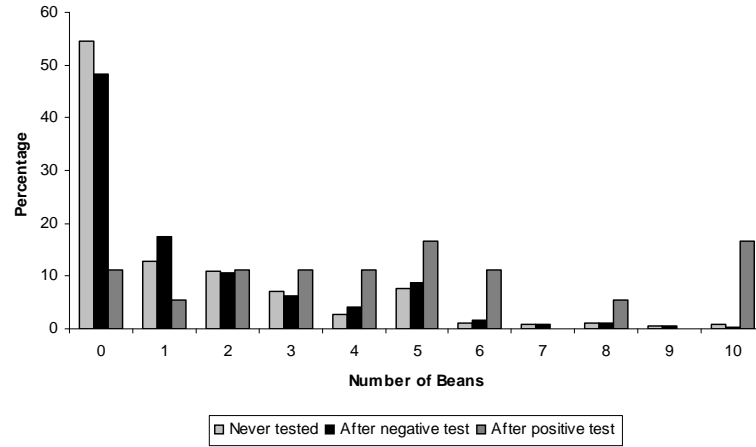


Table 4: Annual Pregnancy Probabilities

		Age Group						
		16-20	21-25	26-30	31-35	36-40	41-45	
All	Prob.	0.405	0.377	0.307	0.273	0.211	0.06	
	N	412	778	698	539	370	351	
Region	Balaka	Prob.	0.414	0.398	0.323	0.31	0.281	0.086
		N	162	226	155	142	121	81
	Mchinji	Prob.	0.369	0.357	0.32	0.267	0.177	0.082
		N	130	319	291	172	113	85
	Rumphu	Prob.	0.433	0.382	0.282	0.253	0.176	0.038
		N	120	233	252	225	136	185
Schooling	None	Prob.	0.375	0.492	0.27	0.263	0.31	0.085
		N	24	61	89	95	84	82
	Primary	Prob.	0.395	0.366	0.315	0.281	0.183	0.053
		N	332	596	520	381	262	247
	Secondary	Prob.	0.482	0.372	0.292	0.238	0.167	0.045
		N	56	121	89	63	24	22
Polygamy	Mono	Prob.	0.412	0.388	0.309	0.285	0.23	0.077
		N	354	605	511	396	235	196
	Poly	Prob.	0.362	0.335	0.299	0.238	0.178	0.039
		N	58	173	187	143	135	155

Table 5: Parameter Estimates of the Survival-after-Infection Function

	Age Group						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
ψ_τ , Weibull scale parameter	16.0	15.4	14.1	12.1	11.0	10.1	7.9
Median survival years after infection	13.3	12.8	11.7	10.0	9.1	8.4	6.6

Estimated by Hallett et al. (2008)

Table 6: Child Mortality Rates per 1000 Person-Year, by Mother HIV status

	Child's Age			
	0	1	2	3-4
Mother HIV-negative at birth	115	26	18	8
Mother HIV-positive at birth	331	128	87	41

Source: Crampin et al. (2003)

Figure 3: Belief Distribution, by Year

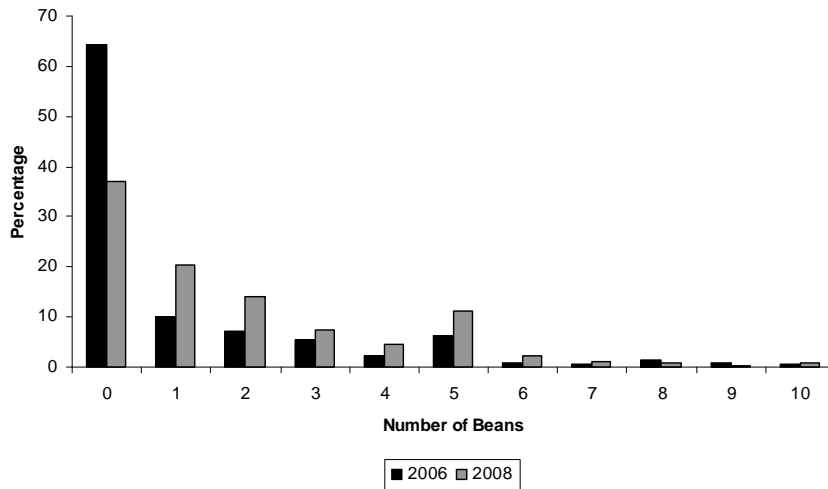


Table 7: Maximum-Likelihood Parameter Estimates: Preferences

Parameter	Description	Estimate	SE
ϕ	CRRA parameter	0.685	0.382
λ_1	Child-consumption interaction	-0.298	0.123
λ_2 - type 0	$N(t)$, Number of children, type 0	1010	220.246
λ_2 - type 1	$N(t)$, Number of children, type 1	1330	299.487
λ_2 - type 2	$N(t)$, Number of children, type 2	620	256.754
λ_2 - type 3	$N(t)$, Number of children, type 3	400	58.818
λ_2 - balaka	$N(t)$, Balaka shifter	163	63.157
λ_2 - rumphi	$N(t)$, Rumphi shifter	480	155.268
λ_2 - primary	$N(t)$, Primary shifter	-34.5	14.192
λ_2 - secondary	$N(t)$, Secondary shifter	550	189.01
λ_2 - poly	$N(t)$, Polygamy shifter	-138	50.903
λ_3	Number of children squared, $N(t)^2$	-2	1.603
λ_3 - balaka	$N(t)^2$, Balaka shifter	10	4.007
λ_3 - rumphi	$N(t)^2$, Rumphi shifter	-70	17.648
λ_3 - primary	$N(t)^2$, Primary shifter	10	4.17
λ_3 - secondary	$N(t)^2$, Secondary shifter	-115	29.511
λ_3 - poly	$N(t)^2$, Polygamy shifter	9.88	3.921
λ_4	Pregnancy, $p(t)$	7380	1260.704
λ_4 - age 20-24	$p(t)$, age 20-24 shifter	-1200	452.497
λ_4 - age 25-29	$p(t)$, age 25-29 shifter	-290	117.357
λ_4 - age 30-34	$p(t)$, age 30-34 shifter	-265	108.07
λ_4 - age 35-39	$p(t)$, age 35-39 shifter	-120	59.593
λ_4 - age 40-44	$p(t)$, age 40-44 shifter	3780	1094.8
λ_5	Consecutive pregnancy, $p(t)p(t-1)$	8030	1276.9
$\lambda_{5,t}$	Consecutive-age interaction, $p(t)p(t-1)t$	-203.5	47.39
σ_p	Standard deviation of preference shock	6600	1204.02
δ	Discount factor	0.892	0.031

Table 8: Ordinary Least Squares Parameter Estimates: Income Function

Parameter	Description	Estimate	SE
θ_1	Constant	10.619	0.135
θ_2	Balaka	-0.0717	0.046
θ_3	Mchinji	-0.092	0.049
θ_4	Land High	0.4295	0.039
θ_5	Number of children, $N(t)$	0.0499	0.011
θ_6	Period, t	0.0118	0.007
θ_7	Period squared, t^2	-0.0002	0.00008
σ_y	Standard deviation of income shocks	0.536407	0.065

Table 9: Maximum-Likelihood Parameter Estimates: Infection Hazard Function, Perceived and Actual

Parameter	Perceived		Actual	
	Estimate	SE	Estimate	SE
Constant, type 0	-14.1	3.456	-9.991	2.455
Constant, type 1	-4.55	0.333	-14.87	1.447
Constant, type 2	2.06	0.774	-20	7.804
Constant, type 3	-2.92	0.532	-12.333	2.504
Period, type 0	-0.2	0.078	0.311	0.258
Period, type 1	-0.0602	0.023	2.125	0.147
Period, type 2	0.0625	0.025	1.358	0.242
Period, type 3	-0.127	0.041	0.422	0.205
Period squared, type 0	0.0174	0.002	-0.005	0.008
Period squared, type 1	0.0033	0.001	-0.114	0.016
Period squared, type 2	-0.0093	0.003	-0.028	0.024
Period squared, type 3	0.0008	0.0003	-0.008	0.005
Duration of marriage	-0.0125	0.0004	0.007	6.89
Primary	0.517	0.179	0.627	0.709
Secondary	-0.562	0.362	0.966	0.718
Land > 1 hectare	-0.0007	0.0003	-0.331	0.37
Polygamy	0.953	0.208	0.7322	0.448
Balaka	-0.59	0.202	0.006	0.548
Rumphhi	-1.181	0.433	-0.151	0.488

Table 10: Maximum-Likelihood Parameter Estimates: Other Parameters Related to Beliefs

Parameter	Description	Estimate	SE
p_{test} , type 0	Test result accuracy, type 0	0.73	0.152
p_{test} , type 1	Test result accuracy, type 1	0.1	0.002
p_{test} , type 2	Test result accuracy, type 2	1	0.248
p_{test} , type 3	Test result accuracy, type 3	0.144	0.045
σ_b	Standard deviation of reporting error	0.25	0.009

Table 11: Maximum Likelihood Parameter Estimates: Type Distribution Parameters

Parameter	Type 1		Type 2		Type 3	
	Estimate	SE	Estimate	SE	Estimate	SE
Constant	-3.07	1.037	-5.12	1.458	-1.9	0.684
Balaka	3.62	1.017	6.37	1.416	5.719	1.153
Rumphhi	0.260	0.106	3.14	1.116	-3.22	0.944
Primary	-3.83	1.082	-6.29	1.369	3.67	0.964
Secondary	0.000	0.495	-11.3	4.888	0.242	0.101
Polygamy	-0.245	0.099	1.552	1.94	3.62	0.896
Age of marriage	0.074	0.046	-0.03	0.012	-0.104	0.043
Year of birth	0.368	0.073	0.191	0.066	-0.626	0.043
$N(\underline{t}) \times \underline{t}$	0.013	0.04	0.0003	0.0001	0.0063	0.002

Table 12: Predicted Selected Characteristics by Unobserved Type

	Type 0	Type 1	Type 2	Type 3
Percent of sample	30.3	48.4	1.3	20
Balaka	2.2	66.1	3.4	28.1
Mchinji	42.7	41.3	0.02	15.9
Rumphu	41	40.9	0.06	17.5
Prob. assigned to being HIV-positive when 30 yr old	0.008	0.11	0.89	0.29
Annual pregnancy probabilities by age groups:				
15-19	0.334	0.416	0.201	-
20-24	0.352	0.405	0.207	0.273
25-29	0.298	0.346	0.191	0.233
30-34	0.272	0.314	0.219	0.218
35-39	0.248	0.302	0.179	0.174
40-44	0.098	0.107	0.039	0.055

Figure 4: Model Fit: Actual and Predicted Reported Beliefs

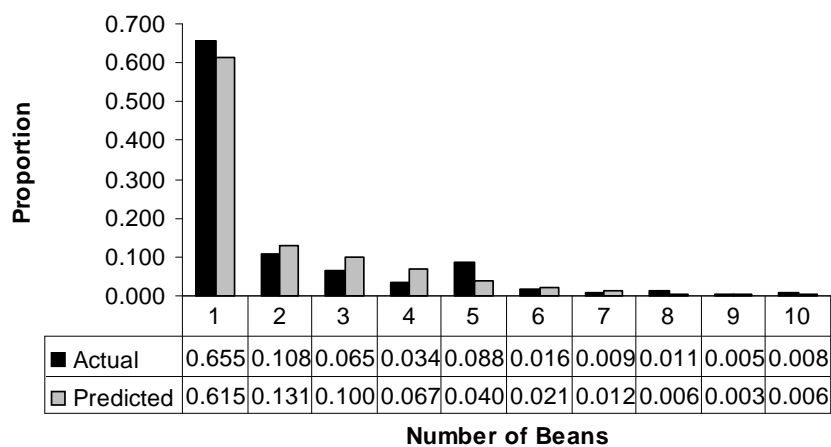


Figure 5: Model Fit: Pregnancy Probabilities, by Age Groups

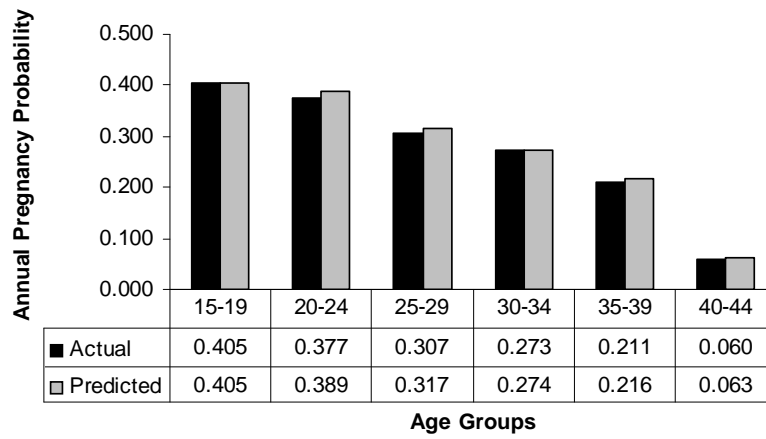


Table 13: Model Fit: Actual and Predicted Pregnancy Probabilities

Region	Age Group		Age Group					
			15-19	20-24	25-29	30-34	35-39	40-44
Region	Balaka	A	0.414	0.398	0.323	0.31	0.281	0.086
		P	0.426	0.421	0.363	0.298	0.232	0.081
		N	162	226	155	142	121	81
	Mchinji	A	0.369	0.357	0.32	0.267	0.177	0.082
		P	0.368	0.374	0.325	0.293	0.211	0.063
		N	130	319	291	172	113	85
	Rumphu	A	0.433	0.382	0.282	0.253	0.176	0.038
		P	0.416	0.379	0.28	0.244	0.205	0.055
		N	120	233	252	225	136	185
Schooling	None	A	0.375	0.492	0.27	0.263	0.31	0.085
		P	0.423	0.403	0.338	0.321	0.246	0.075
		N	24	61	89	95	84	82
	Primary	A	0.395	0.366	0.315	0.281	0.183	0.053
		P	0.4	0.386	0.314	0.267	0.205	0.06
		N	332	596	520	381	262	247
	Secondary	A	0.482	0.372	0.292	0.238	0.167	0.045
		P	0.427	0.398	0.312	0.246	0.236	0.062
		N	56	121	89	63	24	22
Polygamy	Mono	A	0.412	0.388	0.309	0.285	0.23	0.077
		P	0.413	0.398	0.331	0.291	0.24	0.072
		N	354	605	511	396	235	196
	Poly	A	0.362	0.335	0.299	0.238	0.178	0.039
		P	0.405	0.389	0.317	0.274	0.216	0.063
		N	58	173	187	143	135	155

A = actual, P = predicted, N = number of observations

Table 14: Counterfactual Analysis: Average Life-Cycle Number of Births, Experienced Child Deaths, and Children Surviving Beyond Age Four

	Environment			
	Baseline	No HIV	No Tesing	No MTCT
Births	7.03	7.27	7.02	7.05
Experienced Child Deaths	1.22	1.14	1.22	1.11
Children Surviving beyond age four	5.81	6.13	5.8	5.94

MTCT = Mother-to-Child Transmission