NBER WORKING PAPER SERIES

HIV AND SEXUAL BEHAVIOR CHANGE: WHY NOT AFRICA?

Emily Oster

Working Paper 13049
http://www.nber.org/papers/w13049

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2007

Gary Becker, Anne Case, David Cutler, Amy Finkelstein, Matthew Gentzkow, Edward Glaeser, Emir Kamenica, Larry Katz, Michael Kremer, Adriana Lleras-Muney, Kevin Murphy, Andrei Shleifer, Jesse Shapiro, Rebecca Thornton, Alwyn Young and participants in seminars at BREAD, Harvard University, the University of Chicago and SITE provided helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

HIV and Sexual Behavior Change: Why Not Africa?
Emily Oster
NBER Working Paper No. 13049
April 2007
JEL No. I10,J17,O12

ABSTRACT

The response of sexual behavior to HIV in Africa is an important input to predicting the path of the epidemic and to focusing prevention efforts. Existing estimates suggest limited behavioral response, but fail to take into account possible differences across individuals. A simple model of sexual behavior choice among forward-looking individuals implies that behavioral response should be larger for those with lower non-HIV mortality risks and those who are richer. I estimate behavioral response using a new instrumental variables strategy, instrumenting for HIV prevalence with distance to the origin of the virus. I find low response on average, consistent with existing literature, but larger responses for those who face lower non-HIV mortality and for those who are richer. I also show suggestive evidence, based on a very simple calibration, that the magnitude of behavioral response in Africa is of a similar order of magnitude to that among gay men in the United States, once differences in income and life expectancy are taken into account.

Emily Oster
University of Chicago
Graduate School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
emily.oster@chicagogsb.edu

HIV and Sexual Behavior Change: Why not Africa?

Emily Oster*
University of Chicago and NBER

April 17, 2007

Abstract

The response of sexual behavior to HIV in Africa is an important input to predicting the path of the epidemic and to focusing prevention efforts. Existing estimates suggest limited behavioral response, but fail to take into account possible differences across individuals. A simple model of sexual behavior choice among forward-looking individuals implies that behavioral response should be larger for those with lower non-HIV mortality risks and those who are richer. I estimate behavioral response using a new instrumental variables strategy, instrumenting for HIV prevalence with distance to the origin of the virus. I find low response on average, consistent with existing literature, but larger responses for those who face lower non-HIV mortality and for those who are richer. I also show suggestive evidence, based on a very simple calibration, that the magnitude of behavioral response in Africa is of a similar order of magnitude to that among gay men in the United States, once differences in income and life expectancy are taken into account.

1 Introduction

Approximately ten percent of adults in Sub-Saharan Africa are infected with the Human Immunodeficiency Virus (HIV) and the primary mode of transmission in the region is heterosexual sex. For this reason, sexual behavior change is a major focus of HIV prevention efforts and understanding changes in behavior is important for both predicting the future path of the epidemic and for developing policy. Existing literature generally points to limited changes in sexual behavior in Africa (Stoneburner and Low Beer, 2004; Lagarde et al, 1996b; Lindan et al, 1991; Bloom et al, 2000; Ng'weshemi et al, 1996; Williams et al, 2003; Caldwell

*Gary Becker, Anne Case, David Cutler, Amy Finkelstein, Matthew Gentzkow, Edward Glaeser, Emir Kamenica, Larry Katz, Michael Kremer, Adriana Lleras-Muney, Kevin Murphy, Andrei Shleifer, Jesse Shapiro, Rebecca Thornton, Alwyn Young and participants in seminars at BREAD, Harvard University, the University of Chicago and SITE provided helpful comments.

et al, 1999a; Thornton, 2006).[1] This fact is surprising in light of extensive behavioral responses among high risk groups – gay men in particular – in the United States (Winkelstein et al, 1987; McKusick et al, 1985; Francis, 2005). However, estimating the response of sexual behavior to a sexually transmitted infection is empirically challenging, so estimates from existing literature primarily describe changes over short periods of time in quite limited groups, which may not be representative of the overall population. Moreover, there has been only very limited efforts to understand how behavioral response might vary across individuals (a notable exception is DeWalque, 2006). This variation is potentially important for designing HIV interventions and for understanding individual behavior in the face of health risks.

In this paper I attempt to address both of these gaps in the literature. I first present new estimates of behavioral response to HIV, which rely on an instrumental variables strategy. I then consider whether variations in behavioral response across individuals are consistent with utility-maximizing choices in the face of HIV. In particular, I focus primarily on whether individuals with longer non-HIV life expectancy (who therefore have more future life years to lose from HIV infection) change their behavior more. I also analyze whether behavioral response differs across income groups, since richer individuals also have a greater future value of life lost if infected.

Section 2 outlines a simple theory of sexual behavior change by forward-looking, utility-maximizing individuals. The basic comparative statics can be illustrated with a simple example. Consider two men, one who expects to live for another eleven years, and a second who expects to live for another fifty years. In a world without HIV, the choice of sexual behavior need not depend on these future life expectancies. However, in a world with HIV, sexual behavior carries a risk of death from HIV, assumed to happen ten years after infection. Introducing HIV will affect the behavior of both men, since both stand to lose future years of life if they are infected. However, for the first man HIV should not affect his behavior very much, since he expects to die in eleven years even without infection. For this man, HIV infection only costs him one year of life. However, for the second man, infection with HIV

[1] An exception to this is Dupas (2006) who finds that telling young girls that older men are high risk sexual partners causes these girls to switch to younger sexual partners. Interestingly, she finds an increase in the total number of partners, perhaps reinforcing existing findings that it is difficult to encourage decreases in partnerships overall.

means losing forty years of life, so he should have a much larger response to the presence of HIV.

A similar intuition applies to individuals with different incomes: a richer man has more income to lose if he dies prematurely so his response to HIV should be larger. The framework in Section 2 formalizes the example above and delivers three comparative statics: on average, individuals should respond to increases in the HIV rate by decreasing risky behavior, and the responsiveness should be larger for those with longer expected future survival and (under reasonable assumptions) for those who are richer.

In light of these comparative statics, Section 3 of the paper turns to estimating behavioral response. There is a serious reverse causality issue inherent in this estimation. Because HIV is a sexually transmitted infection, areas with initially higher levels of sexual behavior will have higher HIV rates. Even if individuals respond to these high rates by decreasing their risky behavior, cross sectional estimates will be biased. I address this by using an instrumental variables strategy, instrumenting for HIV prevalence with distance to the origin of the virus in the Democratic Republic of the Congo. In principle, if the virus takes time to travel, moving from person to person, areas further from its origin should have lower prevalence on average.

Using data on regional-level HIV rates, I show that there is a strong correlation between distance and HIV prevalence. This correlation remains when controlling separately for latitude and longitude, as well as some simple demographic characteristics of the country. Further, it is robust to including country fixed effects in the regression, thus identifying off of only variations in distance and HIV prevalence within countries. This distance measure appears to be uncorrelated with the incidence of premarital sex in the period before HIV appeared, as well as uncorrelated with income and non-HIV life expectancy. The instrumentation strategy is discussed in more detail in Section 3.

Section 4.1 begins by using this instrumentation strategy to explore overall response to HIV. I find that, for most groups, behavioral response to HIV is negative on average – more HIV leads to less risky behavior – but the point estimates are small and not significant. In general, this is in line with much of the existing literature, which indicates very limited behavioral response overall.

I then turn to testing the comparative statics on life expectancy and income. I begin with the most direct measures of these variables available: income based on durable good ownership and life expectancy based on area-wide mortality profiles. I find significantly larger response to HIV among those with higher income, and a larger (although generally not significantly) response for those with higher life expectancy. However, coefficients on both measures are subject to omitted variable bias, as both may be correlated with education or knowledge of the epidemic, which are likely to independently impact behavior change. Further, the measure of life expectancy is subject to significant measurement error, which would bias the coefficient towards zero and could explain the lack of significance.

To address both of these identification issues in the case of life expectancy, I consider how responsiveness varies with two explicit mortality shifters: malaria prevalence and, for young women only, maternal mortality.[2]Both of these variables are much more precisely measured than the inferred future life expectancy variable. In addition, I argue that these analyses are less contaminated by the omitted variable bias. Malaria prevalence is largely determined by climate, which should be exogenous. In the case of maternal mortality, although the *level* of mortality is unlikely to be exogenous, I take advantage of the fact that it only contributed to mortality risk for young women. By comparing the response of young women to older women, and then that difference in response to the difference between young and old men, I am able to employ a difference-in-difference identification strategy which avoids issues created by the correlation in levels.

Both of these analyses can be thought of as an instrumentation strategy – instrumenting for life expectancy with the mortality risks that contribute to it – or as a direct test of a competing mortality risk hypothesis (Dow et al, 1999). In support of the theory, for both mortality risks – malaria and maternal mortality – I find strong evidence that there are larger responses to HIV for individuals who face lower levels of these other mortality risks.

I address the concerns about omitted variable bias in the case of income more directly by showing results with and without controls for knowledge of the epidemic or for education. The effect of income is somewhat robust to these education controls, although this is not true

[2]Both of these are significant mortality risks in Africa, even for adults. For individuals aged 15-60, roughly 3% of all non-HIV deaths are estimated to be from malaria. For women of childbearing age (15-44), 30% of all non-HIV deaths are in childbirth (Lopez et al, 2006).

in all sub-samples. As a side note, the effect of malaria and maternal mortality seem to be largely unaffected by the inclusion of education controls, supporting the claim that those analyses are not subject to significant omitted variable bias.

In addition to the results from the instrumental variables analysis, I use repeated cross sectional data from a number of countries to test the theoretical predictions. These data are somewhat problematic, since they do not cover a long time period and the measures of behavior change over time. Nevertheless, to the extent that I can interpret the results, these data also provide support for the comparative statics: individuals who are richer and live in lower mortality areas have greater behavioral response.

The primary results focus on differences in behaviorial response *within* Africa. However, it may be interesting to consider whether this explanation has power to explain some of the difference in behavioral response *between* individuals in Africa and a high response group, gay men in the United States.[3] In Section 5, I calibrate a version of the model in Section 2 and derive a value of future utility losses from HIV infection, based on individual income and survival probability.[4] I then estimate responsiveness to this utility loss. Since this measure adjusts for differences in income and survival, under the assumptions of the model we expect similar behavioral response across groups even if they differ on these key variables. I find that, after this adjustment, the magnitude of response among men in the two areas is of a similar order of magnitude. Although it is obviously sensitive to the assumptions of the model, this result suggests that differences in the cost of infection may play a non-trivial role in understanding some of the differences in behavioral response across areas. This final analysis is also relevant as an input to policy. I show simple simulations with endogenous behavioral response and consider the overall and distributional consequences of various policy interventions.

Within economics, this work is perhaps most closely related to the theory of competing mortality risks outlined in Dow, Philipson and Sala-I-Martin (1999). The evidence on life expectancy, especially to the extent it relies on malaria and death in childbirth as

[3]For overall information on responsiveness to HIV among gay men, see Winkelstein et al, 1987; McKusick et al, 1985; Francis, 2005

[4]This value is obviously sensitive to the assumptions of the model; to the extent possible, I draw these assumptions from existing literature (Becker, Philipson and Soares (2005) and Murphy and Topel (2006).)

shifters of mortality, is largely a test of this competing mortality risk theory. The broader argument, and in particular the calibration, are also closely related to Becker et al (2005) and Murphy and Topel (2006). However, while those papers are generally concerned with gains associated with increases in life expectancy, this work focuses on the converse: losses associated with a decline in life expectancy. Finally, these results have the potential to contribute to the literature on the relationship between socioeconomic status and health (Grossman, 1972; Fuchs, 1982; Becker, 1993; Grossman and Kaestner, 1997; Kenkel, 1991; Lleras-Muney, 2005; Cutler, Deaton and Lleras-Muney, 2006), providing support for the theory that one reason we see more health investment behaviors among richer people is that they have more to lose from early mortality (Becker, 1993).

This paper also has implications for more specific studies of the HIV/AIDS epidemic. Much of the existing work on behavior change focuses on cultural barriers to changing behavior – fatalism, low levels of female bargaining power and others (Amuyunzu-Nyamongo et al, 1999; Varga, 1999; Orubuloye and Oguntimehin, 1999; Caldwell, Orubuloye and Caldwell, 1999; Lagarde et al, 1996b; Philipson and Posner, 1995). Certainly the results here do not rule out a role for these variables. However, the results do suggest that standard economic theory may provide significant insight and explanatory power, *without* having to rely on cultural or taste-based differences across areas.

Finally, from a policy standpoint, the results here may provide guidance to HIV/AIDS prevention campaigns, especially since changing sexual behavior is a primary focus of many of these efforts (Green, 2003; Stoneburner and Low Beer, 2004). To the extent that other mortality risks lower the incentive to avoid HIV, interventions that decrease these risks could have positive externalities for HIV prevention.

2 Theoretical Framework

This section outlines a simple theoretical framework for analyzing choices of sexual behavior in a world with HIV. An individual lives a maximum of two periods. He lives for certain in period 1, and there is a chance, p, of surviving through period 2. Each individual receives utility from certain income, y, which is the same in each period, as well as from sexual

partners in each period, σ_1 and σ_2. Total utility in period i is $u(y, \sigma_i)$, and I will assume that $u(.)$ is concave in both y and σ_i. In addition, I will assume that $u_{\sigma\sigma\sigma}(\sigma, y) = 0$.

In a world without HIV, total lifetime utility can be written

$$U_{tot} = u(y, \sigma_1) + pu(y, \sigma_2) \tag{1}$$

Income is fixed in each period so the only choices individuals make are about sexual behavior. The first order condition defining the choice of σ_i is $u_{\sigma_i}(y, \sigma_i) = 0$. Note that optimal choice of σ_i can vary with y in this framework. The direction of that relationship will depend on the sign of $u_{\sigma y}$: if the cross partial is positive, richer people will have more sexual partners; if it is negative, they will have fewer.

I assume that if an individual is infected with HIV in period 1 they have no chance of living until period 2. Given σ_1 sexual partners in period 1, an HIV rate of h and a transmission rate (chance of infection per partnership with an infected person) of β, the chance of infection is approximately $\sigma_1\beta h$.[5] The overall chance of survival to period 2 is $p(1 - \sigma_1\beta h)$. In this case, total lifetime utility is

$$U_{tot} = u(y, \sigma_1) + p(1 - \sigma_1\beta h)u(y, \sigma_2) \tag{2}$$

The choice of σ_2 is unaffected, as sexual partners in the second period do not affect survival. However, the choice of σ_1 is now defined by a new first order condition:
$u_{\sigma_1}(y, \sigma_1) - p\beta h u(y, \sigma_2) = 0$.

I am interested in three comparative statics: the effect of changes in the HIV rate on behavior, and the mediating effects of income and non-HIV life expectancy on this relationship. These comparative statics are summarized in Proposition 1.

Proposition 1. *1. $\frac{d\sigma_1}{dh} < 0$: on average, individuals should decrease their number of sexual partners when the HIV rate increases.*

2. $\frac{d(d\sigma_1/dh)}{dp} < 0$: people with greater non-HIV life expectancy decrease their risky behavior more in response to increases in the HIV rate

[5]The infection probability $h\beta\sigma$ is exactly correct for the first sexual partner for each individual. For someone with n existing sexual partners, then the additional probability of infection with any new parter is $\left(\left(1 - (1 - h\beta)^{n+1}\right) - (1 - (1 - h\beta)^n)\right)$. At low values of n, $h\beta$ will be an extremely good approximation to this; it fails to be a good approximation as n increases into the double and triple digits. However, since very few people in this sample have more than two partners total, the assumption seems reasonable.

3. $\frac{d(d\sigma_1/dh)}{dy} < 0$ *if* $u_{\sigma\sigma}(y,\sigma_1)\,u_y(y,\sigma_2) - u(y,\sigma_2)\,u_{\sigma\sigma y}(y,\sigma_1) < 0$*: people who are richer will decrease their risky behavior more in response to increases in the HIV rate as long as the condition holds.*

Proof. All three results follow from differentiating the first order condition.

1. $\frac{d\sigma_1}{dh} = \frac{p\beta u(y,\sigma_2)}{u_{\sigma\sigma}(y,\sigma_1)}$. Since $u(.)$ is concave in σ the denominator is negative. The numerator is positive, since utility is assumed to be positive, as are the probability of living to period 2 and the transmission rate.

2. $\frac{d(d\sigma_1/dh)}{dp} = \frac{\beta u(y,\sigma_2)}{u_{\sigma\sigma}(y,\sigma_1)}$. Again, the denominator is negative, and the numerator is positive.

3. $\frac{d(d\sigma_1/dh)}{dy} = p\beta\frac{u_{\sigma\sigma}(y,\sigma_1)u_y(y,\sigma_2) - u(y,\sigma_2)u_{\sigma\sigma y}(y,\sigma_1)}{(u_{\sigma\sigma}(y,\sigma_1))^2}$. In this case, the denominator is positive, as is the $p\beta$ multiplier so the overall sign depends on the sign of the fraction numerator.

□

Two of the comparative statics are unambiguous: risky behavior should decrease with increases in the HIV rate on average, and this decrease should be larger for those with higher non-HIV life expectancy. The comparative static on income is ambiguous and hinges, in large part, on the sign and magnitude of $u_{\sigma\sigma y}$. If this is positive (i.e. concavity of $u(.)$ with respect to sigma is decreasing with income), zero, or negative and small, then the comparative static holds. On the other hand, if $u_{\sigma\sigma y}$ is negative and large it is possible that the comparative static could be positive. In Appendix A, I argue that the data suggests a value of $u_{\sigma\sigma y}$ very close to zero, in which case this condition holds and we would expect richer people to decrease their behavior more in response to increases in the HIV rate. Going forward in the paper, I will rely on that analysis and expect a negative relationship between income and behavioral responsiveness. However, it is important to keep in mind that there are certainly utility functions and calibrations that would *not* deliver a comparative static in this direction.

This very simple framework formalizes the intuition from the example in the introduction: on average, we expect behavioral responses to HIV. However, we should not expect these responses to be the same for all individuals. People with greater non-HIV life expectancy are expected to respond more to the epidemic, as are people who are richer. The next two sections focus on testing these predictions in data from Africa. In Section 5 I then consider whether this type of model may help us understand some of the difference in behavioral response between Africa and the United States.

3 Data and Instrumentation Strategy

3.1 Data

The data used in this paper come from the Demographic and Health Surveys (DHS), which are household surveys that have been run in a number of countries in Africa beginning in the late 1980s. The surveys focus on fertility, contraception and child health. As a corollary, questions are asked about sexual behavior; in later surveys these include questions about extramarital sex, as well as premarital sex and sex within marriage. In the most recent surveys, modules have been added about HIV and there are fairly detailed measures of HIV knowledge. A set of controls for all regressions are drawn from the DHS, including information about socioeconomic status (education and possessions, although not directly about income), age, number of children and work activities. The analysis is limited to surveys run between 1999 and 2003, for which the questions on sexual behavior are directly comparable and there is good information on HIV knowledge. The countries included are Benin, Burkina Faso, Ethiopia, Ghana, Kenya, Malawi, Mali, Namibia and Zimbabwe.

The analysis focuses on two dependent variables: a dummy for reporting more than one partner in the last year, and the total number of partners in the last year. Both measures ignore the issue of condom use, which is one possible margin on which individuals change their behavior. Reliable data on condom use are not available in the surveys here, making it difficult to consider the issue of condom use simultaneously with the issue of number of sexual partners. In Appendix C, I analyze condom use separately using several proxies from the surveys. In general, there seems to have been little movement on this margin, so ignoring it for the main analysis is unlikely to significantly affect the results.

Table 1 presents summary statistics. Roughly 78% of the sample are women, reflecting the fact that the survey focuses on women and generally has much larger female samples. Average age of those sampled is 28.4 for women and 30.9 for men, with 68% of women and 60% of men currently married. Women in the sample average 3.6 years of education, about 1.5 years less men. About 30% of the sample is urban and the average number of children is around 3. Finally, the share of women having more than one partner is only around 2%, although for men it is over 11%, reflecting the much higher rates of risky

behavior among men.

As I detail in the next subsection, the instrumental variables strategy in this paper relies on distance from viral origin as the shifter of HIV prevalence. Crucially for this procedure, a large subset of the DHS have GPS data on location. For each survey cluster (about 200 per survey) information is reported on the longitude and latitude. This is used to calculate the distance to each cluster from the virus origin.

Evaluating the comparative statics in Section 2 also requires data on income and some measure of future life expectancy for each individual. Neither of these is directly measured in the data. The best non-education measures of income in the DHS are data on durable good ownership and housing characteristics. As the primary measure of income, I therefore sum the answers to the questions about durable good ownership (i.e. the sum of the response to whether the household has a TV (0/1), has a radio (0/1) and so on) and whether the house has electricity.[6] In Section 5, when I discuss calibration, it will be necessary to have a dollar value of income. For that, I impute income based on another survey, the Living Standards Measurement Survey (LSMS), which was run in the Ivory Coast and South Africa during this period. The LSMS surveys have information on income and expenditures, as well as data on durable good ownership. I regress log per capita household expenditures on durable good ownership, GDP and individual education in the LSMS and then use the coefficients to predict income for individuals in the DHS, based on their country-level GDP, individual durable good ownership and education.

There is also no direct measure of future life expectancy in the DHS. The ideal measure – the one which maps most directly to the theory – would be expected future years of life given an individual's current age and the non-HIV mortality path in their region. The challenge is to identify the appropriate mortality path by region. More specifically, it is to identify the appropriate life table for each area, which would provide mortality probabilities by age without HIV.[7] In order to identify the mortality profile that most closely matches each

[6]The goods are: TV, radio, refrigerator, motorbike, car and bicycle. An alternative would be to use a Filmer-Pritchett strategy (Filmer and Pritchett, 2001) to extract the first principal component of these ownership variables and use that. This gives similar results (available from the author).

[7]Generically, model life tables use data on mortality by age from censuses and other sources to construct expected probabilities of death, by age, for given levels of development (for a more detailed description, see Indepth, 2006). I use the non-HIV INDEPTH Model Life Tables (Indepth, 2006) which are specifically developed to reflect patterns of mortality in Africa in the absence of HIV. The INDEPTH Model Life Tables provide over

area, I take advantage of the fact that mortality among young children is unlikely to be heavily affected by HIV, but child mortality *will* be highly correlated with adult mortality. Using child mortality histories from women in the DHS, I construct mortality rates for children aged 1-5, 5-10 and 10-14 in each region in recent years. Each region is assigned the life table that most closely matches the mortality in these three groups.[8] The information from the life tables will generate expected future years of life for each individual in the sample.

Although this measure of future years of life closely matches the intuition behind the comparative statics, the construction of the variable introduces significant noise and, further, there are concerns about omitted variable bias. As discussed, I therefore focus in addition on malaria and maternal mortality as shifters of life expectancy. The data on malaria is drawn from the MaraLite Database,[9] which provides information on malaria prevalence by region for many, although not all, of the country-regions in the sample. At the country level, these measures are closely correlated with climate-determined malaria susceptibility, as used in Sachs and Malaney (2002).

Maternal mortality is based on sibling histories from the DHS. In a subset of the DHS surveys, women are asked to list each of their siblings and report when the sibling was born and when they died (if deceased). When the sibling is dead, and is a woman, the individual surveyed is also are also asked about whether the sibling's death was related to pregnancy. Using these data, I create a measure of the chance of dying in childbirth (or shortly after), by region, for the subset of countries in the sample with the sibling histories provided. This maternal mortality is very likely to be correlated with socioeconomic status in the area. However, the identification in this case is based on the fact that maternal mortality shifts life expectancy only for a select group of individuals – namely, young women. I therefore estimate whether this variable affects young women more than old women and, as a falsification, whether it affects young men more than old men (essentially, this is a difference-in-difference technique).

---

100 different possible mortality paths, differentiated based on level of development.

[8]Formally, denoting actual mortality among age group a as m_a and life table mortality among that age group as l_a, I minimize: $(m_{1-5} - l_{1-5})^2 + (m_{5-10} - l_{5-10})^2 + (m_{10-14} - l_{10-14})^2$

[9]Available at http://www.mara.org.za/.

3.2 Instrumentation Strategy

The primary results in this paper will focus on estimating what determines the choice of sexual behavior. The goal is to estimate an equation of the form:

$$sex_{i,r} = \gamma_0 + \gamma_1 (hiv_r) + \Psi \mathbf{X_{i,r}} + \upsilon_{i,r} \tag{3}$$

where $sex_{i,r}$ is the sexual behavior of individual i in region r, hiv_r is the HIV prevalence in that region and $X_{i,r}$ is a set of individual and regional-level controls.

There is a reverse causality issue inherent in the estimation. HIV is a sexually transmitted infections: areas where people have a lot of sex are more likely to have high rates of HIV. Even if people respond to the epidemic by decreasing their risky behavior, OLS estimates may well be biased toward finding a positive relationship between HIV and sexual behavior. I address this reverse causality issue in two ways. The primary approach used is to instrument for HIV. Later in the paper, I will also take advantage of repeated cross sectional surveys to estimate changes over time within particular areas.

Using an instrumental variables strategy to estimate the causal effect requires an instrument which is correlated with the HIV rate but (excluding the effects of HIV) uncorrelated with sexual behavior. To identify a reasonable instrument, I first note that (very broadly) two factors determine HIV prevalence within a given area: the speed at which the prevalence increases and the date at which the virus is introduced. The speed of increase, in turn, is determined by sexual behavior and the viral transmission rate. Obviously sexual behavior is not a valid instrument. However, either the viral transmission rate or the arrival date of the virus, are potentially plausible instruments. In this paper, I will focus on the virus arrival date: the earlier the virus arrives in a region, the higher we expect HIV prevalence to be, all else equal.[10]

In theory, it is possible to use the date of virus arrival in each area directly as the instrument; at the country level, Oster (2005) shows a high correlation between the first date at which the virus was observed and prevalence in the late 1990s. In practice, using date

[10] An alternative would be to instrument with circumcision, which has been shown to shift the transmission rate of the virus (Auvert et al, 2005) and used elsewhere as an instrument (Abhuja et al, 2006). However, since circumcision is highly correlated with ethnic group, which is likely to be correlated in turn with behavior, this seems a less plausible instrument in this context.

directly is problematic for two reasons. First, testing early in the epidemic is very limited (in some cases, the virus arrived well before the disease was even discovered) so there are very few areas in which early epidemic levels are known. Second, due again to the limited testing, it is likely that the first date that the virus is observed is correlated with sexual behavior, since we are unlikely to observe the presence of the virus until infection is at a significant level. Instead of using arrival date directly, therefore, I take advantage of the fact that since interactions between people are more frequent when they live closer to each other, arrival date should be closely related to the distance of each region to the viral origin. Areas closer to where the first cases of HIV were discovered should see detectable HIV rates earlier than areas farther away. Unlike the first virus arrival date, however, distance is both well measured and unlikely to be influenced by sexual behavior (although it may be *correlated* with behavior; this possibility will be addressed in more detail later).

Empirically, distance from this viral origin seems to be closely linked with HIV prevalence in the late 1990s. To demonstrate this, I use data on HIV rates by region calculated based on testing of pregnant women reported in the U.S. Census HIV/AIDS Surveillance Database (USCSD). The database aggregates a large share of studies that have been done on HIV prevalence in Africa. For many of the studies, testing was limited to a specific area. I map these areas to larger regions in each country and aggregate all the studies run between 1998 and 2002 in each region to get a regional HIV rate (a region, in this case, is an area like Copperbelt in Zambia). I rely on estimates for pregnant women because they are the most widely and consistently available. For each region, distance is calculated from the center of the Democratic Republic of the Congo (DRC). The first cases of HIV were found in this country, in the late 1950s and early 1960s, scattered in cities on both sides of the country (Vangroenweghe, 2001; Sharp et al, 2001) . Additional evidence suggests that the virus originated with a group of chimps in South-Eastern Cameroon (Gao et al, 1999; Keele et al, 2006), up the river from Kinshasa. This evidence supports the theory that the virus originated from individuals eating infected monkey meat, and then spread through close urban centers. Since early cases were observed on both sides of the Democratic Republic of the Congo, I use a point in the middle of the country as the origin (-6.31, 23.59).[11]

[11]This origin point is not the area where high HIV prevalence was first observed – the first *significant* human

Figure 1 shows the relationship between log HIV prevalence and distance to the center of the Democratic Republic of the Congo.[12] The relationship is clearly downward sloping and roughly linear. Figure 1 also gives some sense of which countries are represented at each distance, by listing the countries with regions in each quantile for a subset of the quantiles. In addition, Appendix W.1 (available on the author's web site) provides a full list of the region-countries in each quantile, as well as their HIV rates.

Table 2 estimates the relationship between prevalence and distance using the data from Figure 1. Column 1 shows the relationship between distance and log prevalence with no controls. The relationship is strongly statistically significant and negative. However, it is clear from Figure 1 that prevalence has at least some geographic component. For example, areas in West Africa are generally further than areas in East Africa; if there are also regional differences in sexual practices or HIV prevention activities, this could bias the results. As a first attempt at adjusting for this, column 2 shows the regression with controls for latitude, longitude and region. Again, the coefficient is negative and strongly significant. It is smaller than the coefficient in Column 1, reflecting the fact that both distance and prevalence co-vary with east-west and north-south orientation within Africa. Column 3 adds simple socioeconomic controls (country GDP, fertility rate and secondary school enrollment) and a control for distance to the country capital city. This limits the sample a bit, and this limitation lowers the coefficient, but it remains of similar magnitude and significance.

Columns 4 and 5 of Table 2 include country fixed effects and identify the effect of distance off of variation across regions within a country. The fixed effects will control for any variables that are constant within a country – for example, any differences associated with ethnic group makeup, or differences in governmental response to the epidemic. Column 4 includes all countries and, again, the coefficient on distance is negative and significant;

nodes of the virus were probably in Rwanda or Southern Uganda. This may suggest that those areas are a more appropriate "origin". However, there are again concerns about *why* these areas had high early rates infection if the virus was first observed elsewhere. In particular, the high rates of infection may be a result of higher rates of sexual behavior, which makes these locations less exogenous origin points.

[12]The particular functional form used here – log prevalence on linear distance – is motivated by simulations. I simulated the growth of prevalence over time in a simple model in which the probability of individual interaction varies with distance between individuals. I simulate using two different functions mapping distance to interaction probability and both suggest that the best (most linear) fit model is log prevalence on linear distance. These simulations are detailed in Appendix B.

Column 5 limits to countries with at least ten regions observed (where we are likely to give enhanced precision), and the coefficient becomes larger and more significant.

These results suggest a strong relationship between distance and HIV prevalence, which does not seem to be driven entirely by regional variation. This relationship is not limited to these data; in Appendix B I use two other datasets – the Demographic and Health Surveys testing data and prevalence derived from mortality rates from Oster (2007) – and confirm the significant effect of distance on prevalence.

Despite the apparently strong relationship, and the fact that it is unlikely that distance actually *drives* sexual behavior, the instrument relies on a single source of variation. This generates at least two primary concerns. The first, specific to this context, is that distance is correlated with sexual behavior in the period before HIV appears. The ideal way to test this would be to look at the correlation between distance and sexual behavior in surveys run prior to the start of the epidemic. Unfortunately, data on sexual behavior before the epidemic is generally not available. As an alternative, I take advantage of the fact that older individuals in the DHS are asked about pre-marital sexual activity (specifically, they are asked about age of marriage and age of first sex, so it is possible to calculate whether each individual had premarital sex). Since premarital sex is one measure of risky behavior, I look at the relationship between this outcome and distance for individuals over 35, who are likely to have been engaging in premarital sex before the epidemic was widely known.

The relationship between premarital sex and distance is shown in Columns 1 and 2 of Table 3, for women and men, respectively. I collapse the data to the cluster-gender level and regress the share of individuals reporting premarital sex on distance, with country fixed effects, mimicking Column 4 of Table 2. In neither case are the coefficients significant.

A second concern is that distance is correlated with socioeconomic factors, particularly those which we will be using to test the comparative statics. Although a correlation would not obviously be a problem for testing these predictions, it would nevertheless be a cause for concern. Columns 3 through 5 of Table 3 show the relationship between distance and income, child mortality rate and malarial prevalence (an observation, as in Columns 1 and 2, is a cluster-gender average; I control for gender in all regressions). Again, none of the relationships are significant. Moreover, they also tell no consistent story in terms

of sign. Overall, these results suggest that (once adjusting for geographic confounds) distance seems to be unrelated to either pre-period sexual activity or measures of income and future life expectancy. This provides confidence in the validity of this distance as an instrument.

There is one additional complication in the estimation strategy. The HIV rate is observed at the regional level but sexual behavior is observed at the individual level. The obvious solution is to merge these two datasets together. However, the first stage estimation shown in Figure 1 and Table 2 utilizes HIV data from a much larger set of regions than those for which sexual behavior data are available. In addition, even to the extent that the countries used in the two stages are the same, the regions within the country often differ. Effectively, there is one sample with data on HIV rates and distance (but not sexual behavior), and another with data on sexual behavior and distance (but not HIV rates). To address this issue, I use a two-sample IV strategy (as outlined in Angrist and Krueger (1992) and used in Dee and Evans (2003), among others).

The estimate of interest is γ_1 in the equation (3) at the start of this section. The first stage equation estimated is

$$hiv_r = \alpha_0 + \alpha_1(distance_r) + \Phi \mathbf{Y_r} + \eta_r \tag{4}$$

Assuming the samples are independent, the coefficient estimate for γ_1 can be generated by estimating the first stage, generating predicted values and using these predicted values in the second stage. This is the standard two stage procedure; the only difference is that I have two samples, so the first stage and second stage samples are not the same. To generate standard errors, I use a Monte Carlo procedure. After estimating the first stage, I take draws from the multivariate normal distribution of the α vector. I use these draws to predict $\hat{hiv}_r$, and use that predicted value in the second stage. I then take a draw from the resulting estimate of $\hat{\gamma_1}$. Doing this 1,000 times, I generate a distribution of $\hat{\gamma_1}$ which provides a coefficient estimate and standard errors.

For the regressions with interactions, the strategy is identical. The only difference is that it is necessary to interact the predicted HIV prevalence with income, life expectancy and the other co-variates. Standard errors are calculated using the same Monte Carlo technique.

4 Results

This section presents three sets of results. I first show estimates of the effect of HIV rates on sexual behavior. I then focus on testing the comparative statics outlined in Section 2. Finally, I move away from the instrumentation strategy outlined above, and show tests of the comparative statics using data from repeated cross sections.

4.1 Response to the HIV Rate

Panel A of Table 4 shows estimates of the effect of HIV rates on a dummy for having multiple sexual partners for single women (Column 1), married women (Column 2), single men (Column 3) and married men (Column 4).[13] Panel B of Table 4 shows the same set of regressions with total number of sexual partners as the dependent variable. With the exception of married men, the effects in this table are all in the expected direction but they are not generally significant. The results are stronger for women in the sense that the magnitudes are larger relative to the average. For both single men and single women, doubling the HIV rate decreases the share reporting multiple partners by (an insignificant) 0.6 percentage points. This effect is much larger relative to the mean for women (1.8%) than for men (11.2%).

These results suggest quite limited behavioral responses to HIV in general, which is consistent with existing literature on limited behavioral responsiveness in Africa. The coefficients may be generally negative, consistent with the comparative statics, but the overall picture certainly does not point to a large shift in sexual behavior. Given these estimates, it is not surprising that large shifts over time in behavior have not been widely observed.

4.2 Response Interactions with Life Expectancy and Income

I turn now to considering whether behavioral response differs across groups based on income and life expectancy. I begin with the most direct measures of income and life expectancy available: durable good ownership and imputed future years of life left. The regressions are

[13]This table, like the rest, show the full second stage estimates with Monte-Carlo standard errors for the coefficients of interest. Web Appendix W.2 shows the full second stage estimates, including controls, although without adjusting the standard errors.

similar to those in Table 4, although I combine the data on single and married individuals and simply include an interaction between marital status and HIV rate.[14] In addition, in this case, all regressions allow for interactions between country dummies and HIV rate. This means that any effect of life expectancy or income should be identified off differences *other than* those that arise from differences across countries.

Table 5 shows the baseline results: the effect of the interaction between HIV rate and income (based on durable good ownership) and expected future years of life (based on calculations from life table). In Columns 1 and 2 the dependent variable is a dummy for having two or more partners; in Columns 3 and 4 it is the total number of partners. Directionally, the comparative statics seem to be supported. All but one of the coefficients are in the expected direction. In the case of income, the results are consistently significant. In the case of life expectancy, however, the coefficients are generally not significant. As discussed, this may well reflect measurement error biasing the coefficient toward zero.

Malaria and Maternal Mortality

In order to avoid both the measurement error and the possible omitted variable bias inherent in the regressions in Table 5, Table 6 shows the effect of malaria prevalence and maternal mortality on behavioral response. Panel A focuses on malaria: the regression includes a control for income interacted with HIV, but replaces the interaction with life expectancy with an interaction with malaria prevalence. In this case, since higher prevalence of malaria lowers life expectancy, the comparative statics would suggest a positive coefficient.

Indeed, the effect of malaria does seem to be positive: for both genders and both measures of behavior, living in an area with higher rates of malaria seems to lead to less behavioral response to HIV and in all cases these results are highly significant. This effect can be seen graphically in Figure 2, which plots the responsiveness to HIV across areas based on their malaria prevalence. The divisions are within country, so the low malaria prevalence regions are the bottom one-third of regions in each country, the medium are the middle one-third and the high malaria are the top one-third (this ensures that differences across

[14]This makes very little difference and is done to simplify the presentation of results. All of the results in Tables 5 and 6 are qualitatively similar if I run the regressions separately by single and married. Details are available from the author.

groups is not driven by differences across countries in overall malaria prevalence). This figure demonstrates increasing coefficients as malaria prevalence rises: behavioral response to HIV is most negative in areas with low malaria prevalence, less negative in the middle regions, and positive (although not significant) in regions that are highly malarial. This is true for both dependent variables considered.

Panel B of Table 6 turns to the effect of maternal mortality. In this case, the coefficient of interest is a dummy for being young (under 21), interacted with maternal mortality and with HIV prevalence. For women, we expect this coefficient to be positive: young women in areas with high maternal mortality should change their behavior more (relative to older women) than young women in areas with low maternal mortality. For men, we do not expect a positive interaction.

The results in Panel B demonstrate strong evidence of this pattern. For women, the triple interaction coefficient is positive and significant; for men, it is negative and largely insignificant. Again, a graphical representation of this can be seen in Figure 3, which plots the responsiveness of young women relative to older women in areas with high, medium and low levels of maternal mortality. As with Figure 2, these divisions are within country. Again, the data suggests the relative responsiveness becomes more positive as we move from regions of low maternal mortality to regions of high mortality.

The magnitude of these effects is difficult to interpret directly since malaria and maternal mortality are not measured on the same scale. However, it is possible to re-scale both coefficients based on the relationship between future years left (the direct life expectancy measure) and either malaria prevalence or maternal mortality. The bolded figures at the bottom of Panels A and B in Table 6 show these re-scaled coefficients. The range of re-scaled coefficients is actually quite similar for malaria and maternal mortality, which is encouraging as it suggests they may both be reflecting the same underlying effects.

Separating Education and Knowledge

The preceding results provide some evidence that life expectancy in general, or competing mortality risks more specifically, causally affect behavioral response to HIV. The results on income in Table 5, however, are still difficult to interpret given the possible omitted variable

bias. In particular, there are concerns that the effect of income is simply an effect of education, which may improve knowledge of the epidemic, or may reflect other individual characteristics (i.e. discount rates) that could affect behavioral response (Fuchs, 1982).

Unfortunately, no obvious instrument for permanent income presents itself. However, it is possible to see whether the results are robust to including controls for epidemic knowledge or for education. Since these are likely to be the primary omitted variables, a finding that the results are robust to inclusion of these controls would provide support for interpreting the coefficients as an effect of income directly.

Table 7 shows the results controlling directly for knowledge of the epidemic interacted with HIV rate. Knowledge in this case is measured based on the response to two questions about HIV: can you get HIV from having multiple partners and can you avoid HIV by using condoms. Panel A simply replicates the analysis in Table 5. Panels B and C show the effects of malaria and maternal mortality with these knowledge controls. Among other things, this provides an additional test of whether those results are driven by omitted variables.

Allowing for the response to vary with knowledge does not seem to have a large effect on the income coefficients. In all three panels, the effect of income is very similar to what we see in Table 5 or Table 6. With the simple control for future life expectancy the coefficients on income are highly significant. With the controls for malaria and maternal mortality the coefficients are generally negative, although not always significant (which is also what we see in Table 6). This may reflect a smaller sample size or correlation between income and disease burden. Regardless, the results in Table 7 suggest that differences in knowledge are *not* driving the coefficients on income.

Table 8 replicates Table 7, but with controls for educational attainment rather than knowledge interacted with HIV. It is important to note that the control for education may actually, in a sense, "over-control". Durable goods are likely to be a noisier measure of income than education. If this is true, then some of the estimated effect of education may actually be an effect of income or mortality, leading us to understate the importance of those variables. Nevertheless, in Panel A the results seem to be largely robust to the inclusion of the education control; three of the four coefficients remain negative and significant. In Panels B and C the significance on the income coefficient is somewhat lower, in line with the previous

evidence from these regressions. It is important to note that the inclusion of the education control has very little impact on the coefficients on malaria prevalence or maternal mortality, suggesting that these results are not driven by any omitted variable bias, which is consistent with the claim that they are well-identified.

In general, the results in Tables 7 and 8 suggest that, although some of the effect of income estimated in Table 5 is likely to be an effect of non-income components of education, there does seem to be evidence of an effect of income on behavior change directly, over and above any education-related determinants of behavioral response.

4.3 Results from Repeated Cross Sections

The analysis thus far has relied on the instrumentation strategy outlined in Section 3. An alternative way to test the comparative statics, which can also address the reverse causality issues, is to take advantage of the fact that a number of countries have multiple DHS surveys from the 1990s. By comparing behavior over time within a particular area it is be possible to estimate behavior change. The primary drawback of this approach is inconsistency in the data over time. Until 1996 no questions were asked about extramarital sex or number of partners. This means that there is only a very limited time period in which I can observe changes in this variable. The surveys do ask consistently about premarital sex so it is possible to use all of the surveys to analyze changes in this variable, which will give some information as long as premarital sex in general is correlated with having multiple partners. I will consider the response of both extramarital sex (after 1996) and premarital sex (in the entire sample).

The repeated cross sectional analysis is run with nine countries: Benin (1996, 2001), Burkina Faso(1992, 1998, 2003), Mali (1996, 2001), Ghana (1993, 1998, 2003), Cameroon (1991, 1998, 2004), Kenya (1993, 1998, 2003), Nigeria (1991, 1999, 2003), Zambia (1996, 2001) and Zimbabwe (1994, 1999). The independent variables of interest are, as above, log HIV rate and log HIV rate interacted with income and expected survival path. Consistent data on HIV rates over time in Africa is difficult to obtain. As an alternative, I use the UNAIDS estimates of prevalence for 2005 and assume that HIV rate increased linearly from the early 1980s to 2005. This is clearly a simplification but it *will* capture both the fact that HIV has increased over time and the fact that this increase has been larger in some countries than in others. As

above, the measure of income is number of durable goods and the measure of life expectancy is child mortality.

Table 9 shows the effect of these variables on extramarital sex for women and men (Columns 1 and 2) and premarital sex for women and men (Columns 3 and 4). All regressions include region specific fixed effects, which serve to address the reverse causality concerns. These results are supportive of the instrumental variables results. The effect of income is negative and significant in all cases, and the effect of child mortality is positive and significant or close to significant.

Overall, the results in this section provide support for the comparative statics outlined in Section 2. On average, behavioral response to HIV is in the expected direction, although it is small and generally not significant. However, looking within particular groups suggests that heterogeneity across individuals is important to take into account. Consistent with a model of forward-looking individuals making utility-maximizing choices about sexual behavior, richer people and those with greater expected non-HIV survival respond more to changes in the HIV rate. The next section discusses the magnitude of these effects and policy implications.

5 Calibration and Policy

Understanding the direction of the effects analyzed above is important for testing the model outlined and for thinking about whether there is a role for a model of optimizing agents in understanding the response of sexual behavior to HIV. These comparative statics, however, tells us relatively little about the magnitude or importance of this explanation. In particular, a contrast is often made between limited behavioral response in Africa and the extensive behavioral response among gay men in the United States. The results in the previous section suggest a possible resolution to some of this apparent puzzle since gay men in the United States are richer and longer-lived than men in Africa. However, the regressions above do not give a sense of *how* much of the puzzle might be resolved by this explanation.

This section takes a first pass at addressing this question by explicitly modeling the future utility loss from an additional sexual partner (the "price" of a partner) for each individual in the sample and estimating response to this utility loss. I compare the magnitude

of this response among individuals in Africa to the magnitude among gay men in the United States. I then use these estimates of magnitude to evaluate policy interventions in a simulation model of the epidemic with endogenous behavioral response.

5.1 Calibrational Model

The calibration relies on the framework set up in Section 2. The survival function is more complex (and draws heavily on the work in Becker, Philipson and Soares (2005) and Murphy and Topel (2006)), but I simplify the model considerably by assuming that utility for sex is separable from utility for income. This is a significant assumption (although somewhat supported in Appendix A) but it is necessary in order to derive straightforward calibration results.

Individuals have certain income y in each period of life and a survival path $S(t)$, which defines a probability of death at each future age t. Income in each period delivers utility $u(y)$ and in the current period individuals can also choose a number of sexual partners σ, which delivers utility $\mu(\sigma)$ (in this version of the model there is no sex in future periods). For an individual currently of age α, with discount rate r, total forward looking utility in the current period without any HIV is

$$U_{tot} = u\left(y\right) \int_{\alpha}^{\infty} \exp\left(-rt\right) S\left(t\right) dt + \mu(\sigma) \tag{5}$$

As in Section 2, the HIV infection probability given σ sexual partners is the product of the prevalence h, transmission rate β and σ. I assume that if an individual contracts HIV this period they die with certainty in ten years. I define discounted survival utility without HIV, $\int_{\alpha}^{\infty} \exp\left(-rt\right) S\left(t\right) dt$, as S_A and discounted survival utility with HIV, $\int_{\alpha}^{\alpha+10} \exp\left(-rt\right) S\left(t\right) dt$, as S_H. Total utility for someone facing a world with HIV rate h is

$$U_{tot} = u(y)\left((\sigma\beta h)S_H + (1 - \sigma\beta h)S_A\right) + \mu(\sigma) \tag{6}$$

The first order condition for the choice of σ is given by equation 7.

$$0 = u\left(y\right)(\beta h)(S_H - S_A) + \mu'\left(\sigma\right) \tag{7}$$

The comparative statics are directionally equivalent to the model in Section 2. In addition, with assumptions about the functional form of utility over income it will be possible to actually estimate a "price" per sexual partner in terms of lost future utility

The price of a sexual partner is simply the future utility losses from HIV infection, multiplied by the probability of infection with a given sexual partner. For an individual with income y, the lost utility from HIV infection is $u(y)(S_A - S_H)$. That is, their total future utility without infection $u(y)S_A$ minus their total future utility with HIV infection, $u(y)S_H$. The probability of infection with a given sexual partner is βh, so the utility cost of a sexual partner, Φ, is described by equation 8.

$$\Phi = \beta h u(y)(S_A - S_H) \tag{8}$$

Given this definition of Φ, $\frac{d\sigma}{d\Phi} = \frac{1}{\mu''(\sigma)}$. This implies that a regression of σ on Φ should deliver coefficients that are comparable across any samples with the same $\mu''(\sigma)$. In other words, for any two groups with the same concavity in their utility for sexual partners, the coefficient should be of the same magnitude. This is not, of course, true for a regression of behavior on HIV rate directly, which will deliver coefficients that depend on this concavity *and* on income and life expectancy.

Given a particular functional form for utility, and individual income and survival path, it is possible to calculate the utility cost of a sexual partner for each individual. Following Becker et al (2005) and Murphy and Topel (2006), I assume CRRA utility $u(c) = \frac{c^{1-1/\gamma}}{1-1/\gamma} + \nu$, where ν is a renormalization so the utility in death is zero. I use values of γ and ν drawn from the existing literature: $\gamma = 1.25$ and $\nu = -14.97$. The survival path relies on the same life tables used to test the comparative statics. As discussed in the data section, the income measure is the imputed income based on individual education, durable goods and country GDP. Transmission rates for Africa are drawn from Oster (2005).

5.2 Response to Price of a Sexual Partner

This section estimates individual responsiveness to the price of a sexual partner. I do this for men and women in Africa and, additionally, for a sample of gay men in the United States. The data on gay men are from an individual-level panel dataset, the MultiCenter AIDS

Cohort Study (MACS), which began in 1984 and followed men through the late 1990s. This study covers roughly 5000 men in Baltimore, Chicago, Pittsburgh and Los Angeles. Men in the study completed surveys about behavior at the baseline and at roughly yearly follow-up sessions.

For gay men, I focus on the period before 1990 to avoid any time when treatment was available. Since the data is an individual-level panel, reverse causality issues can be avoided by controlling for individual-specific fixed effects. The dependent variable is a dummy for reporting two or more partners in the last two months, roughly parallel to the dummy for reporting multiple partners in the African data. This is a very high risk sample of men, with 60% of them reporting multiple partners in an average survey round. The large difference in risk profiles leads to questions about the comparability of the samples but the comparison exercise may still be meaningful. The MACS lacks direct data on income so I match individuals to the census IPUMS for 1980 based on their education and occupation group and assign them the average income in that group. I assume that everyone in the sample has the same mortality profile and use the 1980 Census-based Life Tables for the U.S. to calculate expected survival.

The only conceptually difficult element of the analysis of gay men is what HIV prevalence to use. The MACS did testing in every period, so one possibility is to use HIV prevalence in the sample as the prevalence in each period. However, during the period from 1984 to 1990 the prevalence increases only from 37 to 40%. This very small increase reflects the fact that in this community prevalence was extremely high before AIDS was widely recognized. In other words, over this period *perceived* prevalence is likely to have increased a lot, even as actual prevalence stayed constant. I run specifications with two possible assumptions about prevalence. First, assuming actual prevalence and perceived prevalence are the same; second, assuming perceived prevalence increased linearly from zero in 1984 (at the start of the sample) to 40% in 1990. It is worth noting, before estimating the response to price, that for the sample of gay men the average response to HIV is very large. Depending on which of the two prevalence paths we use, the coefficient in a regression of behavior on log HIV rate ranges from -3.18 to -0.17, compared with a range from -0.006 to 0.13 for men in Africa.

The results of the estimation are shown in Table 10. In this case, there are no

interactions since the HIV rate is a part of the price per partner. For both men and women in Africa, the coefficient on utility cost is negative and statistically significant. The coefficients for gay men in the United States are also highly significant. If we assume actual prevalence is equal to perceived prevalence (Column 3), the coefficient for gay men is higher than in Africa, although it is much closer in magnitude than if we simply regress behavior on HIV rate without any income or survival adjustments. When we assume a more realistic increase in perceived prevalence (Column 4), the coefficients for men in both groups are actually quite similar, and not statistically distinguishable.

If we take the calibrational model seriously, this result suggests that at least some fraction of the difference in behavioral response between the United States and Africa may be explained by differences in the future value of life. This conclusion relies largely on comparing across the two groups of men. For women responsiveness is much smaller. Of course, it is extremely difficult to draw strong conclusions from these results. The model contains serious simplifications, and the empirical work relies on a large number of assumptions. Nevertheless, the fact that the coefficients are of the same order of magnitude – and that we see decreases in sexual behavior among both groups – suggests that this explanation could have some quantitative importance.

5.3 Policy Interventions

The focus of this paper is primarily positive. However, the results may be an important input into evaluating the effect of different policy interventions. The magnitudes calculated above can be seen as inputs to allowing endogenous behavioral response in an epidemiological simulation model of the epidemic.

I set up a simulation model similar to that in Oster (2005). Using data on sexual behavior by gender, income and age group, I simulate a large population of individuals, keeping track of their sexual behavior and HIV status over time. Although the simulation parallels Oster (2005) in most dimensions, the results in this paper allow me to add in individual behavioral responses to HIV. I assume individuals are myopic and behavior in each period is simply a function of HIV rate in that period.[15] For each individual in the simulation,

[15]This is as opposed to individuals choosing a time path of behavior for all future periods based on the HIV

in each period, I calculate their cost of an additional sexual partner (based on their income and age) and then use the coefficients in Table 10 to alter their sexual behavior in response to the epidemic. The model will output HIV prevalence over time, broken down by age group, income group and gender.

The first thing to note, based on the simulation model, is that including endogenous behavior change makes relatively little difference in the magnitude of the HIV rates predicted by the model. Figure 4 shows predicted HIV prevalence by the model with and without behavior change. Even 35 years into the epidemic the predicted difference in prevalence is only around 0.5 percentage points. This reflects the relatively small degree of behavior change estimated.

Perhaps more importantly, the results here have implications for evaluating interventions to limit the impact of HIV. Existing discussions of interventions such as antiretrovial treatment rarely address the possibility of endogenous behavioral response. This paper allows for endogenous behavioral response and presents the first estimates that will allow me to estimate a magnitude for the responsiveness. The results here also allow me to evaluate distributional consequences of these interventions – how might a particular intervention shift the burden of the disease toward the richer individuals in society, for example.

I use the simulation model to evaluate three interventions. The first is drug treatment. I assume that when drugs are introduced time to death increases by five years, so the cost of risky behavior goes down. The second intervention is education. In the baseline simulations, I assume that individual perception of the HIV rate is lower than it actually is – 50% of the truth. The educational intervention is assumed to increase perception to 100% of the truth. Finally, I consider what would happen if everyone had $1000 extra income per year.

I simulate the model for 35 years and introduce the intervention at year 25. Figure 5 shows the HIV incidence by year (beginning 15 years in) for the case of no intervention, and for each intervention considered.[16] All three interventions have some impact on prevalence,

rate now and what it would be in the future, which is a function of their behavior now.

[16]I use incidence rather than prevalence to make the drug intervention more comparable. The prevalence will increase faster in the drug intervention case simply because fewer people are dying. I am not interested in capturing changes in prevalence that result from this fact.

although not in the same direction or the same magnitude. The educational intervention would decrease incidence, as would making everyone wealthier. Both of these effects, however, are extremely small. The effect of treatment is somewhat larger in the other direction. This is true despite the fact that changes in *behavior* are relatively small. The key is that when treatment is introduced not only does risky behavior increase, but the prevalence in the population increases because fewer individuals are dying. This means that the infection rate among sexual partners goes up: not only are there more risky acts but the risk of infection per act is increasing.

The interventions also vary somewhat in their distributional consequences. Figure 6 shows the difference in prevalence rates for the wealthiest and least wealthy groups over time under each intervention. On average, the epidemic is skewed towards the wealthier early on but they change their behavior more and the relationship between income and prevalence switches sign. This effect is exacerbated by the interventions that increase income and by educational campaigns. Treatment, however, has the opposite effect. Introducing treatment means more for richer people, since the potential lost future income is higher for them which produces a mitigation of the trend toward poorer people having higher prevalence.

These results point to the value of the parameters in this paper for thinking about policy. Although the magnitude of these effects is relatively small, there is *some* cause for concern about increases in risky behavior when treatment is introduced or made more widely available. This is particularly true among the wealthier or younger cohorts. This argues, among other things, for extremely carefully considered interventions and, perhaps, simultaneous emphasis on positive behavior change. This model also gives some sense of the value of educational interventions and suggests, in particular, that these interventions will be the most effective for wealthier or younger individuals.

6 Conclusion

The HIV epidemic has devastated Africa and understanding behavioral responses to HIV rates is crucial both for predicting the future path of the epidemic and for preventing its future spread. This paper presents new estimates of behavioral response among the general

population in a broad group of countries in Africa. I find relatively limited behavioral response overall. However, consistent with rational choices of sexual behavior in a world with the risk of HIV, I find behavioral response is larger for those who have higher non-HIV life expectancy and those who are richer. The magnitudes of these effects seem to be relatively large, and may explain some fraction of the apparent puzzle of low behavioral response in Africa relative to the United States.

This paper focuses exclusively on responsiveness to HIV. However, the general message – that responsiveness of health behaviors should be higher among those with higher value future of life – clearly applies to other behaviors. For example, other research suggests that seatbelt use varies with income in the U.S. (Lerner et al, 2001; Shinar, Schechtman and Compton, 2001) and is higher in developed countries than in less developed countries like South Africa or China (Olukoga and Noah, 2005; Zhang et al, 2006). There is also evidence in the surveys used here that individuals do not always undertake beneficial health behaviors even when they are available. For example, 32% of individuals in the DHS report not using a bednet for their children on the previous night *even conditional on owning one*. It is possible that the framework outlined here may help us understand some of these other health behaviors in the developing world.

The results here also have the potential to provide policy guidance. As we consider interventions like drug treatment, which change the cost of infection, the possibility of endogenous behavior change is very real. The parameter estimates in this paper can be used to understand how important this endogenous behavior change is and which groups are most likely to have this response.

References

Amuyunzu-Nyamongo, M, L Tendo-Wambua, B Babishangire, J Nyagero, N Yitbarek, M Matasha, and T Omurwa, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

Angrist, Joshua and Alan Krueger, "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 1992, *87*, 328–336.

Auvert, B, D Taljaard, E Lagarde, J Sobngwi-Tambekou, R Sitta, and A Puren, "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial," *PLoS Medicine*, 2005, *2* (11), e298.

Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd ed., Chicago, IL: University of Chicago Press, 1993.

Becker, Tomas Philipson Gary and Rodrigo Soares, "The Quantity and Quality of Life and the Evolution of World Inequality," *American Economic Review*, March 2005, *95* (1), 277–291.

Bloom, SS, C Banda, G Songolo, S Mulendema, AE Cunningham, and JT Boerma, ""Looking for change in response to the AIDS epidemic: trends in AIDS knowledge and sexual behavior in Zambia, 1990 through 1998."," *Journal of Acquired Immune Deficiency Syndrome*, 2000, *25* (1), 77–85.

Caldwell, John, I.O. Orubuloyeb, and Pat Caldwell, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

___ , Pat Caldwell, John Anarfi, Kofi Awusabo-Asare, James Ntozi, I.O. Orubuloye, Jeff Marck, Wendy Cosford, Rachel Colombo, and Elaine Hollings, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

Cutler, Angus Deaton David and Adriana Lleras-Muney, "The Determinants of Mortality," *Journal of Economic Perspectives*, 2006, *20* (3).

Dee, Thomas and William Evans, "Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates," *Journal of Labor Economics*, 2003, *21* (1), 178–209.

DeWalque, Damien, "How Does the Impact of an HIV/AIDS Information Campaign Vary With Educational Attainment: Evidence from Rural Uganda," 2006. Mimeo, World Bank.

Dow, William, Tomas Philipson, and Xavier Sala-I-Martin, "Longevity Complementarities Under Competing Risks.," *American Economic Review*, 1999, *89* (5), 1357–1372.

Dupas, Pasqualine, "Relative Risks and the Market for Sex: Teenagers, Sugar Daddies and HIV in Kenya," 2006. Mimeo, Dartmouth College.

Filmer, Deon and Lant Pritchett, "Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India," *Demography*, 2001, *38* (1), 115–132.

Forston, Jane, "Mortality Risks and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa," 2006. Mimeo, Princeton University.

Francis, Andrew, "The Economics of Sexuality: The Effect of HIV/AIDS on Sexual Behavior, Desire, and Identity in the United States," 2005. Mimeo, University of Chicago.

Fuchs, Victor, *Economic Aspects of Health*, Chicago, IL: University of Chicago Press, 1982.

Gao, Feng, Elizabeth Bailes, David Robertson et al., "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes," *Nature*, 1999, *397*, 436–441.

Green, Edward, *Rethinking AIDS Prevention*, Westport, CT: Praeger, 2003.

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 1972, *80* (2), 223–255.

___ **and R. Kaestner**, *Effects of Education on Health*, University of Michigan Press, 1997.

INDEPTH Network, *INDEPTH Model Life Tables for Sub-Saharan Africa*, Ashgate, 2006.

Kalemli-Ozcan, Sebnem, "AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa," 2006. Mimeo, University of Houston.

Keele, Brandon, Fran Van Heuverswyn, Yingying Li et al., "Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1," *Science*, 2006, *313* (5786), 523–526.

Kenkel, Donald, "Health Behavior, Health Knowledge and Schooling," *Journal of Political Economy*, 1991, *99* (2), 287–305.

Lagarde, Gilles Pison Emmanual and Catherine Enel, "Knowledge, attitudes and perception of AIDS in rural Senegal: relationship to sexual behaviour and behaviour change," *AIDS*, March 1996, *10* (3), 327–34.

___ **and** ___ , "A Study of Sexual Behavior Change in Rural Senegal," *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology*, 1996, *11*, 282–287.

Lerner, E. Brooke, Dietrich Jehle, Anthony Billittier, Ronald Moscati, Cristine Connery, and Gregory Stiller, "The influence of demographic factors on seatbelt use by adults injured in motor vehicle crashes," *Accident Analysis and Prevention*, 2001, *33*, 659662.

Lindan, C, S Allen, M Carael, F Nsengumuremyi, P Van de Perre, A Serufilira, J Tice, D Black, T Coates, and S Hulley, "Knowledge, attitudes, and perceived risk of AIDS among urban Rwandan women: relationship to HIV infection and behavior change," *AIDS*, 1991, *5* (8), 993–1002.

Lleras-Muney, Adriana, "The Relationship Between Education and Health in the United States," *Review of Economic Studies*, 2005, *72*, 189–221.

Lopez, AD, CD Mathers, M Ezzati, DT Jamison, and CJLMurray, *Global burden of disease and risk factors*, Oxford University Press and the World Bank, 2006.

McKusick, Leon, William Horstman, and Thomas Coates, "AIDS and sexual behavior reported by gay men in San Francisco," *American Journal of Public Health*, May 1985, *75* (5), 493–496.

Murphy, Kevin and Robert Topel, "The Value of Health and Longevity," *Journal of Political Economy.*

Ng'weshemi, JZ, JT Boerma, R Pool, L Barongo, K Senkoro, M Maswe M, R Isingo, D Schapink, S Nnko, and MW Borgdorff, "Changes in male sexual behaviour in response to the AIDS epidemic: evidence from a cohort study in urban Tanzania," *AIDS*, October 1996, *10* (12), 1415–20.

Olukoga, Abiodun and Mongezi Noah, "The Use of Seat Belt by Motor Vehicle Occupants in South Africa," *Traffic Injury Prevention*, 2005, *6*, 398–400.

ORC Macro, "HIV Notes from Measure DHS," Technical Report, USAID 2006.

Orubuloye, I.O. and Folakemi Oguntimehin, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

Oster, Emily, "Sexually Transmitted Infections, Sexual Behavior and the HIV/AIDS Epidemic," *Quarterly Journal of Economics*, May 2005, *120* (2), 467–515.

___ , "Routes of Infection: Exports and HIV Incidence in Africa," 2007. Mimeo, University of Chicago.

Philipson, Tomas and Richard Posner, "The Microeconomics of the HIV/AIDS Epidemic in Africa," *Population and Development Review*, December 1995, *21* (4), 835–848.

Pickering, H, M Okongo, B Nnalusiba, K Bwanika, and J Whitworth, "Sexual networks in Uganda: casual and commercial sex in a trading town," *AIDS Care*, 1997, *9* (2).

Sachs, Jeff and Pia Malaney, "The Economic and Social Burden of Malaria," *Nature*, 2002, *415* (6872).

Sharp, Paul, Elizabeth Bailes, Roy Chaudhuri, Cynthia Rodenburg, Mario Santiago, and Beatrice Hahn, "The origins of acquired immune deficiency syndrome viruses: where and when?," *Philosophical Transactions of the Royal Society of London B*, 2001, *356*, 867–876.

Shinar, David, Edna Schechtman, and Richard Compton, "Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population," *Accident Analysis and Prevention*, 2001, *33*, 111–116.

Stoneburner, Rand and Daniel Low-Beer, "Population-Level HIV Declines and Behavioral Risk Avoidance in Uganda," *Science*, April 2004, *304*, 714–718.

Thornton, Rebecca, "The Demand for and Impact of Learning HIV Status: Evidence from a Field Experiment," 2006. Mimeo, Harvard Unviersity.

Vangroenweghe, Daniel, "The earliest cases of human immunodefciency virus type 1 group M in Congo-Kinshasa, Rwanda and Burundi and the origin of acquired immune deficiency syndrome," *Philosophical Transactions of the Royal Society of London B*, 2001, *356*, 923–925.

Varga, Cynthia, *Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries*, Health Transition Centre, 1999.

Williams, BG, D Taljaard, CM Campbell, E Gouws, L Ndhlovu, J Van Dam, M Carael, and B Auvert, ""Changing patterns of knowledge, reported behaviour and sexually transmitted infections in a South African gold mining community."," *AIDS*, 2003, *17* (14), 2099–2107.

Winkelstein, Warren, Michael Samuel, Nancy Padian, James Wiley, William Lang, Robert Anderson, and Jay Levy, "The San Francisco Men's Health Study: III. Reduction in human immunodeficiency virus transmission among homosexual/bisexual men, 1982-86," *American Journal of Public Health*, June 1987, *77* (6), 685–689.

Young, Alwyn, "In Sorrow to Bring Forth Children: Fertility amidst the Plague of HIV," 2006. Mimeo, University of Chicago.

Zhang, Wei, Yueng-Hsiang Huang, Matthias Roetting, Ying Wang, and Hua Wei, "Drivers views and behaviors about safety in China What do they NOT know about driving?," *Accident Analysis and Prevention*, 2006, *38*, 22–27.

Table 1. ***Summary Statistics for African Data***

	Female	**Male**
Share Female	78.7%	
Age	28.41	30.88
Married	68%	60%
Years of Education	3.57	5.05
Urban	29.7%	29.1%
# of Children	3.04	3.11
Have Multiple Partners	1.9%	11.2%

Notes: This table reports summary statistics based on the sample from the Demographic and Health Surveys in Africa. Countries included are Benin, Burkina Faso, Ethiopia, Ghana, Kenya, Malawi, Mali, Namibia and Zimbabwe.

Table 2. ***First Stage Estimates – HIV Rate and Distance to Virus Origin***

Dependent Variable: Log HIV Rate in Region					
	(1)	(2)	(3)	(4)	(5)
Sample	All	All	All	All	$\geq$ 10 Regions
Explanatory Variables:					
Distance (in 1000 km)	−.7514***	−.4457***	−.303***	−.3607**	−.4247***
	(.052)	(.069)	(.075)	(.141)	(.15)
East Region		−.28	−.4666**		
		(.233)	(.232)		
South Region		.3152	.1063		
		(.242)	(.254)		
Center Region		−.1792	−.1916		
		(.194)	(.214)		
Longitude		.0139**	.0312***		
		(.007)	(.007)		
Latitude		−.027***	−.0238***		
		(.006)	(.006)		
Log GDP			.2272**		
			(.096)		
Sec. School Enroll.			−.0087		
			(.007)		
Fertility Rate			.0059		
			(.106)		
Log Dist. to Cap'l			−.0001		
			(0)		
constant	3.755***	2.442***	.459	2.907***	3.099***
	(.12)	(.229)	(1.22)	(.308)	(.313)
Country FE	NO	NO	NO	YES	YES
Year Controls	NO	YES	YES	NO	NO
Number of Obs.	467	467	442	467	413
R^2	.31	.56	.6	.68	.64

standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: HIV rates are estimated from the U.S. Census HIV/AIDS Surveillance Database. Distance is calculated from the center of the region or country to roughly the center of the Democratic Republic of the Congo. Columns 1-4 contain all observations. Column 5 includes only countries with at least 10 regions observed. All data used for this table are shown in Appendix W.1.

Table 3. *First Stage Falsification,*

Dependent Variable	*Had Premarital Sex*		*# Durables*	*Mortality Ages 1-14*	*Malaria Prevalence*
	(1)	(2)	(3)	(4)	(5)
	Women, > 35	Men, > 35	All	All	All
Explanatory Variables:					
Distance	$-.0434$	$-.1005$	.3163	.00240	.094
	(.106)	(.117)	(.215)	(.0069)	(.134)
Male			$-.0833^{*}$	.00002	.0002
			(.05)	(.00004)	(.001)
constant	$.689^{***}$	1.042^{***}	.647	$.0482^{***}$	.14
	(.26)	(.299)	(.533)	(.0178)	(.348)
Country FE	YES	YES	YES	YES	YES
Number of Obs.	3151	2897	6318	6358	3372
R^2	.27	.12	.25	.64	.7

[a] standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: All regressions are run at the cluster-gender level, so the dependent variable represents the cluster-gender average (i.e. the average premarital sex for women in cluster 5 in Burkina Faso). The dependent variable in Columns 1 and 2 is a dummy for reporting premarital sex for individuals over 35 at the time of the survey. In Column 3 it is durable good ownership plus electricity; Column 4: region mortality rate among children 1-14; Column 5: Malaria prevalence in region.

Table 4. *Effect of HIV on Sexual Behavior – IV Estimates*

Panel A: Risky Behavior Dummy				
Dependent Variable: More than One Partner (0/1)				
	Single Women	Married Women	Single Men	Married Men
Log HIV Rate	−0.0064	−0.0056	−0.0060	0.090
	(.010)	(.024)	(.060)	(.060)
Mean of Dependent Variable	1.8%	2.1%	11.2%	13.4%
Number of Observations	28,071	58,351	13,506	15,929
Panel B: Total Number of Partners				
Dependent Variable: Total Sexual Partners				
	Single Women	Married Women	Single Men	Married Men
Log HIV Rate	−0.122	−0.0176	−0.356	0.122
	(.085)	(.029)	(.396)	(.104)
Mean of Dependent Variable	0.29	1.02	0.64	1.18
Number of Observations	28,071	58,346	13,505	15,851

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the Monte Carlo procedure outlined in Section 3. Second stage regressions with controls (although not adjusted standard errors) are shown in Appendix tables W.2.1 and W.2.2, in web Appendix W.2. Single individuals are defined as having more than one partner if they report two or more partners in the last year. Married individuals are defined as having more than one partner if they report any extramarital sex in the last year. Total number of partners includes spouse for individuals who are married.

Table 5. *HIV and Sexual Behavior – Interactions with Income and Life Expectancy*

Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	(1)	(2)	(3)	(4)
	Women	Men	Women	Men
Income (Durables) × Log HIV	-0.0021^{**}	-0.011^{***}	-0.010^{***}	-0.022^{**}
	(.0007)	(.0024)	(.0021)	(.0084)
Yrs. Left. × Log HIV	-0.0018^{***}	-0.0006	-0.0011	0.0005
	(.0005)	(.0010)	(.0013)	(.0040)
Number of Observations	83,627	22,631	83,623	22,563

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the Monte Carlo procedure outlined in Section 3. Second stage regressions with controls are shown in Appendix Table W.2.3 in web Appendix W.2. Other controls include: standard demographics, latitude, longitude, country dummies and country interacted with HIV, age and marital status interacted with HIV. Income is the number of different types of durable goods reported, plus a dummy for whether the house has electricity. Years left is the expected number of years of life left. This is based on child mortality levels by region and mortality patterns from the INDEPTH Model Life Tables. The level effects of HIV is not reported, since by including interactions between distance and country dummies I am effectively allowing each country to have its own HIV response on average, so the overall average does not have any obvious interpretation.

Table 6. *HIV Rate and Sexual Behavior – Interactions with Malaria and Maternal Mortality*

Panel A: Malaria				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0007	−0.0046	−0.0072**	−0.0049
	(.0011)	(.0041)	(.0027)	(.0183)
Malaria × Log HIV	0.057***	0.111**	0.141***	0.458***
	(.012)	(.038)	(.023)	(.120)
Implied Effect of Yrs. Left	**-0.0033**	**-0.0065**	**-0.0081**	**-0.0272**
Number of Observations	43,043	11,854	43,043	11,808
Panel B: Maternal Mortality				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0048***	−0.0085*	−0.0155**	−0.0395
	(.0012)	(.0054)	(.0037)	(.0364)
Maternal Mortality × Log HIV	−0.458**	−0.059	−0.826**	0.768
	(.169)	(.479)	(.317)	(1.619)
Age <21 × Income (Durables) × Log HIV	0.0017	−0.0012	0.0008	0.0005
	(.0015)	(.0079)	(.0062)	(.0376)
Age <21 × Maternal Mortality × Log HIV	0.493***	−0.523*	1.774***	−1.025
	(.137)	(.335)	(.423)	(1.003)
Implied Effect of Yrs. Left	**-0.0070**		**-0.0252**	
Number of Observations	28,071	58,346	13,505	15,851

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the a Monte Carlo procedure. Second stage regressions with controls are shown in Appendix Tables W.2.4 and W.2.5, in web Appendix W.2. Other controls: main effect of HIV, standard demographics, latitude, longitude, country dummies and country interacted with HIV, age and marital status interacted with HIV. "Malaria" is malaria prevalence in the region reported by the MaraLite database. In Panel B, the omitted age category is 21-35. Maternal mortality is the share of women who die in childbirth in the region, based on sibling mortality histories in the DHS. Income is the number of different types of durable goods reported, plus a dummy for whether the house has electricity. The level effect of HIV is not reported, since by including interactions between distance and country dummies we are effectively allowing each country to have it's own HIV response on average.

Table 7. *HIV Rate and Sexual Behavior – Controls for Knowledge*

Panel A: Years Left				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0020**	−0.0101***	−0.0102***	−0.0207*
	(.0009)	(.0029)	(.0027)	(.0112)
Years Left × Log HIV	−0.0023***	−0.0013	−0.0027*	0.0015
	(.0006)	(.0012)	(.0015)	(.0056)
Knowledge × Log HIV	−0.0008	−0.0119*	0.0009	−0.0162
	(.0023)	(.0069)	(.0052)	(.0276)
Number of Observations	83,627	22,631	83,623	22,563
Panel B: Malaria				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0005	−0.0041	−0.0067**	−0.0037
	(.0012)	(.0040)	(.0027)	(.0178)
Malaria × Log HIV	0.055***	0.114**	0.137***	0.460***
	(.011)	(.037)	(.024)	(.119)
Knowledge × Log HIV	0.0017	−0.0113*	0.0110	−0.0323
	(.0035)	(.0098)	(.0074)	(.0269)
Number of Observations	43,043	11,854	43,043	11,808
Panel C: Maternal Mortality				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0039**	−0.0018	−0.0132**	−0.0260
	(.0016)	(.0082)	(.0041)	(.0486)
Maternal Mortality × Log HIV	−0.501***	−0.209	−0.821**	0.660
	(.167)	(.482)	(.347)	(1.700)
Age <21 × Income (Durables) × Log HIV	0.0017	−0.0081	−0.0025	−0.0082
	(.0017)	(.0112)	(.0072)	(.0539)
Age <21 × Maternal Mortality × Log HIV	0.536***	−0.090*	2.058***	0.547
	(.140)	(.376)	(.374)	(1.034)
Knowledge × Log HIV	−0.0033	−0.0588**	−0.0168**	−0.1572**
	(.0038)	(.0121)	(.0081)	(.0400)
Number of Observations	28,071	58,346	13,505	15,851

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the a Monte Carlo procedure. Second stage regressions with controls are shown in Appendix Tables W.2.6, W.2.7 and W.2.8 in web Appendix W.2. Other controls: main effect of HIV, standard demographics, latitude, longitude, country dummies and country interacted with HIV, age and marital status interacted with HIV. Knowledge is based on response to two questions: can you get HIV from not using a condom and can you get AIDS from having multiple sexual partners. Malaria and maternal mortality are measured as described in Table 6; the measure of years left is described in Table 5.

Table 8. *HIV Rate and Sexual Behavior – Controls for Education*

Panel A: Years Left				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0003	−0.0077***	−0.0068***	−0.0156**
	(.0008)	(.0025)	(.0022)	(.0073)
Years Left × Log HIV	−0.0016***	−0.0002	−0.0006	0.0015
	(.0005)	(.0010)	(.0013)	(.0041)
Education × Log HIV	−0.0074***	−0.0151***	−0.0183***	−0.0292***
	(.0013)	(.0038)	(.0028)	(.0104)
Number of Observations	83,627	22,631	83,623	22,563
Panel B: Malaria				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	0.0011	−0.0012	−0.0047**	−0.00004
	(.0011)	(.0040)	(.0023)	(.0184)
Malaria × Log HIV	0.048***	0.090**	0.129***	0.418***
	(.011)	(.036)	(.024)	(.112)
Education × Log HIV	−0.0090***	−0.0158**	−0.0114***	−0.0237*
	(.0021)	(.0064)	(.0037)	(.0134)
Number of Observations	43,043	11,854	43,043	11,808
Panel C: Maternal Mortality				
Dependent Var.	*> 1 Partner (0/1)*		*# of Partners*	
	Women	Men	Women	Men
Income (Durables) × Log HIV	−0.0014	−0.0006	−0.0074**	−0.0232
	(.0011)	(.0050)	(.0029)	(.0374)
Maternal Mortality × Log HIV	−0.459***	−0.162	−0.845**	0.574
	(.151)	(.508)	(.316)	(1.731)
Age <21 × Income (Durables) × Log HIV	−0.0006	−0.0107	−0.0057	−0.0176
	(.0013)	(.0079)	(.0058)	(.0420)
Age <21 × Maternal Mortality × Log HIV	0.474***	−0.581*	1.745***	−1.103
	(.126)	(.350)	(.420)	(1.112)
Education × Log HIV	−0.0145***	−0.0323**	−0.0333***	−0.0699*
	(.0022)	(.0061)	(.0052)	(.0106)
Number of Observations	28,071	58,346	13,505	15,851

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the a Monte Carlo procedure. Second stage regressions with controls are shown in Appendix Tables W.2.9, W.2.10 and W.2.11, in web Appendix W.2. Other controls: main effect of HIV, standard demographics, latitude, longitude, country dummies and country interacted with HIV, age and marital status interacted with HIV. Education is in categories: no education, primary education and secondary education. Malaria and maternal mortality are measured as described in Table 6; the measure of years left is described in Table 5.

Table 9. ***Regressions with Repeated Cross Sectional Data***

Dependent Var.:	*Extramar. Sex (0/1)*		*Premarital Sex (0/1)*	
	(1)	(2)	(3)	(4)
	Women	Men	Women	Men
Explanatory Variables:				
Log HIV Rate	−.021	.4419	.5962***	1.5843***
	(.104)	(.369)	(.15)	(.438)
Income × Log HIV	−.0014*	−.0129**	−.0091***	−.0129***
	(.001)	(.005)	(.003)	(.004)
Yrs Left × Log HIV	−.0037*	−.0054	−.0081**	−.0171***
	(.002)	(.005)	(.003)	(.003)
Income	−.0012	.0203**	−.0027	.0149**
	(.001)	(.009)	(.004)	(.005)
Yrs Left	.0095***	.0002	.0016	.0062**
	(.003)	(.005)	(.003)	(.003)
Muslim (0/1)	−.002	−.0235*	−.0223	−.0109
	(.002)	(.014)	(.015)	(.017)
Education	.0056	.0225	.008	.032**
	(.006)	(.014)	(.012)	(.012)
Urban	.0106	.012	.0546***	.0316**
	(.007)	(.011)	(.012)	(.014)
Work (0/1)	−.0021	.0109	.0513***	.0915***
	(.003)	(.01)	(.009)	(.017)
# Kids	−.0035***	.0017**	.048***	.0826**
	(.001)	(.001)	(.017)	(.031)
year	.0141*	−.0192*	−.0143*	−.0578**
	(.007)	(.011)	(.007)	(.021)
constant	−.278*	.876**	−2.596***	−2.855***
	(.143)	(.348)	(.39)	(.288)
Region FE	YES	YES	YES	YES
Number of Observations	65,787	16,448	32,219	12,426
R^2	.07	.12	.2	.28

standard errors in parentheses, clustered by country-year; * significant at 10%; ** significant at 5%; *** significant at 1%

Notes: All regressions also include a cubic in age, and age interacted with HIV rate. In the first two columns, the dependent variable is equal to one if the individual (married) reports any extramarital sexual activity in the last year. In Columns 3 and 4, it is a dummy equal to one if the individual (unmarried) reports any sexual activity in the last year, and zero otherwise. Income is simply the sum of number of durable goods owned and a dummy for having electricity. Years left is the expected number of years of life left. This is based on child mortality levels by region and mortality patterns from the INDEPTH Model Life Tables. Education is in categories: no education, primary education and secondary education. Columns 1 and 2 are limited to countries with more than two observations after 1995.

Table 10. ***Effect of the Price (Utility Cost) of Sexual Partners on Behavior***

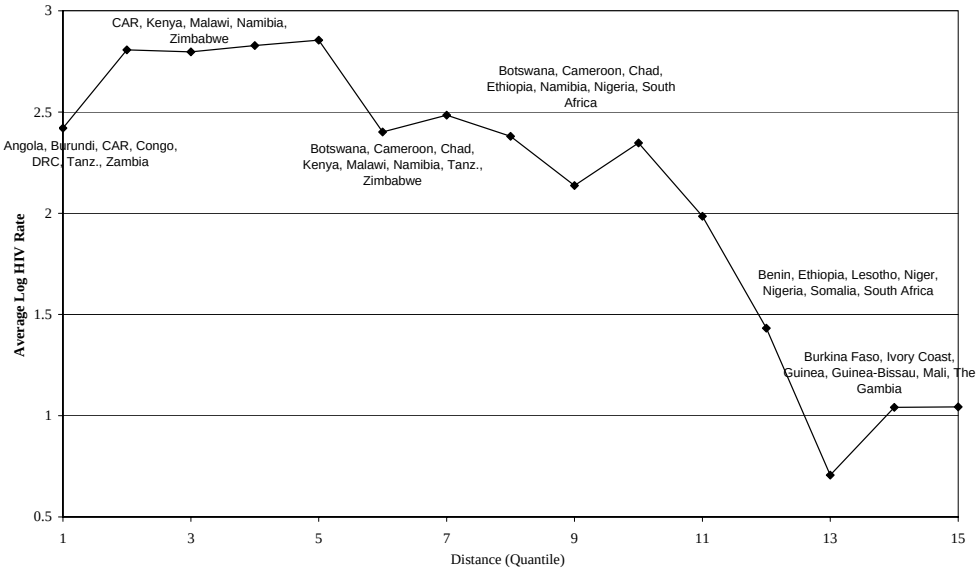
	Dependent Variable: More than One Partner (0/1)			
	Women, Africa	Men, Africa	Gay Men, US	
			Perceived HIV =Actual HIV	*Linear Increase in Perceived HIV*
Utility Cost of Sex	-0.0045^{***}	-0.018^{***}	-0.209^{***}	-0.0324^{***}
	(.0011)	(.0052)	(.0397)	(.0038)
Number of Observations	82,798	22,356	24,502	24,502

standard errors in parentheses, clustered by region

* significant at 10%; ** significant at 5%; *** significant at 1%

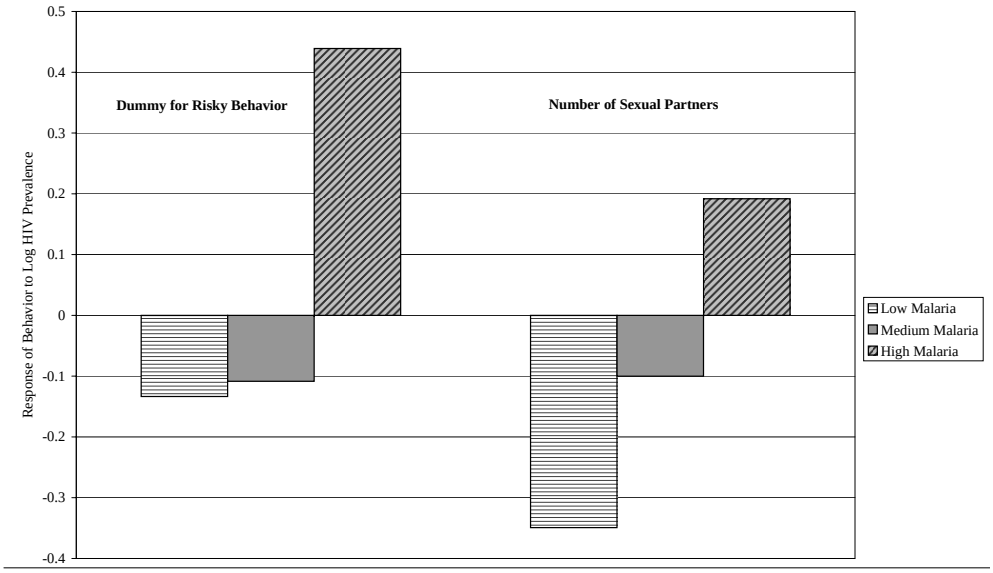
Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the Monte Carlo procedure outlined in Section 3. Second stage regressions with controls (although not adjusted standard errors) are shown in Appendix Table W.2.12, in web Appendix W.2. Other controls include: main effect of HIV, standard demographics, latitude, longitude, country dummies and country interacted with HIV, age and marital status interacted with HIV. Income is imputed from durable good ownership and GDP, using the relationship from the South African and Ivory Coast LSMS. For Africa, survival path is based on mortality patterns from the INDEPTH Model Life Tables. In Columns 3 and 4 the data on gay men is from the MACS, limited to the period before 1990 and to uninfected people. Both regressions include individual-specific fixed effects. The HIV rate used in Column 3 is the actual rate in the sample; in Column 4, I assume that the HIV rate perceived (the relevant rate) increases linearly from the start of the sample to reach the actual rate (40%) in 1990. Income is imputed by matching individuals to the average income in the Census IPUMS, based on education and occupation group. Survival path for gay men is based on the U.S. Census life tables for 1980.

Figure 1:
Log HIV Rate and Distance from Origin



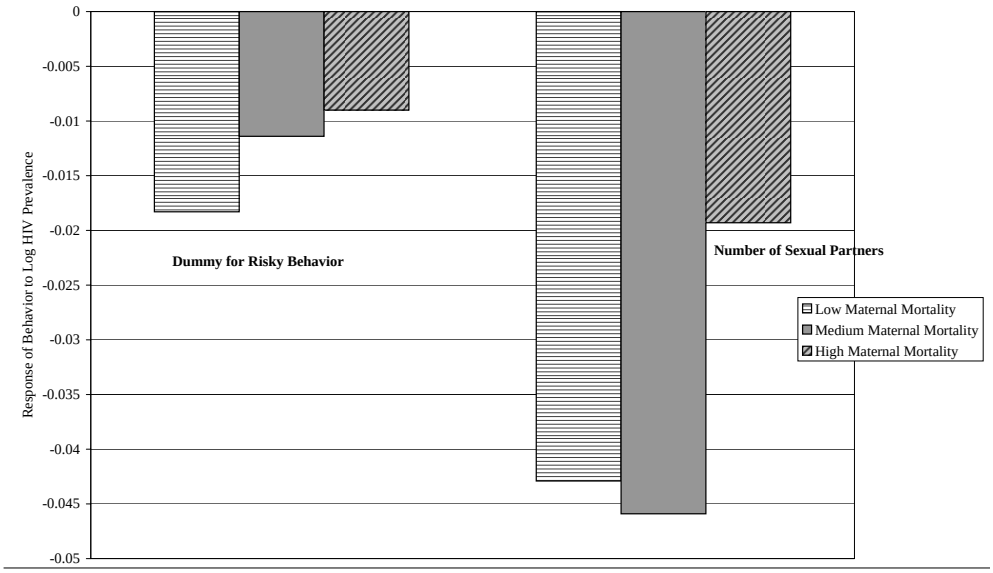
Notes: Figure shows the relationship between log HIV prevalence (by region) and distance quantile. The countries with regions in each quantile are listed for every third quantile. A full list of regions appears in Appendix W.1, available on the authors website.

Figure 2:
Response to HIV by Malarial Prevalence



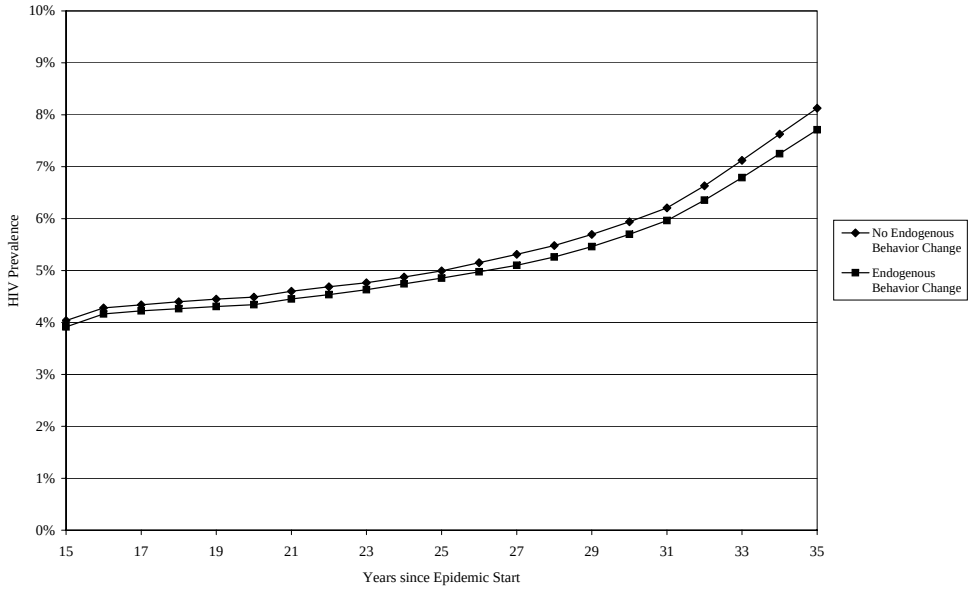
Notes: This figure shows the effect of HIV rate on sexual behavior (the coefficient from a regression) in regions with low, medium and high malarial prevalence relative to their country. Areas with low prevalence are defined as the bottom 1/3 of regions in each country, medium are the middle 1/3 and high are the highest 1/3 in each country.

Figure 3:
Response to HIV by Maternal Mortality



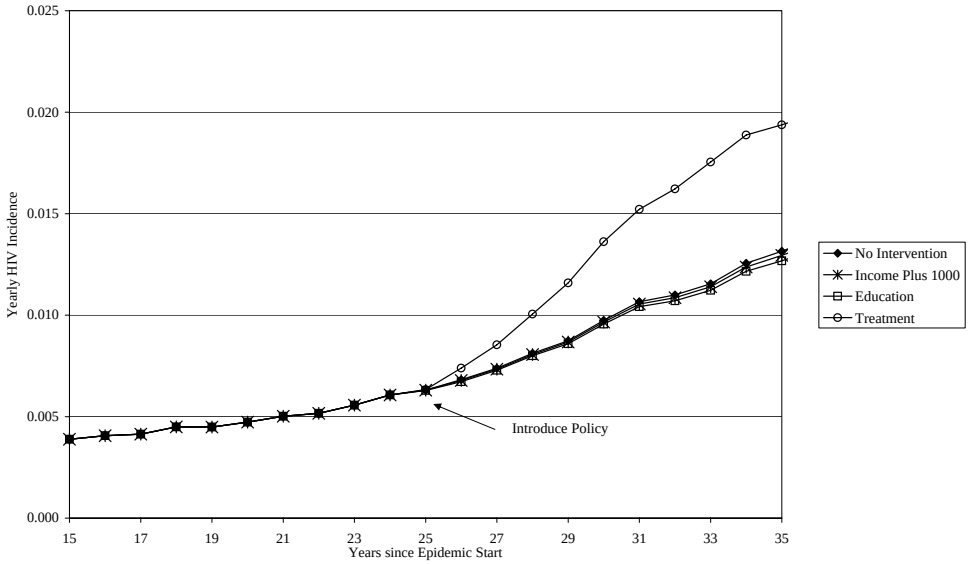
Notes: This figure shows the effect of HIV rate on sexual behavior for young women relative to old women in regions with low, medium and high maternal mortality relative to their country. Areas with low mortality are defined as the bottom 1/3 of regions in each country, medium are the middle 1/3 and high are the highest 1/3 in each country.

Figure 4:
HIV Prevalence, With and without Endogenous Behavior Change



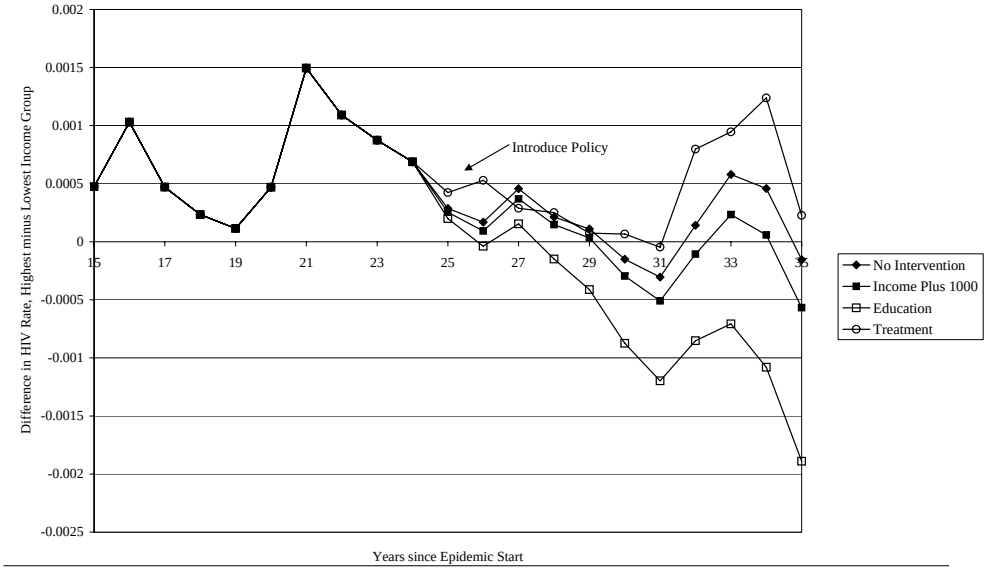
Notes: This figure shows the yearly HIV prevalence from a simulation model (described in Section 5), with and without allowing for endogenous behavioral response of the magnitude estimated in Section 5.

Figure 5:
Yearly HIV Incidence, Simulated Under Intervnetions



Notes: This figure shows the yearly HIV incidence from a simulation model, for a baseline case and three interventions (described in Section 5). Treatment is introducing drug treatment to everyone, doubling time to death with HIV; education is modeled as improving individual estimates of the HIV rate (increasing them); income improvement is giving everyone an extra $1000 per year in income. Incidence is defined as the number of new infections during the year divided by the uninfected population at the start of the year.

Figure 6:
Difference in HIV Prevalence, High and Low Income Groups



Notes: This figure shows changes over time in the difference in prevalence rates between highest income and lowest income groups in the simulation for a baseline case and three interventions (described in Section 5). Treatment is introducing drug treatment to everyone, doubling time to death with HIV; education is modeled as improving individual estimates of the HIV rate (increasing them); income improvement is giving everyone an extra $1000 per year in income.

Appendix A: Income and Sex Utility

As noted in Section 2, the comparative static on income will be negative only if $u_{\sigma\sigma}(y,\sigma_1)\,u_y(y,\sigma_2) - u(y,\sigma_2)\,u_{\sigma\sigma y}(y,\sigma_1) < 0$. Whether or not this holds depends largely on the relationship between income and sexual behavior. This appendix presents some evidence on whether this assumption is reasonable.

I note first that this condition will hold with certainty if $u_{\sigma\sigma y} = 0$. This is difficult to test directly. However, a straightforward (but stringent) test asks whether $u_{\sigma y} = 0$; if this holds then $u_{\sigma\sigma y} = 0$ and this condition is easier to test. Abstracting away from HIV risk, individuals will choose σ such that $u_\sigma = 0$. The values of σ that we observe chosen will tell us something about $u_{\sigma y}$: if u_σ depends on y, so $u_{\sigma y} \neq 0$, then we should see the observed σ depending on income.

I test this by regressing the choices of sexual behavior on income. The results above suggest that any relationship between these will be mediated by HIV, implying that the number of sexual partners in Africa is probably not a good dependent variable to test this. Instead, I consider the relationship between income and frequency of sex within marriage. If richer people have a higher utility for sex, they should have sex more with their spouse. The results of this regression are shown in Appendix Table 1, which displays regressions of frequency of marital sex for women on income and a number of additional controls. The results point to virtually no relationship between income and sexual frequency – a increase of $10,000 a year in income (most of the range of the data) is predicted to increase sexual encounters by 0.58 per year, on a base of 43. The small magnitude and insignificance of this relationship suggests that $u_{\sigma y}$ is at least close to zero and supports the separability assumption.

Appendix Table 1. *Income and Frequency of Marital Sex*

Dependent Variable: Frequency of Marital Sex	
Explanatory Variables:	
Income (in $10,000)	.5818
	(.36)
Age	$-.5025^{***}$
	(-8.34)
Education	$.6586^{***}$
	(5.03)
Urban	.4928
	(.59)
# Kids	$-.6783^{***}$
	(-2.99)
# Kids Home	$-.906^{***}$
	(-3.38)
Muslim	1.3985
	(1.36)
Currently Pregnant	-12.7569^{***}
	(-13.38)
Birth in Last Year	-7.0889^{***}
	(-7.23)
Husband at Home	34.9292^{***}
	(35.75)
constant	35.235^{***}
	(19.5)
Number of Observations	21471
R^2	.12

[a] t-statistics in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: The dependent variable is imputed frequency of sex in marriage, based on reports of the last time you had sex with your spouse and assuming that the survey is randomly placed relative to sex. Income is predicted based on durable good ownership and GDP, using data from the South African and Ivory Coast Living Standard Measurement Surveys.

A more sophisticated way to address this is to directly estimate the degree of complementarity between income and sex from the data. In particular (abstracting away from any dependence on p) assume that individuals have a CES utility function over consumption and sexual activity:

$$u(c, \sigma) = \left(\alpha c^{-\rho} + (1-\alpha)\,\sigma^{-\rho}\right)^{-\frac{1}{\rho}}$$

If a given sexual encounter has a price r, so $c = y - r\sigma$ then, with an assumption about r, it is possible to use the data on income and marital sex frequency to estimate the α and ρ which best fit the data.[17]

[17] r is the price per sexual act. I assume that $r = 3$, based on Pikering et al, 1997, who report this as the average price per encounter, based on interviews with sex workers in Uganda.

Using the same data on marital sex frequency and income, I estimate the best fit α and ρ. If both are allowed to vary, the best fit estimates are $\alpha = 0.65$ and $\rho = -0.2$. If we assume that $\alpha = 0.5$ and estimate only ρ, I find $\rho = -0.7$. Using these values of ρ, along with average income, it is clear that $u_{\sigma\sigma}u_y - u(y)u_{\sigma\sigma y} < 0$. This is true for the whole range of HIV rates. Looking carefully at the components of this expression, it is clear why this is true. With these values of ρ, $u_{\sigma\sigma y}$ is negative but extremely close to zero. With $\rho = -0.7$, for example, we are very close to the assumption of separability, where $u_{\sigma\sigma y} = 0$.[18]

Both pieces of evidence here suggest that the assumption required for the comparative static on income in Section 2 to hold is likely to be satisfied empirically. In addition, the calibration of ρ suggests that, at least in the data here, it may not be very inappropriate to assume that utility for income and sex are separable, as is done in Section 5.

Appendix B: Instrumental Variables Details and Robustness

B.1. Functional Form Relationship Between HIV and Distance

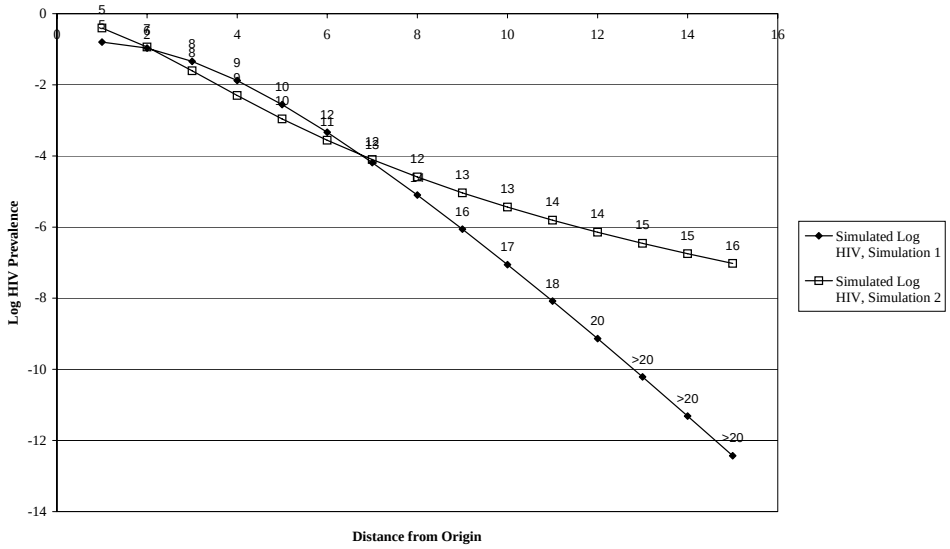
This subsection briefly discusses the choice of functional form for the relationship between HIV and distance. Ex ante, it is not obvious what the shape of that relationship might be. To get some sense of the most appropriate relationship, I develop a simple simulation model of epidemic spread, relying on two different assumptions about the relationship between distance between individuals and their probability of interaction.

Assume that individuals are arrayed discretely (i.e. some individuals at point 1, some at point 2, etc) along a line of length n, where the distance between any two individuals i and j is d_{ij}. The key to the simulation is that the chance that individuals have a sexual relationship is declining (according to some function $f(d_{ij})$) as they are farther away from each other. I assume that if two individuals meet for a sexual relationship, and one of them is infected with HIV, the disease is passed to the other individual with some probability p. I assume that HIV is introduced to one individual at one point along the line at time 0, and then follow the disease over time.

I make two possible assumptions about the functional form of $f(d_{ij}$. First, I assume $f(d_{ij}) = p^{1+dist*.25}$; second, $f(d_{ij}) = \frac{p}{dist^4}$. Appendix Figure B.1. below shows the relationship between log HIV rate and distance after 20 periods for both of these functional form assumptions. The relationship is downward sloping and roughly linear, suggesting that if we use this functional form we should expect the linear regression to fit well. The figure also shows the first date at which the virus is observed at at least 0.1% in each area, which is clearly later in areas further from the origin. This makes explicit the link between distance and time discussed in Section 3.

[18] An alternative way to get these estimates would be to use the number of sexual partners for people in low HIV countries, under the assumption that their behavior is relatively unaffected by HIV. If I do this, the estimated ρ is extremely similar: assuming that $\alpha = 0.5$, I estimate $\rho = -0.6$.

Appendix Figure B1:
Simulated Relationship Between Distance and Log HIV Prevalence



Notes: The figure shows simulated HIV rates, based on simulations outlined in Appendix B, graphed against distance. The graph shows HIV rates 15 years into the simulated epidemic. The number labels show the first year at which HIV rate in that distance is at least .1% in the simulation.

B.2. Alternative Data on Prevalence

The first stage regressions use data from the U.S. Census HIV/AIDS Surveillance Database. This data is useful in the sense that it covers a wide swath of areas. However, it presents clear issues of comparability across space since the data comes from many different sources. As long as this is mostly random, that will just induce larger standard errors, but it is worth considering whether more precise data show similar results.

I test for the HIV-distance relationship using two other data sources, both of which are smaller sample but more comparable across space. The first is the DHS population testing data in six countries (Burkina Faso, Cameroon, Ghana, Kenya, Mali, Zambia). This gives prevalence estimates by region within the country which are clearly comparable over space. The second data source is prevalence for regions across nine countries derived based on mortality data (Oster, 2007). Again, since the technique for deriving prevalence is the same over space, these estimates are consistently comparable.

Appendix Table 2 shows regressions of log HIV rate on distance for the DHS (Columns 1 and 2) and the mortality data (Columns 3 and 4). Columns 1 and 3 contain no controls, and Columns 2 and 4 control for latitude, longitude and urbanization. In all four columns, the effect of distance is strongly statistically significant, and is of similar size to what is estimated in the primary first stage regressions in Table 2.

Appendix Table 2. *HIV and Distance, Alternative Data Sources*

Dependent Variable: Log HIV Prevalence in Region				
	(1)	(2)	(3)	(4)
Estimates From:	DHS	DHS	Mortality Data	Mortality Data
Explanatory Variables:				
Distance (in 1000 km)	$-.6792^{***}$	$-.5766^{**}$	$-.6885^{***}$	$-.5391^{**}$
	(.087)	(.23)	(.1)	(.223)
Latitude		$-.003$		$-.0366^{***}$
		(.019)		(.012)
Longitude		.0032		$-.0166$
		(.013)		(.013)
Urban Pct.		1.1471^{***}		$-.1031$
		(.312)		()
constant	2.716^{***}	2.043^{***}	2.6^{***}	2.466^{***}
	(.242)	(.681)	(.225)	(.715)
Number of Observations	59	59	48	48
R^2	.52	.61	.51	.60

[a] standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix C: Other Behavioral Outcomes

In the primary analysis, I have focused exclusively on the choice of whether or not to have multiple sexual partners. In many ways, this is the most obvious risky behavior and it allows for a natural cross sample comparison. However, there are other behaviors that we might expect to be affected by the epidemic. In this section, I briefly discuss the effect of the epidemic on condom use and choice of partner type.

Condom Use

A large part of the response to HIV in the developed world has been an increase in the use of condoms. In the developing world, overall condom use is much lower (only 19% of men and 6% of women in the DHS survey report using a condom with their last partner). However, it is still possible that condom use responds to the epidemic.

Measures of condom use in the DHS are imperfect. It would be ideal to know what share of the time individuals use condoms with their partners; what type of parters they are more likely to use them with and so on. Unfortunately, questions like this are not posed in the surveys. I use a proxy for condom use: reporting condom use during the last sexual encounter. This is the most direct measure of condom use available. It does not capture consistency of use, which is potentially a problem, but it should at least be correlated with overall rates of usage. In all regressions, I limit the sample to single individuals because the choice of condom use among married individuals is much more complicated, primarily because it involves issues of fertility.

Panel A of Appendix Table 3 reports the responsiveness of condom use to the HIV rate. The results are quite weak; there are no statistically significant effects on condom usage. The coefficients are not consistent directionally (although, on average, both genders increase condom use in response to HIV). Taken together with the previous results, this might suggest that people are more willing to substitute away from multiple partners than towards condom use, which seems somewhat surprising. It is perhaps more plausible that this simply reflects the relatively noisy measure of condom use used here.

Partner Type

The risk of HIV infection depends not just on how many partners an individual has and whether they use condoms, but also on what kind of parters these are. More casual partnerships – brief acquaintances or commercial sex workers – are likely to be more risky because the partner's sexual history is unknown and selection would suggest they are more promiscuous. Given this, we might expect to see substitution away from these more risky partnerships in response to HIV.

Again, I consider the effect of the HIV rate, both alone and interacted with income and child mortality on whether the individual has had sex with a casual acquaintance or commercial sex worker in the last year (as opposed to a spouse, boyfriend, fiance or other longer-term partner).[19] The results are reported in Panel B of Appendix Table 3. For women the mean of the variable is extremely small; fewer than 1 percent of women report sex with one of these groups and we generally see no movement in partner type. For men, there is some suggestive evidence of response: overall, men seem to decrease their partnerships with unknown partners in response to increases in the HIV rate. And this response is larger for those who are richer and larger for those with longer life

[19]The analysis uses all individuals in the sample who are sexually active, not just those with non-marital partners. The results are very similar if we restrict to those with non-marital partners.

expectancy. However, as with the results on condom use, these estimates are very imprecise and none are statistically significant.

Appendix Table 3. ***Response of Condom Use, Partner Type to HIV Rate***

Panel A: Condom Use				
Dependent Variable: Used Condom in Last Sex (0/1)				
	Single Women	Single Men	Single Women	Single Men
Log HIV Rate	0.0333	0.162		
	(.109)	(.114)		
Income (# Durables) × Log HIV			−0.0061*	0.0020
			(.0035)	(.0058)
Yrs Left × Log HIV			−0.00017	−0.0054**
			(.0014)	(.0026)
Number of Observations	9117	6105	8640	3574
Panel B: Partner Type				
Dependent Variable: Sex with Risky Partner Type (0/1)				
	Women	Men	Women	Men
Log HIV Rate	−0.004	−0.029		
	(.0029)	(.047)		
Income (# Durables) × Log HIV			0.0001	−0.0010
			(.0002)	(.0012)
Yrs Left × Log HIV			−0.0001	−0.00007
			(.0001)	(.0009)
Number of Observations	61,712	21,321	59,212	16,486

standard errors in parentheses, clustered by region. * significant at 10%; ** significant at 5%; *** significant at 1%

Notes: Notes: First stage regressions appear in Table 2. Coefficients and standard errors here are produced using the Monte Carlo procedure outlined in Section 3. Income is the number of different types of durable goods reported, plus a dummy for whether the house has electricity. Years left is the expected number of years of life left. This is based on child mortality levels by region and mortality patterns from the INDEPTH Model Life Tables. Other controls include: main effect of HIV, standard demographics, latitude, longitude, country dummies and country interacted with distance, age and marital status interacted with distance. The dependent variable in Panel A is a dummy for reporting condom use with the last partner. The dependent variable in Panel B is a dummy equal to 1 if the individual reports sex in the last year with either a casual acquaintance or a commercial sex worker, and zero if they are sexually active but do not report that. In Columns 3 and 4, the level effects of HIV are not reported, since by including interactions between distance and country dummies we are effectively allowing each country to have it's own HIV response on average.